

AISB2016

Symposium sur les principes de la robotique

4 avril 2016, Sheffield, Royaume-Uni.

Édité par
Tony J Prescott

Comité d'organisation
Tony J Prescott
Alan Winfield
Madeleine de Coq Buning
Joanna J Bryson
Noël Sharkey

Partie de la Convention 2016 de la Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB)

À propos du Symposium

Cinq ans se sont écoulés depuis la publication des « Principes de la robotique »¹ élaborés par un groupe d'éminents experts britanniques en robotique et en intelligence artificielle lors d'une retraite financée par l'EPSRC/AHRC.

1. Les robots sont des outils à usages multiples. Les robots ne devraient pas être conçus uniquement ou principalement pour tuer ou blesser des êtres humains, sauf dans l'intérêt de la sécurité nationale.
2. Les êtres humains, et non les robots, sont les agents responsables.
3. Les robots sont des produits.
4. Les robots sont des artefacts manufacturés. Ils ne devraient pas être conçus dans une manière trompeuse d'exploiter les utilisateurs vulnérables ; au lieu de cela, leur nature de machine devrait être transparente.
5. La personne légalement responsable du robot doit être désignée.

Les principes ont eu un impact significatif dans la recherche robotique au Royaume-

Uni et continuent de provoquer des débats substantiels. À un moment où l'inquiétude du public concernant le développement des technologies

- b. Suffisance/Généralité – le principe est-il suffisant et couvre-t-il tous les problèmes importants qui pourraient survenir dans la réglementation de la robotique dans le monde réel ou des préoccupations importantes sont-elles ignorées ?
- c. Utilité — sont les principes d'utilisation pratique pour les développeurs de robots, les utilisateurs ou la loi

décideurs, indétérminant les stratégies pour les meilleures pratiques en robotique, ou les normes ou les cadres

Un symposium d'un jour a eu lieu le 4 avril dans le cadre de la conférence AISB 2016 à Sheffield, au Royaume-Uni. Ces actes contiennent des commentaires sur le principe qui ne sont pas sollicités avant la réunion.

¹ <https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>

AISB2016SymposiumonthePrinciplesofRobotics

SoumisCommentaires

- 1 Joanna Bryson
LasignificationdesprincipesEPSRCdelarobotique
- 2 Amadou Gning, Darryl Davis, Yongqiang Cheng et Peter Robinson
Robotiquerechercheéthiquediscussion
- 3 Vincent Muller
Règles juridiques, exigences éthiques, enjeux futurs
- 4 Tony Prescott
Les robots ne sont pas que des outils
- 5 Michel Szollosy
Défendre un human(isme) obsolète ?
- 6 Aurore Voiculescu
Réguler les robotstowns : Réflexionssurlesprincipesdelarobotiquedepuis
nouveau côté éloigné de la loi
- 7 Paula Boddington
Commentairesurlaresponsabilité,laconceptionduproduitetlesnotionsdesécurité
- 8 RoelanddeBruinandMadeleinedeCoqBuning
Commentaireduprincipe2
- 9 Burkhard Shafer et Lilian Edwards
Traitement équitable des données et robotique
- 10 Amanda Sharkey
Les robots peuvent-ils être des agents moraux responsables ? Et pourquoi devrions-nous nous en soucier ?
- 11 Tom Sorell et Heather Draper
Réflexions sur la vie privée, la sécurité et la tromperie
- 12 Emilie Collins
Commentaireduprincipe4
- 13 Andreas Theodorou, RobertWortham et Joanna Bryson Pourquoi mon
robot se comporte-t-il comme ça ? Concevoir la transparence pour l'inspection en temps réel
de robots autonomes
- 14 Robert Wortham, Andreas Theodorou et Joanna Bryson
Robottransparence,confianceetutilité

Le sens des principes EPSRC de la robotique

Joanna Bryson, Université de Bath et Princeton Center for Information Technology Policy

Introduction

En revisitant les principes de la robotique, il est important d'examiner attentivement leur pleine signification. Ici, je visite brièvement d'abord le sens du document dans son ensemble, puis celui de ses éléments constitutifs. Les principes EPSRC de la robotique ont été générés sous forme de livrable par un groupe assemblé avec peu de conseils et aucun livrable requis. L'intention initiale de l'événement robotique epsrc semble avoir été uniquement la discussion elle-même, ou peut-être même uniquement le fait de la réunion. Les universitaires présents voulaient quelque chose à montrer pour le temps qu'ils avaient passé, et par conséquent, une quantité substantielle de temps de tous ceux qui étaient présents le dernier jour a été consacrée à la création des [trois versions des principes et de leur documentation](#). Une partie de la [documentation](#) a été prolongé – là encore par consensus – après

la réunion. Il est juste et approprié qu'il y ait un moyen d'examiner et même de mettre à jour ou de maintenir le document. Même les constitutions nationales ont des moyens d'entretien. Cependant, il est essentiel pour l'efficacité des documents de politique qu'ils ne soient pas faciles à modifier. Ils devraient fournir un gouvernail pour empêcher le tramage et, en tant que tels, sont généralement plus difficiles à modifier qu'ils ne l'étaient à instancier en premier lieu. Notez que certains pays et d'autres unions politiques n'ont pas trouvé facile de créer même leurs constitutions initiales pour cette raison même. Il est donc important de bien réfléchir à la signification des principes.

Les principes comme politique

La politique technologique, et la politique plus généralement, est une chose étonnamment amorphe. Comme d'autres aspects de l'intelligence naturelle, la politique ne réside pas toujours dans la loi ou même dans la gouvernance. Une grande partie de la politique n'est pas écrite et même pas explicitement connue. Le Royaume-Uni est en fait remarquable dans son innovation de la common law, qui reconnaît cela et l'importance de la culture et des précédents. Néanmoins, à la lumière froide d'un comité travaillant sur des cas d'impact REF, nous devons nous demander si les principes sont politiques ? Je pense que la réponse est oui". Il s'agit d'un ensemble de directives convenues par une fraction substantielle, voire arbitraire, de la communauté qu'elles affectent, et elles sont publiées sur les pages Web du gouvernement. Toute politique a trois composantes : allocative, distributive et stabilisatrice. L'allocation est le processus de détermination des problèmes qui valent la peine de consacrer du temps et d'autres ressources. Dans le cas des principes, cela a été initié par l'EPSRC (ou une organisation au-dessus d'eux) par crainte que le public britannique puisse rejeter la robotique car ils avaient des aliments génétiquement modifiés. On nous a dit que le rejet de la robotique était considéré comme une grave menace pour l'économie britannique. Notez également que chacun des participants (du moins ceux qui n'étaient pas spécifiquement payés pour y assister) a également fait des investissements individuels, allouant du temps au problème de l'éthique des robots, même si pour beaucoup cela a été confondu avec une opportunité de se faire mieux connaître par leur

principal organisme de financement. L'élément stabilisateur est celui qui garantit que la politique, une fois définie, est intégrée dans la société de telle manière qu'il est peu probable qu'elle soit rapidement annulée ou qu'elle devienne un

une responsabilité ou un sujet de controverse. Dans le cas des principes, cela a évidemment été réalisé au moins dans une certaine mesure depuis que nous célébrons leur cinquième anniversaire. D'après mes discussions avec d'autres auteurs, je n'en connais aucun entièrement amoureux du produit final, mais tous respectent le processus démocratique (certes représentatif) par lequel ils ont été réalisés et l'importance de l'engagement mutuel de leurs collègues envers le produit final. Pour ma part, j'aimerais voir les principes davantage réifiés en politiques ou même en lois, mais je n'ai pas encore découvert le processus par lequel cela pourrait être accompli. Cependant, ils ont été et continuent d'être portés à l'attention de divers comités de normalisation et d'enquêtes parlementaires ainsi que de la presse et d'autres universitaires. Je laisse pour la fin l'aspect le plus controversé de la politique : le distributif. À la base, toute politique consiste à sélectionner des actions, ce qui implique l'allocation ou plutôt la réallocation des ressources. La politique essaie d'effleurer cela, car cela va nécessairement à l'encontre de ceux dont les ressources sont réaffectées, même dans les cas où ces individus devraient en tirer un bénéfice net. Nous détestons perdre le contrôle, mais les politiques sont pour le contrôle. "Essayer d'effleurer" est en fait un euphémisme ; rendre la redistribution acceptable peut être le projet central des politiciens. Dans ce cas, le gouvernement avait des préoccupations très spécifiques concernant les personnes qui avaient été dans les médias faisant la promotion de la peur des robots, et étaient très clairs dans leur désir de trouver des moyens de détourner l'attention des médias et les impressions du public vers la sécurité de la robotique. En revanche, ce sont vraiment les participants qui ont évoqué les autres grands glissements du sensationnalisme au pragmatisme, l'affirmation que les robots ne sont pas responsables devant la loi, et qu'il ne faut pas tromper les utilisateurs sur leurs capacités. Les représentants du conseil savaient que cette redistribution du pouvoir mettrait en colère certains de leurs bénéficiaires de financement exceptionnels, et les participants savaient la même chose de certains de leurs collègues. Néanmoins, il y avait une unanimité frappante parmi les universitaires sur le fait que les plus grands risques moraux des robots étaient leur nature charismatique et l'incroyable empressement de nombreuses personnes à investir leur propre identité dans les machines, ce qui a conduit à une confusion frappante sur leur nature dont nous avons tous été témoins. Ce charisme et cette confusion ont laissé la porte ouverte à toutes sortes de manipulations par les entreprises et les gouvernements, où les robots pourraient être érigés en responsables ou même en substitut de vies ou de valeurs humaines.

Le principe de tuer

Les robots sont des outils polyvalents. Les robots ne doivent pas être conçus uniquement ou principalement pour tuer ou blesser des humains, sauf dans l'intérêt de la sécurité nationale.

Les trois premiers principes étaient destinés à corriger les lois d'Asimov. Les robots ne sont pas des parties responsables, ils ne peuvent donc pas tuer. Au lieu de cela, les robots ne devraient pas être utilisables comme outils pour tuer. Cette règle simple rendait clair le transfert de la subjectivité morale, et répondait simultanément aux désirs pacifistes de la plupart des présents. Cependant, pragmatiquement, les robots étaient déjà utilisés comme armes de guerre, et une loi inapplicable est discutable. Nous étions persuadés que diriger avec un principe connu pour être faux réduirait considérablement nos chances d'avoir un impact culturel. Le sens du premier principe pourrait donc sembler neutralisé par le compromis de l'exception, mais que les robots ne soient pas des armes dans la société civile reste un point social important. Au-delà de cela, le fait que la politique pratique doit prendre en compte les besoins du gouvernement pour répondre à la fois à la sécurité et à l'industrie (le Royaume-Uni est le cinquième plus grand pays marchand d'armes au monde) a également un sens. Aussi purement académiques que certains d'entre nous puissent

souhaite que notre discipline soit, le fait que nombre de ses produits aient une utilité immédiate signifie que nous ne pouvons pas éviter d'avoir un impact sur notre monde.

Le principe de conformité

Les humains, et non les robots, sont des agents responsables. Les robots doivent être conçus ; exploités dans la mesure du possible pour se conformer aux lois existantes et aux droits et libertés fondamentaux, y compris la vie privée.

La deuxième loi d'Asimov a à voir avec les instructions suivantes, mais même la notion d'obéir à une agence morale implicite. La signification originale de cette loi était que les robots sont une technologie ordinaire et se conforment aux normes et lois ordinaires. Dans la formation des principes en tant que suite, le deuxième principe est devenu celui qui communiquait davantage une partie du péril de l'IA en général, et de l'IA confondue avec un sujet moral en particulier. L'accent mis sur la vie privée reflète la préoccupation particulière d'un agent physique intelligent percevant occupant exactement le même espace qu'une famille humaine. La technologie est fondamentalement immergée dans l'environnement humain, plus que n'importe quelle technologie ou animal de compagnie antérieur, peut-être même plus que certains humains dans un foyer comme les enfants. Il a accès au langage écrit et parlé, aux informations sociales, aux horaires observés, etc. De plus, il peut être confondu avec un animal de compagnie ou un autre membre de la famille de confiance, ses capacités spéciales pour une communication parfaite avec le monde extérieur temporairement oubliées, ou ses capacités à apprendre des régularités. et classer les stimuli. Dans ces cas, les informations des primates peuvent être stockées involontairement dans un cloud public, ou même un cloud supposément privé susceptible d'être piraté. Forcer une telle technologie nouvelle et humaine à se conformer aux normes légales standard de confidentialité et de sécurité n'est pas une tâche triviale.

Le principe de banalisation

Les robots sont des produits. Ils doivent être conçus à l'aide de processus garantissant leur sûreté et leur sécurité.

La dernière loi Asimov est l'autoprotection, mais les robots n'ont pas de moi. Au lieu de cela, cette loi s'est concentrée sur la protection des humains contre les robots au niveau de la solidité de base du robot. Le principe nous amène à nouveau à prendre conscience de la nature manufacturée non spéciale du robot, dans une tentative d'éviter l'évitement de la responsabilité légale en affirmant que les robots ont une nature unique. Le fabricant d'un robot devrait avoir exactement autant de responsabilité pour les machines fonctionnant selon les spécifications que le fabricant d'une voiture ou d'un outil électrique. En fait, les robots peuvent être des voitures ou des outils électriques, mais si tel est le cas, ils devraient être plus sûrs que la variété conventionnelle des deux.

Le principe de transparence

Les robots sont des objets manufacturés. Ils ne doivent pas être conçus de manière trompeuse pour exploiter les utilisateurs vulnérables ; au lieu de cela, leur nature de machine doit être transparente.

Les trois premiers principes ont établi le cadre juridique de la fabrication et de la vente de robotique comme étant identique à d'autres produits. Les deux derniers visent à garantir que l'état est également communiqué à l'utilisateur. Le principe de transparence vise à garantir que les individus ne surinvestissent pas dans leur technologie, par exemple en engageant un home sitter pour éviter que le robot ne se sente seul. Certains roboticiens s'opposent à ce principe parce que la tromperie est nécessaire à l'efficacité de leur application prévue, comme faire en sorte que les gens ne se sentent pas seuls afin qu'ils soient moins déprimés. D'autres soutiennent que ce principe nie la possibilité que les robots soient plus que des machines ordinaires. Le premier argument est ouvert à l'expérimentation. Tout d'abord, il faut établir qu'il n'y a aucun moyen de déclencher un engagement émotionnel sans tromperie, ce qui semble peu probable compte tenu de l'étendue de l'engagement émotionnel établi avec des personnages fictifs et des objets clairement non conscients. Si une exigence de tromperie est établie expérimentalement, alors le compromis entre les coûts et les avantages de la tromperie peut être débattu. Le second est cependant incontournable. La paternité que nous avons sur les artefacts est un élément fondamental de leur nature de machine. L'IA est par définition un artefact. Dans une certaine mesure, nous pourrions même affirmer que ce principe est autolimitatif. Si l'IA devait vraiment être capable de modifier ce que signifie être une machine, alors communiquer cette nature de machine modifiée répondrait toujours à ce principe.

Le principe de la responsabilité juridique

La personne ayant la responsabilité légale d'un robot doit être désignée.

Enfin, le cinquième principe communique le statut des robots en tant qu'artefacts de la manière la plus fondamentale possible. Ils sont détenus et cette propriété doit être légalement attribuée. Le fait que les robots soient construits et possédés est la raison pour laquelle j'ai précédemment soutenu que nous sommes éthiquement obligés de ne pas en faire des personnes parce que posséder des personnes est incontestablement contraire à l'éthique. L'argument n'est pas qu'il existe des robots ressemblant à des personnes que nous devrions rétrograder légalement, mais plutôt que le statut juridique nécessairement rétrogradé signifie que nous ne devrions pas faire en sorte que la ressemblance à une personne soit une caractéristique de tout robot fabriqué légalement. Cependant, les principes de la robotique ne vont pas jusqu'à cet extrême du futurisme. Comme je l'ai dit plus tôt, ils se concentrent sur la communication de la réalité actuelle à une population si désireuse de posséder et de s'identifier au surhumain qu'elle pourrait facilement être amenée à croire qu'un robot mal fabriqué ou mal exploité est lui-même responsable des dommages qui lui sont infligés. Si vous entendez un bruit horrible et trouvez une voiture écrasée dans votre maison, vous pouvez rapidement et facilement identifier le propriétaire de la voiture, même si la voiture est actuellement vide, simplement grâce à ses plaques d'immatriculation ou, dans le pire des cas, grâce aux numéros de série. L'idée est que la même chose devrait être vraie si vous trouvez un robot intégré dans votre propriété. Les participants à la retraite sur la robotique ont prédit avec précision un problème déjà présent dans notre société à cause des drones, et qui est maintenant résolu dans certains pays avec des licences obligatoires comme le comité l'a recommandé.

Conclusion

En résumé, les principes de l'EPSRC ont de la valeur parce qu'ils représentent une politique élaborée à un coût important pour les contribuables et les particuliers. Bien qu'aucune politique ne soit parfaite, idéalement, elles

ne devrait être remplacé que par une nouvelle politique avec un niveau d'investissement équivalent ou supérieur à la fois par le gouvernement et les experts du domaine. Leur objectif est de donner confiance aux consommateurs et aux citoyens dans la robotique en tant que technologie fiable apte à devenir omniprésente dans notre société. Les principes individuels représentent chacun des préoccupations substantielles des experts et des parties prenantes, bien que parfois cette représentation elle-même ne soit pas parfaitement transparente. L'objectif global était de communiquer clairement que la responsabilité d'une fabrication et d'un fonctionnement sûrs et fiables des robots n'était pas différente de celle de tout autre objet fabriqué et vendu au Royaume-Uni, et par conséquent, les lois existantes du pays devraient être adéquates pour couvrir à la fois les consommateurs et les fabricants. . Il est important de réaliser que ce n'est pas le cas pour tous les robots imaginables. Il est facile de concevoir des œuvres d'art uniques qui sont considérées comme des robots et non comme des produits banalisés, ou de concevoir des robots qui sont simplement construits de manière dangereuse ou irresponsable. Ce que les gens ont plus de mal à conceptualiser, c'est qu'il peut y avoir des propriétés cognitives telles que la souffrance qu'il serait possible d'intégrer dans un robot, mais cela serait aussi contraire à l'éthique que de mettre des freins défectueux sur une voiture. Les principes de la robotique ne cherchent pas à déterminer ce qui est possible, ils cherchent à communiquer des pratiques conseillées pour intégrer la robotique autonome dans le droit du territoire.

Robotique Recherche Éthique Discussion

A. Gning, D. Davis, Y Cheng, P. Robinson, Département d'informatique,
Université de Hull.

e.gning@hull.ac.uk

Introduction

Dans le monde moderne avec le développement des ressources technologiques, la robotique a donné lieu à de nombreuses applications [1] et à un déploiement souvent imprévu dans la vie réelle (par exemple l'utilisation croissante des drones dans les applications civiles). Pour accompagner cette ère de la robotique, la communauté scientifique et la société en général doivent définir des principes éthiques suffisamment généraux pour être robustes aux évolutions dans le temps et adaptés à l'éventail des applications possibles.

Les principes éthiques doivent être suffisamment vulgarisés et universalisés pour que les concepteurs et fabricants de robots soient conscients des réglementations et des limites à respecter.

De manière générale, les applications de la robotique peuvent être classées en quatre groupes : robots domestiques ou d'assistance humaine [2] [3] [4], robots médicaux [5] [6] [7] [8], robots de défense [9] [10] et robots industriels [11] [12]. La discussion se concentre sur les trois premiers groupes de robots, car les robots industriels sont souvent limités à des zones limitées avec des ensembles prédéfinis de tâches limitées. et ne sont pas en interaction directe avec la société.

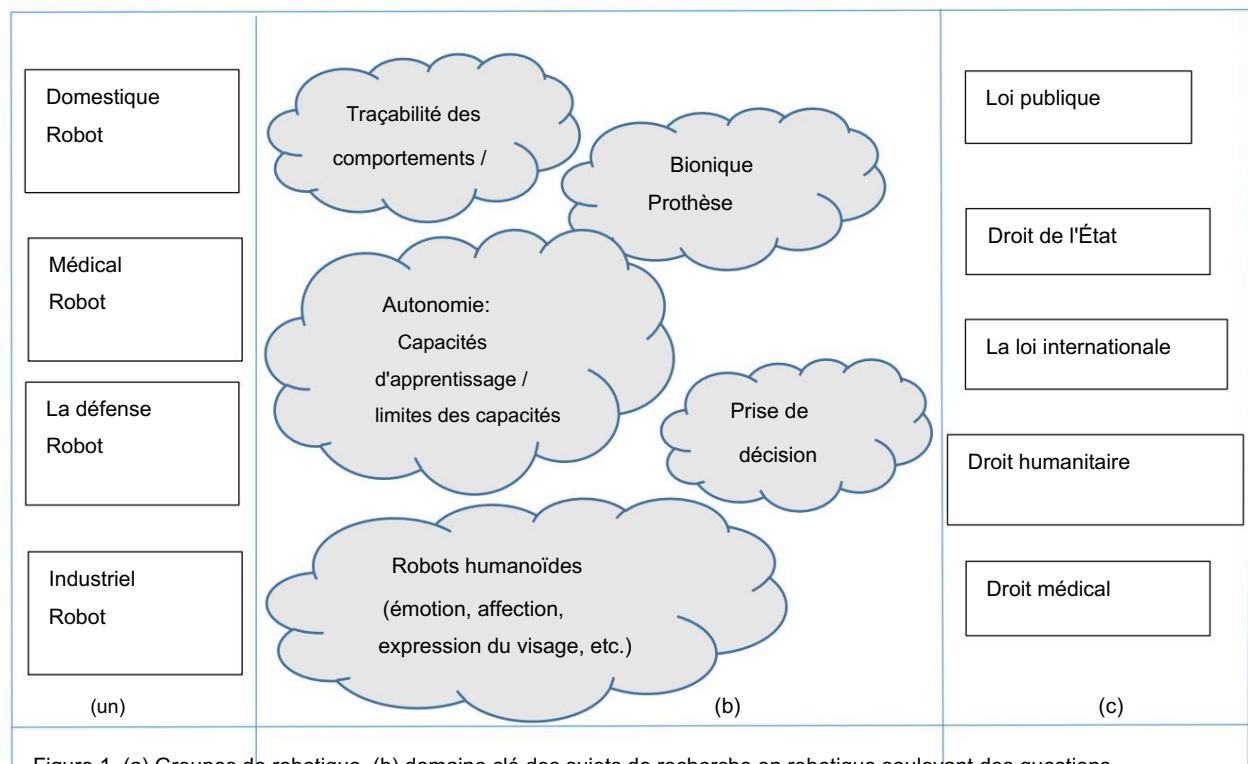


Figure 1. (a) Groupes de robotique, (b) domaine clé des sujets de recherche en robotique soulevant des questions d'éthique, (c) domaine de la réglementation à prendre en compte.

La figure 1 énumère la tâche difficile de réglementer l'éthique pour différents groupes de robotique en ce qui concerne cinq sujets de recherche populaires. Dans chaque groupe de robotique (figure 1-(a)), des questions éthiques peuvent être posées et les principales sont représentées (figure 1-(b)) : comment concevoir des robots toujours

capable d'interpréter leur comportement, de connaître les limites et les limites des robots apprenants ; les robots humanoïdes imitant les gestes humains et l'animation du visage peuvent pousser l'interaction avec l'homme (en particulier les enfants et les personnes handicapées) vers de nouvelles frontières, ce qui nécessite de réglementer et d'anticiper les implications possibles ; De nos jours, les humains peuvent grandement bénéficier de la prothèse grâce aux progrès de la recherche en robotique. Cependant, cela peut amener les personnes non handicapées à rechercher des prothèses susceptibles d'améliorer leurs capacités, ce qui rend à nouveau nécessaire de nouvelles réglementations. Enfin, le plus grand défi concerne les capacités de décision des robots. Ces décisions impliquent directement la vie humaine et soulèvent donc des questions juridiques résultant d'actions exécutées.

En plus de ces sujets de recherche, l'éthique de la robotique doit être compatible avec un ensemble de domaines du droit (figure 1 –(c)). Il faut tenir compte du fait que les robots doivent être compatibles avec différents niveaux de lois qui peuvent changer, par exemple, d'une région à l'autre ou d'un État à l'autre.

Il y a cinq ans au Royaume-Uni, un panel d'éminents experts en robotique et en intelligence artificielle a publié les Principes de robotique de l'EPSRC sous la forme de cinq règles et de sept messages de haut niveau. Nous proposons de discuter de ces règles en mettant l'accent sur la structure transversale - entre la robotique groupes, thèmes de recherche et cadres juridiques présentés dans la figure 1 - et dans le respect des trois critères de validité, de suffisance et d'utilité.

Discussionsurlafixationdesrègles

Le commentaire général qui peut être fait à propos de l'ensemble de cinq règles est qu'il est plutôt ambitieux de penser qu'il est possible de donner des orientations communes/uniformes à tous les types de robots, à travers toutes les évolutions possibles de la recherche et tous les cadres juridiques. Il serait plus naturel de rechercher des règles d'orientation reflétant la nature transversale de la robotique illustrée à la figure 1. Nous pensons que les cinq règles ne sont pas assez générales et les règles devraient être précisées et particularisées pour chaque groupe de robots de la figure 1 (a) en prenant compte de la spécificité des lois et des pistes de recherche impliquées dans la figure 1 (c) fin (b).

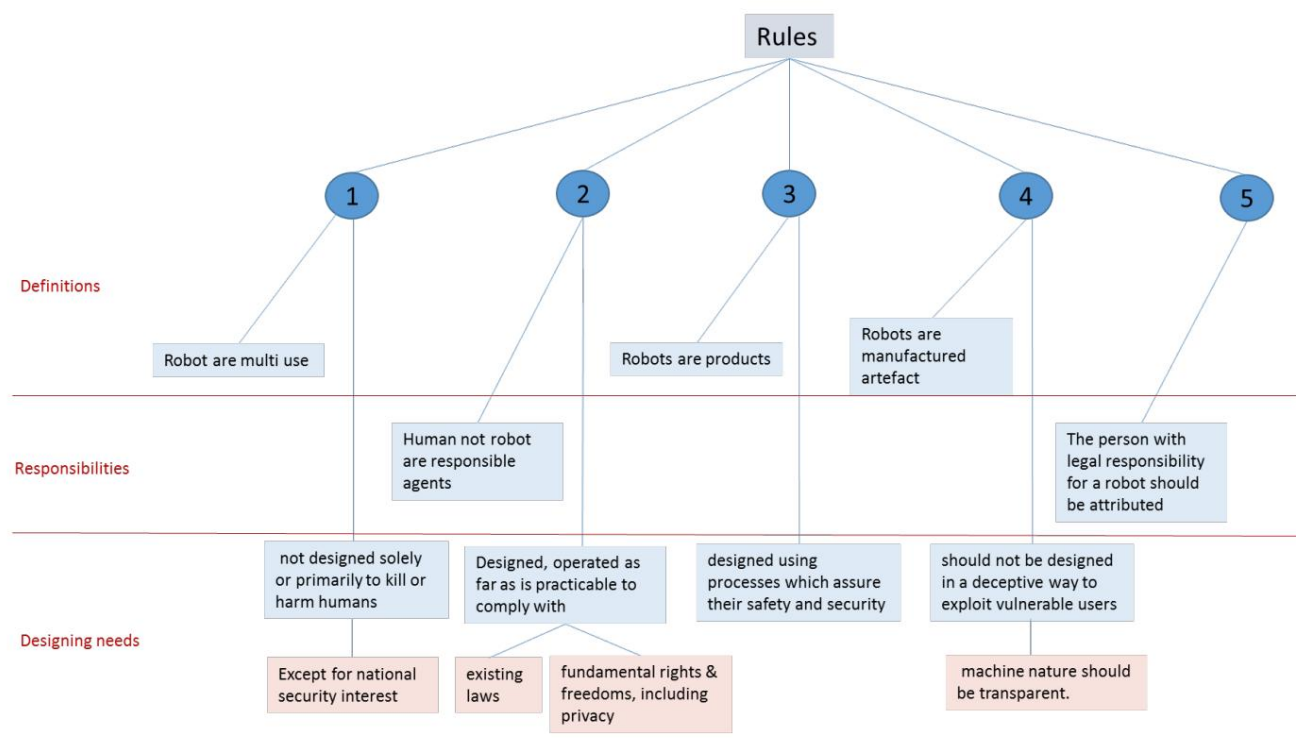


Figure 2. L'ensemble des cinq règles

La figure 2 représente de manière graphique l'ensemble des cinq règles. On peut voir qu'il existe un schéma commun aux cinq règles : les premières phrases sont souvent des généralités et des définitions sur les robots tels que "les robots sont des produits" ou déclarent que les responsabilités ne sont pas imputables au robot mais à l'homme à la place. Les dernières phrases énoncent les besoins de conception et oscillent entre les questions de sécurité, les questions juridiques et les questions de transparence.

Il est clair que les critiques concernant la structure illustrée à la figure 2 peuvent être répertoriées comme suit

- Les cinq règles se chevauchent parfois. Par exemple, « se conformer à la loi existante » dans la règle 2 englobe « non conçu uniquement ou principalement pour tuer ou blesser des êtres humains » dans la règle 1 ; "la personne ayant la responsabilité légale d'un robot devrait être attribuée" dans la règle 5 peut encapsuler implicitement "l'homme et non le robot sont des agents responsables" dans la règle 2.
- Les cinq règles ne sont pas générales. Par exemple, on voit à peine comment la prothèse bionique peut être intégrée dans les règles sous sa forme actuelle (par exemple, la recherche en robotique peut permettre à l'avenir de modifier le corps humain afin de gagner plus de puissance, de vitesse, etc.). La sexualité artificielle est un autre exemple de recherche controversée qui peut entraîner des questions éthiques.
- Les Cinq règles ne précisent pas que la loi peut être contradictoire selon le domaine considéré parmi les communes, les régions, les pays et le continent. Un exemple similaire souvent rencontré dans la recherche est celui des demandes de brevet pour lesquelles plusieurs études spécifiques sont nécessaires pour s'appliquer dans des régions données du monde. Les lois peuvent être encore plus compliquées puisque les croyances religieuses et les us et coutumes des gens dicteront la notion d'éthique.

conclusion

Dans cette discussion, nous avons brièvement fourni des arguments sur la nécessité d'une formulation différente pour les cinq règles. Il est démontré à travers une représentation imagée des cinq règles qu'elles ne sont en fait pas suffisantes, se chevauchent et ne reflètent pas explicitement les véritables enjeux de l'éthique robotique.

Nous avons basé une partie de notre raisonnement sur le caractère transversal de l'éthique robotique selon trois axes : les groupes qui constituent la robotique, les voies futures de la robotique qu'il est essentiel de saisir dans la définition de l'éthique et le caractère structuré du droit. Nous préconisons une reformulation naturelle qui différenciera l'éthique pour chaque groupe de robotique tout en étant inflexible sur les contradictions et fortes contraintes qui peuvent exister du fait de la nature structurelle du droit.

Bibliographie

- [1] G. Bekey, Robotics: State of the art and future challenges., californie: London Imperial 2008.
Presse universitaire. ,
- [2] K. Dautenhahn, S. Woods, C. Kaouri, ML Walters, KL Koay et I. Werry, Qu'est-ce qu'un robot compagnon-ami, assistant ou majordome ?, IIC o. RI a. Systèmes, éd., 2005.
- [3] J. Forlizzi et DC, Robots de service dans l'environnement domestique : une étude de l'aspirateur roomba dans la maison, 1ère conférence ACM SIGCHI/SIGART sur l'interaction homme-robot. AMC, 2006.
- [4] JY Sung, RE Grinter, HI Christensen et L. Guo, Femmes au foyer ou technophiles ? : comprendre les propriétaires de robots domestiques, 3e Conférence internationale ACM/IEEE sur l'interaction homme-robot (HRI), 2008, pp. 129-136.
- [5] GP Moustris, SC Hiridis, K. Deliparaschos et K. Konstantinidis, Evolution of systèmes chirurgicaux robotisés autonomes et semi-autonomes : une revue de la littérature,

- vol. 7, The International Journal of Medical Robotics and Computer Assisted Surgery, 2011, pp. 375-392.
- [6] J. Rassweiler, J. Binder et T. Frede, "Robotic and telesurgery: will they change our futur ? », vol. 11, n° 3, pp. 309-320, 2001.
- [7] G. Kwakkel, KBJ et KHI, Effets de la thérapie assistée par robot sur le membre supérieur récupération après un AVC : une revue systématique, Neuroréhabilitation et réparation neurale, 2007.
- [8] K. Cleary et C. Nguyen, « État de l'art en robotique chirurgicale : applications cliniques et défis technologiques », vol. 6, non. 6, pages 312-328, 2001.
- [9] PW Singer, « Robots en guerre », Wilson Quarterly, 2008.
- [10] TK Adams, « La guerre du futur et le déclin de la prise de décision humaine », Paramètres, vol. 31, non. 4, 2001.
- [11] P. Leitão, "Contrôle de fabrication distribué basé sur des agents : une enquête sur l'état de l'art", Applications d'ingénierie de l'intelligence artificielle, vol. 22, non. 7, pages 979-991, 2009.
- [12] ZM Bi, SY Lang, W. Shen et L. Wang, « Systèmes de fabrication reconfigurables : l'état de l'art », International Journal of Production Research, vol. 46, non. 4, pages 967-992, 2008.

Les robots ne sont pas que des outils

Tony J. Prescott, Université de Sheffield

Au cœur des principes EPSRC de la robotique (désormais « les principes ») sont un nombre d'affirmations ontologiques concernant la nature des robots, également sur ce qu'ils ne sont pas. Les affirmations sur ce qu'ils ne sont pas. "not people" (commentaire du principe 3), et que l'intelligence du robot ne peut donner qu'une « impression d'intelligence réelle » (commentaire du principe 4).

À la première lecture, ces affirmations semblent être des affirmations simples de vérités évidentes. Je soutiendrai que ce n'est pas le cas. en réfléchissant attentivement au statut ontologique des robots.

Si nous regardons la façon dont les principes sont présentés, il semble qu'il s'agisse d'un processus implicite d'induction au travail qui permet d'interpréter les déclarations sur ce que sont la plupart des robots actuels, comme des déclarations sur ce que les robots doivent essentiellement être. Bien qu'il soit facile d'être d'accord avec une affirmation générale selon laquelle les robots sont des outils à usages multiples, en particulier dans le contexte d'une discussion sur la double utilisation (principe 1), l'affirmation beaucoup plus forte que les robots ne sont que des outils, ou simplement des outils, nie qu'ils pourraient raisonnablement appartenir à d'autres catégories disjointes.

Prenons la catégorie « compagnon » par exemple. Il y a un effort majeur autour du développement de robots compagnons qui peuvent fournir un soutien social et émotionnel aux personnes, comme partiellement reconnu dans la discussion du principe 4. Le fait que les robots pourraient appartenir à ces deux catégories soulève des questions importantes et intéressantes qui sont obscurcies en insistant sur le fait que les robots ne sont que des outils.

En effet, conformément à la vision des robots comme des outils, la discussion sur le compagnonnage des robots dans les principes est assez dédaigneuse, décrivant les jouets qui pourraient procurer un certain plaisir aux personnes qui ne peuvent pas ou ne peuvent pas se permettre de garder des animaux de compagnie.

La fausse nature des compagnons robots, selon certains, crée un problème d'éthique dans la mesure où les compagnons robots sont potentiellement trompeurs et devraient donc être conçus de manière à ce que leur « nature machine soit transparente ».

Le problème ontologique concerne en particulier l'affirmation selon laquelle les robots ne pourraient jamais posséder de capacités psychologiques telles que les « vraies » émotions ou l'intelligence.

En effet, il existe des affirmations contraires selon lesquelles les robots, convenablement configurés, peuvent avoir des émotions[1], alors que l'avenir de l'intelligence artificielle, en tant qu'intelligence, n'a pas de plafond évident au-dessous du niveau humain.

Un autre problème concerne l'hypothèse sur la façon dont les gens verront les robots - en particulier, que les robots seront vus comme des outils s'ils sont montrés de manière transparente. par exemple, les gens peuvent anthropomorphiser des robots, même s'il s'agit manifestement de produits manufacturés.

Les animations Heider-Simmel de figures géométriques simples [2] (voir figure), montrent à quel point cette information peut être grossière et pourtant nous verrons toujours l'intentionnalité, la motivation, voire l'émotion. à nous-mêmes.

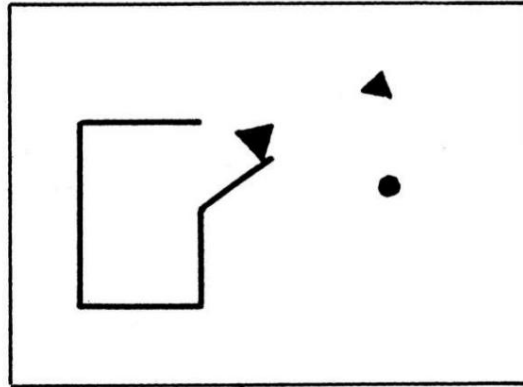


Figure. Les formes géométriques se déplaçant dans une animation simple ont été interprétées comme « des êtres animés, principalement des personnes », dans cette célèbre étude de 1944 menée par Heider et Simmel.

L'analyse des problèmes ontologiques et psychologiques dans l'interaction homme-robot a déjà été faite par Kahn et ses collègues [4]. Suivant un courant de pensée similaire, nous pouvons décrire quatre manières générales de combiner les perspectives ontologiques sur ce que sont les robots et les perspectives psychologiques sur la façon dont les robots sont vus. Celles-ci sont illustrées dans le tableau suivant ainsi que certains des problèmes éthiques qu'ils soulèvent.

I. Les robots ne sont que des outils (o), et les gens verront les robots comme de simples outils, à moins qu'ils ne soient induits en erreur par une conception robotique trompeuse (p). Problèmes éthiques : Nous devons aborder les responsabilités humaines en tant que fabricants/utilisateurs de robots et le risque de tromperie en fabriquant des robots qui semblent être quelque chose qu'ils ne sont pas.	II. Les robots ne sont que des outils (o), mais les gens peuvent les voir dotés de capacités psychologiques importantes, quelle que soit la transparence de leur nature-machine (p). Problèmes éthiques : nous devons prendre en compte la façon dont les gens voient les robots, par exemple, qu'ils peuvent se sentir avoir des relations significatives et précieuses avec les robots, ou qu'ils peuvent voir des robots avoir des états internes importants, comme la capacité de souffrir, bien qu'ils n'aient pas de telles capacités.
III. Les robots peuvent avoir des capacités psychologiques importantes (o) mais les gens n'y verront toujours que des outils (p). Enjeux éthiques : Nous devrions analyser le risque lié au traitement d'entités qui peuvent avoir des capacités psychologiques importantes, comme la capacité de souffrir, comme s'il ne s'agissait que d'outils, et les dangers inhérents à la création d'une nouvelle classe d'entités dotées de capacités psychologiques importantes, comme l'intelligence humaine, sans reconnaître qu'elles le faisaient.	IV. Les robots peuvent avoir des capacités psychologiques similaires à celles des humains (o), et les gens verront eux-mêmes de telles capacités (p). Problèmes éthiques : nous devrions envisager des scénarios dans lesquels les personnes devront coexister aux côtés de nouveaux types d'entités psychologiquement significatives sous la forme de futurs robots/IA.

Tableau. Comment les perspectives ontologique(o) et psychologique(p) sur les robots peuvent se combiner (d'après Kahn et al., 2007).

Notez qu'un seul quadrant de ce tableau (I) est traité dans les principes, mais que II, III et IV sont tous possibles, du moins en théorie.

Dans le quadrant II, des questions intéressantes se posent sur la façon dont les robots doivent être traités, non pas parce qu'ils sont des agents sensibles, mais parce que les gens choisiront de les traiter tels quels. Les robots-chiens Sony Aibo [6], ne semblent pas si étranges lorsqu'ils sont vus du point de vue de la façon dont les robots sont vus par les gens plutôt qu'en termes de ce qu'ils sont. nous nous développons avec d'autres biens de valeur, Par contre, pour certains robots, ils peuvent ressembler davantage aux relations que nous entretenons avec des animaux de compagnie, y compris, par exemple, en souhaitant les soutenir et les nourrir (ce que nous pouvons nous-mêmes trouver gratifiant). Enfin, certaines relations homme-robot peuvent partager des similitudes avec les relations homme-homme. plus généralement, ce qui peut être nécessaire, pour développer des principes éthiques adaptés, est de développer une taxonomie des différentes formes de liens affectifs qui pourraient exister entre les robots et les personnes et d'analyser les facteurs qui pourraient sous-tendre le développement et le maintien de telles relations[7].

Quadrant III Concernent la possibilité de dispositifs de disposition destinés à un psychologique des capacités qui sont dans la gamme de ce qui est aimé par les gens. Enclusing suffering sans trahir de manière insensible à la manière dont les duet à percer, ce qui est possible, il est possible que ce soit un nombre de troubles. OTS est devenu adémorétique. 9] affirme un rôle critique pour l'intégration de l'information

qui ne nécessite pas nécessairement un substrat biologique.

des cerveaux plus petits que le nôtre, comme les poissons, peuvent être sensibles de manière significative (par exemple, ils peuvent éprouver de la douleur) [11]. Ces développements suggèrent que la conscience pourrait être possible dans un agent artificiel sans avoir à correspondre à la taille ou à la complexité d'un cerveau humain intact.

[12], et Bryson [13] ont proposé que les robots d'aujourd'hui pourraient déjà avoir des formes simples de conscience qui répondent à certains critères couramment proposés.

L'une des conséquences de la vision des robots comme de « simples outils » est le rejet implicite de la possibilité d'une IA forte, à savoir que les futurs robots pourraient avoir une intelligence générale de niveau humain, ou au-delà du niveau humain. La ularité pourrait inverser la relation maître-esclave entre les humains et les androbots. une super-IA auto-amorçante. Une approche éthique inciterait sûrement à plus de vigilance.

Le débat porte sur la perspective du « cerveau global », proposée par Heylighen[15] et d'autres, selon laquelle les humains et l'IA avancée pourraient coexister pour notre bénéfice mutuel.

Bien que les scénarios des quadrants III/IV puissent sembler farfelus ou du moins lointains, de telles préoccupations ont captivé l'imagination du public et suscité d'importants appels au débat (par exemple [16]). Il se peut qu'il soit peu aisé d'apaiser les voix et qu'il puisse apparaître comme hégémonique et condescendant.

Une approche plus candide peut être de reconnaître que, alors que la plupart des robots ne sont actuellement guère plus que des outils, nous entrons dans une ère où il y aura de nouveaux types d'entités qui combinent certaines des propriétés des machines et des outils avec des capacités psychologiques que nous pensions auparavant réservées aux organismes biologiques complexes tels que les humains. elle vit de la même manière que les organismes biologiques, ni simplement mécanique comme avec une machine traditionnelle.

La liminalité des robots les rend à la fois fascinants et effrayants par nature, et un paratonnerre pour nos craintes plus larges concernant les effets déshumanisants de la technologie[18].

L'Association of Manhattan Scientists a écrit en 1945 [19] sur leur sentiment de responsabilité collective pour leur rôle dans le développement d'une technologie avec un « potentiel de grande nuisance ou de grand bien » (l'énergie atomique) et de leur « conscience particulière » qu'elle pourrait conduire à « l'avancée de quatre civilisations ou à sa destruction totale ». la responsabilité de comprendre et de parler franchement de ce que l'avenir de la robotique pourrait apporter et de ses avantages et menaces potentiels.

Les références

1. Fellous, J.-M., From human emotion to robot emotions, in AAAI Spring Symposium on Architectures for modeling emotions: Cross-disciplinary foundations E. Hudlicka and L. Cañamero, Editors. 2004, AAAI Press: Menlo Park, CA. p. 37-47.
2. Heider, F. et M. Simmel, Une étude expérimentale du comportement apparent. *The American Journal of Psychology*, 1944. 57(2): p. 243-259.
3. Levy, D., *Love and Sex with Robots*. 2007, Londres : HarperCollins.
4. Kahn, J., Peter H., et al., What is a Human?: Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies*, 2007. 8(3): p. 363-390.
5. Lovgren, S., Robot Code of Ethics to Prevent Android Abuse, Protect Humans, dans *National Geographic News*. 2007.
6. Brown, A., *Tomorrow's robot dog is truly human* in *Guardian*. 2015 : Manchester.
7. Collins, EC, A. Millings et P. T. J. Attachment to Assistive Technology: A New Conceptualisation. en *technologie d'assistance : de la recherche à la pratique : AAATE 2013*. 2013.
8. Metzinger, T., *Le tunnel de l'ego : la science de l'esprit et le mythe du soi*. 2009, New York : Basic Books.
9. Tononi, G., *Consciousness as Integrated Information: a Provisional Manifesto*. *The Biological Bulletin*, 2008. 215(3): p. 216-242.
- dix. Qiu, J., Sonder les îlots de conscience dans le cerveau endommagé. *The Lancet Neurology*. 6(11): p. 946-947.
11. Seth, AK, Why fish pain cannot and should not be ruled out *Animal Sentience*, 2016. 2016. 020.
12. Dennett, D., *Les exigences pratiques pour fabriquer un robot conscient*. *Philosophical Transactions of the Royal Society of London A*, 1994. 349: p. 133-146.

13. Bryson, JJ Crude, Cheesy, Second-RateConsciousness. à VienneConferenceonConsciousness. 2008.
14. Bostrom, N., Superintelligence : Chemins, Dangers, Stratégies. 2014, Oxford : Oxford University Press.
15. Heylighen,F., TheGlobalBrainasaNewUtopia, in Zukunftsfiguren, R.MareschandF.Rötzer,Editors.2002,Suhrkamp:Frankfurt.
16. Institut FutureofLife. UneLettreOuverte :PrioritésdeRecherchepourUneIntelligenceArtificielleRobusteeetBénéfique. 2015 ; Disponible sur : <http://futureoflife.org/ai-open-letter/>.
17. Kang,M., SublimeDreamsofLivingMachines:TheAutomatonintheEuropeanImagination. 2011, Cambridge, MA : Harvard University Press.
18. Szollosy, M., Freud, Frankenstein et notre peur des robots : projection dans notre perception culturelle de la technologie. IA&SOCIÉTÉ, 2016 : p.1-7.
19. Association des scientifiques de Manhattan. Déclaration préliminaire. 1945 ; Disponible sur : <https://www.gilderlehrman.org/history-by-era/postwar-politics-and-origins-cold-war/resources/physicists-predict-nuclear-arms-race->

Atelier AISB sur les principes EPSRC de la robotique

Défendre un human(isme) obsolète ?

Michael Szollosy, Sheffield Robotique

Introduction

L'atelier EPSRC 2010 visant à définir un ensemble de principes pour le développement responsable de la recherche sur les robots et les systèmes robotiques.

Il ressort clairement de l'examen des principes EPSRC qu'ils peuvent et doivent remplir une fonction vitale dans protégeant les êtres humains des recherches irresponsables ou simplement irréfléchies sur les technologies qui pourraient avoir des conséquences très réelles et très négatives pour l'humanité, à un niveau personnel, sociétal ou même à l'échelle de l'espèce.

Cependant, il est également clair que ce qui est protégé par les Principes de l'EPSRC est tout être humain spécifique, ou du moins, toute conception spécifique de ce qui constitue un être humain. .

(Cependant, je suis d'avis qu'il aurait dû y avoir un préambule, énonçant ces hypothèses.)
document enraciné dans un contexte historique et culturel particulier, ce qui rend peu probable que ces Principes, dans leur forme actuelle, perdurent à moyen et à long terme.

Les Principes de l'EPSRC formulent certaines hypothèses, très spécifiques, mais complètement tacites, sur ce qui constitue un « être humain ». Et les Principes conviennent donc très bien à l'être humain. pour concevoir, construire et acheter des robots, et doit en conserver l'entière responsabilité légale.

Les principes tels qu'ils sont actuellement articulés seront probablement suffisants à court terme, peut-être même à moyen terme, pour traiter la plupart des problèmes liés aux nouvelles technologies qui émergent de la robotique et des laboratoires informatiques à travers le Royaume-Uni.

créature transitoire, relativement nouvelle invention, étant, en

outre, qu'il n'est, ni actuellement, qu'il n'a jamais été cohérent en interne et unitaire, et sera constamment refaçoné et transformé

mais aussi –

plus simplement, de nouveau autant sur soi, la société et de nouvelles façons de penser notre place dans le monde, des changements qui peuvent être humains – ou les êtres humains – dont il est question dans les Principes de l'ESPRC ne sera pas la première, ni la dernière, formulation de ce que signifie être humain.

Quel humain ?

Au cœur (implicite) des principes de l'ESPRC se trouve un être humain particulier défini au cours des derniers siècles par ce qui est devenu l'humanisme. Cet être humain est un agent à part entière, un être indépendant qui ne doit pas être gouverné par d'autres forces, métaphysiques ou surnaturelles.

Cet être humain est au centre des systèmes juridiques, éthiques, économiques et politiques européens; cependant, il est essentiel des

Il y a peu de consensus sur la naissance de l'être humain – certains revendiqueraient la Renaissance, d'autres diraient les Lumières, et d'autres encore diraient que le sujet humaniste n'est devenu central dans la façon dont nous pensons à nous-mêmes qu'au XIXe siècle. En se concentrant sur nous-mêmes, il est certain que cette conceptualisation de l'être humain est une invention, non donnée ; l'humain humaniste n'est pas " naturel ", ni même une " interprétation " correcte " de notre nature humaine.

Et lorsque nous contextualisons l'humanisme de cette manière, et le considérons comme historiquement, en notant à quel point les différentes idées de ce que signifie être humain (ou même " humaniste ") ont radicalement changé au cours des siècles, il devient évident que nous ne parlons pas seulement d'un être humain, ou d'une idée de ce que cela signifie d'être humain. parler de n'est pas un être humain unique, mais de nombreux êtres humains, pas une nature humaine unique, inaliénable et évidente, mais des humains changeant de conception d'eux-mêmes dans des contextes particuliers . Chaque personne peut être un membre du public, un citoyen, un spécialiste de différents contextes, un consommateur, un producteur ; nous pouvons être des criminels, des patients, des clients, des contribuables, des parties prenantes, des étudiants, des ouvriers ou des dirigeants ou toutes ces choses à la fois, ou aucune d'entre elles, selon les contextes. 'ora' La femme est très différente aujourd'hui de ce qu'elle était il y a deux cents, ou même cinquante, voire dix ans.

Les hypothèses qui sous-tendent notre conception d'une conception solitaire de ce que signifie être « humain » ne sont pas viables sous l'examen approfondi de nouvelles façons de penser à soi, et aussi comme les nouvelles technologies nous obligent à nous penser différemment.

Les bâtons pour aider à la chasse à la navette volante ont transformé la façon dont le tissu a été fabriqué lors de la révolution industrielle.

plus étroitement intégrés à leurs outils car les bâtons deviennent des prothèses et le travail humain est complètement remplacé par des machines automatisées. Et ces développements créeront de nouveaux êtres humains et de nouvelles façons de penser à nous-mêmes.

Cependant, la technologie ne se manifeste pas toujours dans des entités physiques ; les avancées technologiques ne prennent pas toujours la forme de nouveaux outils ou de nouvelles machines : l'invention des lois et du système juridique étaient de nouvelles technologies qui ont eu un impact considérable sur la façon dont nous construisons notre moi humain et social, de la même manière que l'invention de la méthode scientifique, les nouvelles relations industrielles ou Facebook ont changé la façon dont nous nous concevons et présentons cette idée de nous-mêmes. le monde. Nos technologies du 21^{esiècle}—

y compris les robots et l'IA plus avancés, mais aussi les systèmes juridiques, les organismes politiques et les systèmes pour la vie éthique—seront des développements qui transformeront radicalement, au fil du temps, la façon dont nous nous voyons et concevons l'idée même de ce que signifie être humain.

De quelles manières implicites et tacites nous avons besoin d'examiner spécifiquement, le principe d'ESPRC se laisse-t-il aller à ces hypothèses humanistes sur la nature des êtres humains ? s—constitutions, chartes, traités ou déclarations de principes—devraient-elles noncer explicitement dès le départ ces vérités qui sont tenues comme évidentes, la base de ce qui suit.

En l'absence d'un sujet clairement défini, les Principes offrent la conception humaniste habituelle et familière de l'humain - l'être humain statique et homogène qui sera très bientôt créé obsolète, s'il n'est pas encore prêt, par les avancées technologiques qu'il cherche à maîtriser. Il existe une conception trop simpliste de la relation entre les êtres humains et leurs outils : une relation à sens unique où les outils sont toujours les serviteurs de leurs maîtres humains, et toujours sous le contrôle d'un agent humain indépendant. Une telle relation entre sujet et objet, agent actif et article passif aurait toujours été naïve. nos nouvelles technologies exigent une réorganisation fondamentale de tout notre mode de vie, et ont insisté sur la façon dont nous concevons toute notre structure sociale.

Dire que notre relation avec nos outils n'est pas une simple relation maître-serviteur, une relation à sens unique ne signifie pas que nos outils sont nos maîtres. (Il n'y a vraiment rien de controversé à dire cela ; c'est quelque chose que Marx a reconnu il y a plus de cent cinquante ans, en expliquant comment le mode de production d'une société

Les relations entre les êtres humains et leurs outils ont toujours été plus complexes qu'on ne l'imaginait par l'humanisme et dans ces principes. et les frontières entre « biologique » et « machine » sont encore plus floues.

Les Principes, par conséquent, malgré de nobles intentions, tentent de donner la maîtrise des technologies futures et évolutives à un être humain obsolète.

La conception de l'être humain proposée dans les Principes partage également avec l'humanisme l'illusion d'offrir un seul sujet homogène, alors qu'en fait ce sujet est une compilation d'êtres multiples, souvent contradictoires. e, voire des cultures différentes au sein d'une même communauté.

Il est peu probable que les progrès de la robotique et de l'IA profitent également à toutes les communautés et à toutes les nations, en particulier à court et à moyen terme, et les principes de développement de la robotique devraient en tenir compte.

Il n'est pas clair non plus à quels individus il est fait référence dans le document ;

- Les Principes font référence aux « citoyens », un sujet d'un certain corps politique (généralement national), bien qu'il ne soit pas clair si quelqu'un peut plus prétendre être un « citoyen » d'un État-nation discret et indépendant, indépendamment d'autres influences. Les avantages et la responsabilité des robots ne seront-ils que pour les citoyens d'un État-nation ou d'un organe politique en particulier ? raisons de sécurité ».
- Les Principes sont inflexibles sur le fait que seuls les êtres humains sont des " agents responsables " dans la loi. De plus, nous pouvons nous demander comment les humains technologiquement améliorés (par exemple, les cyborgs) peuvent être considérés dans la loi comme des agents tout aussi (moins ? plus ?) responsables.
- Les Principes insistent sur des considérations de confidentialité, même si nous pouvons déjà voir que pour beaucoup de gens les frontières entre « soi » et « public » sont floues, et la notion de vie privée a Les médias sociaux, la promesse des « maisons intelligentes » et les problèmes de sécurité ont signifié que, culturellement, nous avons une idée très différente de ce que signifie la « vie privée » et de son importance pour nous.
- Les Principes ont établi une distinction claire entre ceux qui " conçoivent " des robots, ceux qui " vendent " des robots et ces " consommateurs " et " utilisateurs ". Les Principes acceptent implicitement que les intérêts de ces groupes puissent se concurrencer. , ex. Paul Mason (2015). .Commenousestomponsdéjàlesfrontièresentreproducteursetvendeursd'une partetconsommateursetutilisateursd'autre part

d'autres (par exemple, Uber, les données de crowdsourcing, Google), ces catégories doivent déjà être beaucoup plus flexibles qu'elles ne l'imaginent dans un humanisme direct et simpliste.

Il convient également de souligner que les principes ESPRCP de la robotique sont, sans surprise, peut-être, très européens et chrétiens. Les robots sont considérés comme des machines, et donc comme de simples objets. Cette propriété immatérielle, métaphysique, unique à la vie ou, au plus profond de l'articulation, unique spécifiquement à l'homme. (Cette idée de faire fi de quelque chose de vitalement humain se trouve dans l'idée même du robot, lorsque le mot a été introduit pour la première fois dans le monde dans la pièce de théâtre de Karl Capek en 1921, RUR). Ces principes ESPRC.

Par contraste, il convient de noter – comme beaucoup l'ont (par exemple, Metzler et Lewis 2008 ; Lee, Sung, Šabanović, Han 2012) – comment les robots et l'IA sont perçus différemment dans différents contextes culturels. Il s'agit de religions « animistes », c'est-à-dire qu'ils croient que tout, y compris les objets inanimés, contient la nature du kami ou de l'esprit. De telles influences, si profondément ancrées, ne sont pas susceptibles d'être transformées trop facilement par l'introduction de nouvelles technologies et idées, mais elles soulignent que les principes ESPRC sont très étroitement liés à un contexte culturel et historique très spécifique, et que nous devons être prêts et désireux d'imaginer d'autres idées et relations, non seulement à l'avenir, mais dès maintenant, si nous essayons d'établir un consensus international sur les principes de la robotique.

Des nouveaux humains ?

Le fait que les Principes " ne soient pas destinés à être des lois dures et rapides, mais plutôt à éclairer le débat et à servir de référence future " démontre la réflexion prospective des délégués à l'atelier, mais les Principes devraient être terminés pour permettre d'aller au-delà de la notion étroite d'" humain " qui les sous-tend dans leur état actuel. -limitation rapide de ce qui peut être conçu, car cette idée « d'humain » définit toutes les relations qui y sont imaginées.

Les penseurs peuvent marchander (et souvent, jusqu'à la nausée) à propos du moment où le consensus soutenant le sujet humaniste s'est effondré, mais il est clair qu'à un moment donné après la Seconde Guerre mondiale, avec la perte de la foi dans les métarécits et les nouvelles herméneutiques radicales de la suspicion (que certains en sont venus à comprendre comme du « postmodernisme »), le sujet humaniste stable tel qu'il était autrefois compris n'était pas long pour ce monde. Il est clair que quelque chose doit venir ensuite, car nous ne pouvons pas procéder à la construction de cadres ou de modèles sans avoir une idée de ce que signifie être humain - il y a beaucoup moins d'accord. humains

en tant que consommateurs, humains en tant que producteurs, humains en tant que concepteurs, en tant que sujets juridiques, en tant que citoyens et sujets de diverses entités politiques... N'importe lequel de ces êtres humains peut battre n'importe quel moment.

On peut chercher à recréer des principes de robotique basés sur un sujet humain qui vient après l'humanisme. On peut vouloir appeler cet être humain le post-humain, ou simplement post-humain. Cependant, ces termes sont compliqués et renvoient à un éventail vertigineux d'idées et d'idéologies différentes (encore plus que ce qui était contenu sous le terme générique « humanisme » qui l'a précédé). pour l'innovation technologique.

Le posthumanisme peut simplement signifier, philosophiquement, culturellement, ce qui vient après l'humanisme ; ce posthumanisme, parfois un anti-humanisme, réfute le genre d'hypothèses stables et singulières sur la nature humaine et humaine qui sont énoncées par l'humanisme. Pour aller un peu plus loin, le posthumanisme accepte la contingence et les contextes des conceptions de l'humain, et remplace une nature humaine statique par quelque chose de plus dynamique et pluraliste.

Le posthumanisme, ou peut-être plus précisément, les posthumanismes, ne sont pas téléologiques ; ils ne partent pas du fait que l'être humain auquel nous sommes arrivés après des millions d'années d'évolution et des milliers d'années de philosophie - nous - est l' être humain, un produit final, fini, poli qui restera désormais à jamais immuable et immuable. La grande force du posthumanisme, telle qu'elle est comprise et articulée ici, est qu'il existe une flexibilité intrinsèque pour s'adapter à de tels changements, et il est important que toute entreprise ambitieuse, telle que l'élaboration d'un ensemble de principes pour la définition de nos relations présentes et futures avec une technologie en constante évolution, a une flexibilité intrinsèque similaire.

Certains, particulièrement optimistes quant à la quasi-émergence de l'IA sensible, et qui pourraient se qualifier de transhumanistes, pourraient considérer les principes ESPRC comme naïvement anthropocentriques, qu'ils ne tiennent pas compte de l'émergence en tant que robots et agents essentiels de l'IA qui méritent une considération (peut-être égale) aux côtés des humains dans la création de tout principe éthique. avançant une croyance que les principes ESPRC sont vraiment quelque peu dépassés et trop étroits dans leur conception de ce qui constitue « l'humain », bien que je sois beaucoup moins optimiste quant à la minence de l'IA sensible, et je ne partage pas la certitude transhumaniste générale que les humains sont radicalement transformés par la technologie (par exemple, les êtres humains sont presque immortels) sont également très proches. Même si nous n'inventons pas de nouveaux robots et ne faisons aucun progrès dans l'intelligence artificielle - ce qui est très peu probable - c'est presque certain que nous, les êtres humains, continuerons à réinventer les autres systèmes et institutions qui définissent qui nous sommes, transformant ainsi les êtres humains et nécessitant un nouvel ensemble de principes plus flexibles pour définir notre relation avec les robots et l'IA.

Les rédacteurs des principes entendent en faire un « document vivant », non pas des lois « absolues », mais la base d'un débat et d'une référence futurs, ce qui est exactement ce dont il a besoin.

nos nouvelles technologies lorsqu'elles prennent comme lieu de départ un sujet humain rigide et déjà obsolète en son cœur.

Il est intéressant de voir que le préambule des Principes mentionne l'omniprésence d'Asimov et de ses Trois Lois. mondes différents, peuplés de différentes sortes d'êtres humains. Ces relations changeantes, et à la fois permettent et imposent des limites à diverses directions de développement. .Nous devons réfléchir avec imagination aux types de robots que nous créerons , mais aussi aux types de personnes que nous deviendrons, et si nous cherchons à élaborer des principes pour le bénéfice de quatre sociétés, nous devons mieux comprendre à quoi ressembleront ces sociétés et les êtres humains qui les peuplent.

Les références

Lee, Sung, Šabanović, Han. 2012. Cultural design of domestic robots : une étude des attentes des utilisateurs dans Corée et États-Unis.

Mason P. 2015. Postcapitalisme : un guide pour notre avenir. Londres : Allen Lane.

Metzler and Lewis. 2008. Points de vue éthiques, points de vue religieux et acceptation des applications robotiques : un projet pilote étude. AAAI. 15–22.

Réguler! Robot! Villes : !! Réflexions! sur! les! principes! de! la! robotique du nouveau côté lointain du droit

Aurore! Voiculescu

Centre! de! droit! & & théorie, Université! de! Westminster

« Ils! m'ont! demandé! où! je! choisirais! de! courir, & qu'est-ce que! favoris ?

Où! robot! & souris! & et! hommes, & j'ai! dit, & courir! & circuler! & dans! robot! & villes.

Mais! est-ce! sage! & ? & Pour! & l'étain! & un! imbécile! & et! le! fer! & n'a! pas! pensé !

Ordinateur! & souris! & peut! & trouver! & moi! & faits! & et! enseigne! & moi! & ce que! je ne suis pas.

Mais! robot! & tout! & inhumain! & est, & tout! & péché! & avec! cog! & et! mesh.

Pas! & si! & nous! & enseignons! & les! & bonnes! & choses! & dedans, & ainsi! & cela! & peut! & enseigner! & notre! & chair! &

[...]

Comme! & l'homme! & lui-

même! & un! & mélange! & est, & exubérant! & paradoxe, & donc! & nous! & devons! & enseigner! & à! & nos! & machines! & folles! & : & de! & se! & lever! &

Venez! & courez! & avec! & moi, & sauvages! & enfants! & hommes, & moitié! & d'ires! & et! & malédictions, & moitié! & de! & clowns.

Pace! & robot! & souris, & race! & robot! & men, & win! & close! & in! & robot! & towns.

Roy! Bradbury¹

Les! principes! de! l'initiative! de! la! robotique! découlent! d'une! réflexion! de! la! mesure! à
quel point les robots! affectent déjà! notre! vie! et! dans! la! plus! grande! mesure! dans! laquelle! ils! est!
prévu ça ! ils! volonté! affecter! il! dans! le! 'robot! les villes!' de! le! relativement! près! avenir.
Que! l'initiale! régulation! en rapport! à! ce! transformateur ! technologie! prendra! la! forme! de!
principes! directeurs! doux, d'instruments! juridiques! internes! durs, ! ou! même! de! complexe!
international! traités ! est! un! difficile! encore, ! à! ce! indiquer, ! un! secondaire! problème. ! Le!
primaire! question! est plutôt! un! (légal! E)! normatif! question, ! visé ! à! délimitant! clair! frontières!
de! le! humain/robot ! coexistence ; ! adressage ! le! normatif! dynamique! de! causalité! et!
responsabilité; ! en essayant! à! identifier! le! lieu ou! lieux de! mensa et! actu dedans ! processus !
et, ! oser! nous! dire, ! relations qui! pourraient! bien! prouver! à! devenez! plus! et! plus! complexe!
avec! le! progrès! de! la! science! et! de! la! technologie.²

Enracinement ! depuis! ce! besoin! pour! normatif! introspection! (dans! notre! psyché! sociale!
bien! plus! que! quoi que ce soit! d'autre), ! ce! papier! est! une invitation! à la réflexion ! sur! le!

¹ Rayon! Bradbury, ! Where! & Robot! & Mice! & and! & Robot! & Men! & Run! & Round! & in! & Robot! &
Towns: & New! & Poems, & Both! & Light! & and! & Dark! (New! York: ! Random! House! Inc, ! 1977). !

² ! Aurore ! Voiculescu, ! "Humain! Droits! Au-delà! le! Humain: ! Herméneutiques! et!
Normativité! à! l'âge! de! l'inconnu, » ! (à paraître). !

proposé ! Des principes de Robotique ! (couvrant ! 5 principes ! et ! 7 ! Niveau élevé ! Messages) ! à venir ! dehors ! de ! le ! pluridisciplinaire ! expert Informé ! EPSRC ! et ! AHRC ! Robotique ! Retraite ! dans ! 2010 ! Le ! complexité ! de ! problèmes ! la ! la couverture est ! tel ! que ! ces ! reflets ! peut ! seulement ! but ! à ! s'engager ! avec ! quoi ! est ! proposé !, avec le ! texte ! offert ! pour ! réflexion !, pris ! dehors ! quelques ! de ! le ! possible ! significations ! ou ! interprétations ! de ! tels ! textes !. Cette analyse est ! proposée comme ! essentielle ! pour ! préparer ! terrain pour ! plus loin ! discussions !, et ! enfin !, pour ! embarquement ! sur ! n'importe lequel ! éventuel ! processus ! réglementaires !.

Principes dans la recherche d'une définition !

Réfléchir sur les principes existants ! invite d'abord à réfléchir à ce qu'est un robot et à la définition que celui-ci ! devrait se contenter de, 3 et donc ! le type d'entités que l'on devrait viser à réguler, devrait être une réponse à ! notre ! état de la Terre ! en ! technologie ! ou ! un ! réflexion ! de ! notre ! state of the E (technologie) ! in ! art !. Autrement dit, sur quel point du spectre entre la science et la science ! fiction ! devrait nous ! lieu ! nous-mêmes ! quand ! conception ! normes ! et ! évaluer ! leur ! efficacité ? Jusqu'où dans l'avenir devrait-on regarder, quand l'avenir pour ! qui ! nous ! réglons ! c'est ainsi ! loin ! que nous ! pouvons ! que ! spéculer ! comme ! à son existence, tandis que ! en même temps nous galopons vers ce futur très futur à une vitesse toujours plus grande ? !

Loi / règlement, ! dur ! ou ! doux, ! nécessite des définitions. Le ! des principes ! ici ! sous ! discussion ! faire ! pas ! immédiatement ! envoyer ! à ! un !. UN ! robot ! est ! défini ! par ! quelques ! comme ! 'un ! machine ! capable ! de ! portant ! dehors ! un ! complexe ! série ! de ! Actions ! automatiquement' ! ou, ! différemment ! nuancé, ! 'un ! mécanique ! ou ! agent ! virtuel ! artificiel, ! généralement ! une ! électro-mécanique ! machine ! que ! est ! guidé ! par ! un ! ordinateur ! programme ! ou ! électronique ! circuit' !. Divers, ! plus ! ou ! moins ! réalisable ! distinctions ! sont ! aussi ! mettre ! avant, ! plus ! notamment ! entre ! industriel ! et ! service ! Machines, ! entre ! très ! machines ! autonomes ! et ! programmes ! informatiques ! cognitifs, ! entre ! incarnés ! et ! des entités ! cognitives ! désincarnées, ! etc. ! la ! NASA ! elle-même ! utilise ! un ! plutôt ! banal ! et ! imprécis ! langue, ! très ! peu serviable ! pour ! le ! régulateur, ! définir ! des robots ! comme ! "des machines qui peuvent être utilisées pour faire des travaux". Certains robots, la ! formulation ! de ! la ! NASA ! continue ! à ! ajouter, ! 'peut ! faire ! travail ! par ! eux-mêmes. ! Autre ! des robots ! doit ! toujours ! avoir ! un ! personne !

3 ! A ! définition ! jamais ! être ! valeur ! neutre, ! toujours ! établissement ! le ! 'in' Es ! et ! le ! 'out' Es ! suivant ! un ! plus ! ou ! moins ! déclaré ! value ! laden ! chemin. ! (Voyez ! Alan ! Norrie ! dans !

Voiculescu 2000)

4 Merriam E Webster ! Dictionnaire, ! "Définition ! de ! "Robot", accédé ! Février ! dix, ! 2016, ! <http://www.merriam-webster.com/dictionary/robot?entry=robot>.

récit! eux! quoi! à faire'.⁵ Tel! un! variété! de! formulations! créer! un! réglementaire! casse-tête! et! rendre! toute!
énoncé! normatif! difficile! à suivre! et! ou! facile! à échapper! à se conformer.!!

Alors que! d'accord! que! là! est! Non! convenu! définition! en soi, ⁶ quelques-uns! mettre! de suite! un!
nombre! de! fonctionnalités! qu'un! robot! aurait, des! fonctionnalités! qui, d'un! organisme! de! réglementation! (et! pas!
uniquement)! perspective, sont! eux-mêmes! dans! besoin! de! définitions! : ! sentant le! alentours! (avoir! une!
"conscience"! intégrée! de! son! environnement);! mouvement, si! rouler, marcher, poussée, ou! peut-être!
même! juste! transmettre des données ;! énergie, être capable! de! se! nourrir! de! façons! cela! dépendra! de! quoi!
le but de! le! robot! est; l'intelligence: étant! fourni! de! "smarts"! par! son! programmeur, ayant! la! capacité! à! évaluer!
alentours, circonstances, complexe! information. ! [(Et! plus! une! machine! est! capable! d'interagir!
indépendamment! avec! un! monde! dynamique, plus! plus! cette! machine/robot! est!,!!) A! vise, l'autre! autres,!
précisément! vers! ça)]!

Donc, un! robot! est! défini! plus! spécifiquement! comme! un! système, un! machine! cela contient! capteurs,!
systèmes! de! contrôle, manipulateurs, alimentations! et! logiciels! tous! fonctionnant! ensemble! à! effectuer! un!
tâche".! Selon! à! tel! un! perspective, "[conception, bâtiment, la programmation! et! essai! un! robot! est! un!
combinaison! de! la physique, mécanique! ingénierie, électrique! ingénierie, de construction! ingénierie,
mathématiques! et! l'informatique. Dans! quelques! cas! biologie, médecine, chimie! pourrait! aussi! être!
impliqué ».! Si! le! étudiant! dans! robotique! peut! activement! se! lancer avec! tous! ces! discipline! "dans! un!
profondément! problème! Eposant! problème! Résoudre! environnement », ⁷ certains! pourrait! justement! dire,!
que! régulateur! des robots! et! 'robot! les villes' a besoin! un! de la même manière! engagement!
interdisciplinaire! complexe! avec! la! plupart! des! si! pas tout! de! ces! champs. #! Pour! le! normatif! discours! (qu'il
soit! dur! réglementaire! ou souple), ! le! fait! le fait que plusieurs de ces définitions aient un certain nombre
de points en commun n'est pas suffisant.

UN! définition! que! est! suffisamment! précis, encore! dynamique! assez! à! capturer! le! essence! de! le!
socioEtechnologique! phénomènes! est! donc! nécessaire! pour! ouverture! les! principes! de! la! robotique! à!
développement! et! problématisation. Cel! besoin! se rapporte! à! des! perspectives! comme! celle! d'Andra! Keay, qui!
parle! des! robots! comme! «... l'un! & environnement; trop! grand! pour! nous! regarder! comme! un! article ».! Bien que!
inévitablement! lié! à! la! progresse! de! technologie! E! "[ce que nous! appel! un! robot! aujourd'hui! c'est! plus!
sophistiqué

⁵ NASA, "Quoi! Est! Robotique? »! La NASA et sait, ! Peut! 18, ! 2015, ! http://www.nasa.gov/audience/forstudents/k4/stories/nasaEknow/what_is_robotics_k4.html. !

⁶ H.! James! Wilson, "Quoi! Est! un! Robot, Quoi qu'il en soit?" ! Harvard! Business! Review, ! Avril! 15, 2015, !
<https://hbr.org/2015/04/whatEisEaErobotEanyway>.

⁷ Idem.

que ce que nous appelions un robot dans les années plus que ça. « Cela a toujours été un problème d'identité » dit Keay. 8 !!

III! devrait! être! dit,! cependant,! ces identités!et! classements ! avoir!toujours! a été!
 problématique! et! problématisé ! quand! partie! de! réglementaire ! initiative,! si! ces!
 avait! à! faire! avec! humains! ou! nonEhumains ! ressemblent.! Loi,! dans! particulier,!
 toujours! finit!de!transformer!toute!identité!en!une!fiction!juridique qui!souvent!a!très!peu!à!
 faire! avec!toute!autre!dimension!physique!ou!scientifique!de!cette!entité. 9!En!même!
 temps,! loi!E!dans!son!sens!le!plus!large!des!impératifs!normatifs!socialement!soutenus!
 E!a!toujours! a prospéré!sur!les!définitions.!L'absence!d'une!«!définition!de!travail» d'une!
 le robot apparaît ! donc! à la fois! comme! témoin! à! les!défis!del'épingleage! technologie!
 dans!son! se précipiter,!et!en!reflet!d'une!possible!faiblesse!à!traiter!dans!!a!proposition!
 document.!!

Dernier!mais!pas!le moindre,!en!plus! au!problème!de! l'absence!d'un!travail!convenu!
définition!(qui!sera!aussitôt!contestée!et!problématisée,!bien!sur!),!là! est aussi! la!
reconnaissance! que! 'définitions!sont! jamais! neutre'.!Cettelidée!était! avancé!
quelques! décennies! il y a! par! Larry ! Peut! quand! réfléchissant ! sur! définir! le!
responsabilité!d'un!organisme!collectif!nonEhumain!(une!innovation!juridique!par
ailleurs pas!sans!intérêt).!Les!définitions,!des!avancées!pourraient,!créer!des!'pseudo-
unités'!qui!sont!proposées! comme! faits,!alors qu'en!réalité,!ils!constituent!des!
oppositions!qui!séparent!arbitrairement!celles! qui!sont!inclus!et! ceux!qui!sont!exclus!
d'une!conceptualisation!partagée! ou! pratique'. 10 ! Ça ! affirmation! volonté! devenir!
plus! et! plus! évident! une fois! le! l'éventail!des!options!disponibles!entre!le!robot!et!l'IA!s'élargit. 11

Que ce soit!prévu!avec!leffroi!ou!avec!l'excitation,! le!défi!de!réglementer! les!multiples!
dimensions!des!interactions!des!humains!Erobot!sont!multiples.!Un!chiffre! de! problèmes!
mettre! avant! pour! réflexion! dans! relation! à! le! donné! cinq! des! principes! sont!
mentionné!brièvement!ici:!

8Moyez! entretien! avec! Andra ! Clé, ! fondateur! de! Robot! Rampe de lancement! et! Gérant!
 Directeur! de! Silicium! Vallée! Robotique ! à Signe ! Brasseur, ! "Quoi! Est! un! Robot?! Le!
 Répondre! 2014 4st! [https://gigaom.com/2014/07/05/](https://gigaom.com/2014/07/05/what-is-ea-robot-the-answer-is-constantly-evolution/) ! Juillet! 5, !
 whatEisEaErobotEtheEanswerEisEconstantlyE evolution/!(Souligné par l'auteur).!

9 David !

Language of a Legal Fiction," *Harvard Law Review* 114, no. 6 (2001): 1745–68.

dix Larry ! Peut,! Partage & Responsabilité,! Nouveau! édition! 1996 ! (Chicago :! Université ! De ! Chicago!Press,!1992),!171ff.

11 Kenneth ! Gradi, ! "Artificiel! Intelligence:! Être! Effrayé,! Être! Très,! Très! Effrayé! (Ou! Pas),"! SeytLines :& Changer& la& Pratique& de& Loi, ! Décembre! 31 ! 2014, ! <http://www.seytlines.com/2014/12/artificialEintelligenceEbeEafraidEbeEveryEveryEafraidEorEnot/>.

Tout d'abord, il est nécessaire d'avoir une perspective un peu plus claire sur ce que c'est et ce que régleme : L'absence mentionnée d'une définition convenue à part, la !! principes mettre avant pour la discussion révèlent le potentiel de confusion quant à la réel ordre du jour : 'régulateur des robots ! dans le réel monde ! a un double sens, ! complet de écueils. Si on ! régule les robots eux-mêmes, le ! le ! apporte une agence implicite ! cela peut être tenu pour acquis dans des contextes où cela peut être indésirable (bien que cette interprétation soit clairement contredite par certains des principes, notamment par !

Principe Non. 2 ! et 5). ! Le deuxième signification, ! plus dans doubler avec quoi ! le cinq des principes eux-mêmes révéler, ! pourrait but ! à 'régulateur le création et utiliser de robots'. Le choix de cette interprétation devrait être plus explicite tout au long de la formulations, évitant la confusion réglementaire. !

Principe hon ! 1 : !

Les robots sont des outils multiples. Les robots ne doivent pas être conçus uniquement et principalement pour tuer et

Le première partie de ce principe est, cependant, encore plus déconcertant. peut trouver le départ déclaration 'les robots sont multi-uses outils' ! comme virtuellement un restriction qui ne sert pas un objectif réel. On ignore pourquoi un robot doit être 'multi-uses' pour être sûr ou, l'inverse, de comment une autrement mortelle robot peut devenir tout moins mortels il est conçu comme 'multi-uses'. ! Cela se rapporte à le prochaine partie du principe : 'les robots ne doivent pas être conçus uniquement et principalement pour tuer et nuire aux humains'. ! Afin de mettre les 'robots tueurs' en conformité avec la lettre ! cette partie du principe, il ! suffirait d'apprendre légalement aux 'robots tueurs' à faire crêpes ou tricoter des chaussettes en laine. ! C'est quoi, ! dans la perspective normative juridique, ! on appellerait une « échappatoire de la conformité créative ». Pour identifier et utiliser une telle échappatoire, un œil juridique n'a pas besoin de regarder plus que une interprétation littérale de le texte. ! Néanmoins, le littéral interprétation est un de le primaire règles de l'interprétation en loi, ! lorsque l'interprétation conforme à ! « esprit de la règle » n'est peut-être pas pratique. ! Les explications données à ce principe dans la Loi de 2010 original document faire pas sembler à vraiment adresse ce ! plutôt basique approche à interpréter les règles et ses conséquences dans ce contexte particulier.

Le! commentaire! à! ce! principe! apparaît ! à! impliquer! autre! potentiel! pièges ! pour! raisonnement! normatif. !
D'abord, comme évoqué! la! propos! des! outils! multi! Euse!, ! il y a! un! effort! pour! mettre! en! avant! l'idée! que! les!
robots! sont! des! outils! comme! les! autres. ! dans! Pour! poursuivre! avec! cette! logique, ! des! équivalences! sont!
cherchées! à! tout! prix! ! En! comparant! un! robot! avec! un! couteau! ou! un! pistolet! utilisé! pour! différents, ! à! la! fois!
relativement! bénins! et! criminels! buts, fait! pas! couverture! pour! le! incohérence! que! là! sont! outils, ! y
compris! des! armes, pour! lesquelles! on! ne! peut! penser! qu'à! d'autres! fins! que! celle! à! première! vue. !!!

Principe! Non. ! 2 : ! Les humains et non les robots sont des agents responsables. Les robots devraient être
conçus et exploités dans la mesure du possible pour respecter les lois existantes et les droits fondamentaux
et les libertés, y compris la confidentialité. ! Le! commentaires! à! ce! principe! semblent! ajouter! plus! de! confusion!
que! de! clarté. ! Tout! d'abord, ! une! relativement! faible! matière, ! il y a! la! présomption! que! personne! n'est! risqué!
délibérément! de! partir! à! construire! un! robot! qui! casse! la! loi. ! Cela! met! une! présomption! cela! n'a! pas!
fondation! dans! le! monde! réel! de! la! déviance! et! la! défiance, ! comme! l'a! révélé! le! par! socio! légal! parmi! la! grande!
population! ainsi! que! parmi! les! cols! blancs. ¹² Deuxièmement, ! le! plus! important, ! la! façon! de! responsabilité! est!
assignée! par! le! commentaires! apparaît ! à! ignorer! les! deux! le! chemin! 'loi! pense' ¹³ comme! Bien!
comme! le! chemin! les! robots! peuvent! ne! pas! réussir! à! atteindre! les! objectifs! et! les! désirs! que! les! humains
ont! spécifiés. ¹⁴

Comme! élément! supplémentaire! ici, ! il! devrait! être! légalement! mentionné! qu'en! l'absence! de! une! définition! de!
travail! claire, ! l'IA, ! les! robots! d'apprentissage, ! etc, ! sont! tous! concernés! par! des! principes! et! leur! paramètres. !
Leur! mécanique, ! cependant, ! peut! Bien! soit! plus! complexe! que! la! loi! discours! normatif! peut! traiter! en!
l'absence! d'une! clarté! définition#. ! Robots ! et! IA ! Machines! peut! Bien! apprendre! à! accord! avec! 'des
exceptions' ! avant! que! la! loi! apprenne! à! gérer! les! différences'. ¹⁵ De même, ! les! autres! disciplines! semblent! à!
indiquer! que! Nombres! (dans! ce! cas! particulier, ! 'programmation')! peut! Bien! être! plus! que! juste! que, !
Nombres! être! inaliénablement ! complété ! par! associé !

¹² Certains! utile! cependant! lâche! exemples! Ryan ! Mathews ! et! Watts ! Wacker, ! L'avantage! du! &
déviant! :& Comment! Fringe! Idées! Créer! Mass! Markets ! (Maison! aléatoire, !
2010); ! Kelly ! Pêcheur, ! "Le! Psychologie! de! Fraude: ! Quoi! Motive! Les! fraudeurs ! à!
Commettre! Crime?" ! (Social! Science! Research! Network, ! March! 31, ! 2015), ! <http://papers.ssrn.com/abstract=2596825>. !

¹³ Gunther ! Teubner, ! "Comment! le! Loi! Pense: ! Vers! un! Constructiviste!

Epistemology! of! Law », ! Law! and! Society! Review 23, no! 5 (1989): ! 727–57.

¹⁴ Voyez! pour! exemple! le! orthogonalité ! théorie! dans! Pseudo! Boström, ! "Le!

Superintelligent ! Volonté! Motivation! et! Instrumental! Rationalité! dans! Avancé!

Artificiel! Agents, ! Esprits! et! Machines, ! 2012, ! <http://www.nickbostrom.com/superintelligentwill.pdf>. !

¹⁵ Ceci! est! construit! sur! les! idées! d'identité! et! de! différence, ! de! Leibniz! à! Kant... ! Gilles!
Deleuze! & Q&Le! Point! de! Vue! (Le! Pli, ! Leibniz! et! Le! Baroque) & 1986 & FRA & Sub! & ITA, ! 2012, ! http://www.youtube.com/watch?v=2ZrA_7ewQGs&feature=youtuve_gdata_play
euh.

avec un! narratif! mouvement! que,! un! pourrait! dire! ici,! peut! être! interprété ! différemment!
par!la!machine!que!par!l'humain,!mais!peut!toujours!être!interprété!par!elle. 16!!!

Principe! Non.! 3 : ! Les robots& sont& des& produits.& Ils& devraient& être& conçus& en
utilisant& des& processus& qui&assurent&leur&sûreté&et&sécurité. Ce!principe!soulève!des!
problèmes!de!distinctions! dans! le! légal! autodéfense ! débat;! délit ! problèmes,! assimilation!
de! responsabilité,! etc.,! problèmes! que! dépendre! très! beaucoup! de! le! contexte! et! de! le!
étendue! à! lequel,! comme! mentionné!ci-dessus,!le!régulateur!sera!disposé!à!interpréter!des!
équivalences!entre! robots! et! autre! les types! de! outils! ou! entre! robots! et! autre! les types!
de! propriété! articles.!

Principe! Non.! 4 :!

Les&robots&sont&fabriqués&artefacts.&Ils&ne&doivent&pas&être&conçus&de&façon&trompeuse&pour&exploiter&le

Principe! Non.! 5 : ! La& personne& ayant& la& responsabilité& légale& pour& un& robot&
devrait& être& attribuée. Ce!principe!apparaît!assez!clair!dans!le!contexte!dans!lequel!nous!
sommes!un! loin!de!!l'IA!capable!de!remplir!techniquement!les!exigences!pour!une! responsabilité.!
At! le même! temps,! compte!des!éléments!complexes!et! des champs! que! entrer! le! maquillage!
de! un! robot (voir! ci-dessus! lors! de! discussions! sur! des! questions! liées! à! la! définition),!
des! défis! à! l'approche! réglementaire! resteront! de! le! indiquer! de! voir! de! pris ! à part! le!
divers! degrés! de! responsabilité,! quand! des choses! aller! faux.! Ayant! un! 'inscrit! gardien',!
porteur! de! responsabilité,! est! seulement une partie ! de! le! solution.! Le! responsabilité!
palier! entité! volonté! exiger! plus loin!

16 Marc ! du! Sautoy, ! Narrative& et& Preuve :& Deux& Côtés& de& la& Même& Équation ?& |
& TORCHE ! (TORCH,! The! Oxford! Research! Centre! in! the! Humanities,! 2015),! <http://www.torch.ox.ac.uk/narrativeEandEproofEtwoEsidesEsameEequationE0>.

17 Voyez ! par!exemple!Cynthia!Breazeal,!“Emotion!and!Sociable!Humanoid!Robots,”!
International&Journal&of&Human&Computer&Studies 59,!no.!1–2! (Juillet!2003):!129ff,!
doi:10.1016/S1071E5819(03)00018E1.

réflexion, ! comme ! volonté ! le ! taper ! de ! dommage(s) ! que ! peut ! être ! attribué ! à ! tel ! entités, ! en ! compte, !
comme ! quelques ! suggérer, ! que ! pour ! le ! d'abord ! temps ! 'le ! la promiscuité ! de ! données ! est ! récemment !
associée ! à ! la ! capacité ! de ! faire ! des ! dommages ! physiques ». 18 !

Au lieu ! de ! conclusions !

Pour boucler la boucle de ces ! notes, ! avec ! une ! autre ! des ! images ! poétiques ! de ! Bradbury, ! on ! peut ! dire ! qu'en !
concernant ! les ! souris ! robots, ! c'est ! très ! vraisemblable ! que ! dans ! le ! contexte ! de ! pouvoirs ! de ! marché ! et ! de !
déréglementation ! des ! capacités ! scientifiques ! combinés, ! nous ! serons ! plus ! sûrement ! d'abord ! 'saut ! désactivé !
le ! faisaient ! et ! construire ! notre ! ailes ! sur ! le ! chemin ! vers le bas !. Dans ! tel ! circonstances, ce qu'on peut
faire, c'est ! s'assurer ! d'être ! relativement ! préparé ! ce ! saut ! par ! déjà ! interrogeant ! et ! problématiser !
notre ! relation ! avec ! science ! et ! technologie ! et ! par ! intensifiant ! notre ! réflexion ! sur ! vie ! dans ! robot !
villes !. Adressant ! 'outils', ! 'produits', ! 'artefacts' ! et ! 'agents', ! il ! faut ! prendre ! en !

St ! celle d'Augustin ! réflexion ! sur ! le ! complexe ! connexion ! entre ! langue ! et ! interprétation ! comme ! un !
chemin ! à ! révélateur ! un ! Plus profond, ! existentiel ! niveau ! de ! compréhension de soi !. La ! façon ! que !
nous ! pense ! normativement ! environ ! human ! E ! robot ! interaction ! 19 ! volonté ! dire ! comme ! beaucoup ! à propos
de ! le ! robot ! comme ! à propos de ! le ! humain !. À ! emprunter ! depuis !
Ginabattista ! celui de Vico ! 1725 ! Nouvelle & Science, 20 ! nous ! besoin ! à ! ours ! dans ! esprit ! que ! notre !
penser ! à ! robots ! est ! ancré ! dans ! un ! donné ! culture ! contexte !. Cela signifie ! que, ! dans ! réfléchissant ! à
propos de ! le ! normatif ! paramètres ! de ! robot ! les villes, ! le ! sociale ! scientifique ! volonté ! pas ! accord ! avec !
un ! champ ! de ! idéalisé ! et ! putativement ! 'sujet ! indépendant ! objets', ! mais ! en ! quête ! sur ! un ! monde ! qui !
est, ! fondamentalement, ! son ! propre !. Le ! processus ! de ! régulateur ! des robots ! est, ! donc, ! un ! processus ! de !
auto-compréhension, ! enraciné ! dans ! un ! étant donné ! le ! contexte ! historique ! et ! la ! pratique !. Une ! compréhension !
qui ! n'aboutit ! pas ! automatiquement ! dans ! des ! propositions ! propres, ! normatives, ! ressemblant ! à !

18 Ryan ! Calo, ! «La robotique ! et ! les ! leçons ! de ! la ! cyberloi», ! California & Law & Review 103 ! (2015): ! 513.

19 Sois ! ce ! interaction ! compris ! sur ! celui de Latour ! spectre ! de ! faits ! et ! agence ; !
Bruno ! Latour, ! "Comment ! à ! Parler ! À propos de ! le ! Corps ? ! Le ! Normatif ! Dimension ! de !
Science ! Études, ! "Corps & Société 10, ! Non ! 2-3 ! (Juin ! 1, ! 2004): ! 205-29, !
doi:10.1177/1357034X04042943; ! Bruno ! Latour, ! "Corps, ! Cyborgs ! et ! la ! Politique ! de ! Incarnation », ! dans !
The & Body: & Darwin & College & Lectures, ! éd. ! Sean ! Sweeney ! et ! Ian !
Hoder, ! 2002, ! 127-41.
20 ! Giambattista ! Vico, ! La & nouvelle & science & de & Giambattista & Vico ! (L'Université de Cornell !
Presse, ! 1744).

Commentaires sur la responsabilité, la conception des produits et les notions de sécurité

Paula Boddington, Département d'informatique, Université d'Oxford.

Les commentaires ci-dessous portent essentiellement sur la règle 2 et sur la règle 3.

Certains de ces points sont illustrés par la prise en compte au sens large de l'utilisation de la robotique dans les contextes des soins infirmiers/personnes âgées et des soins sociaux.

Règle 2 : les êtres humains, et non les robots, sont les agents responsables.

Si les humains sont des agents responsables, mais que les robots ne le sont pas, cela implique que là où les robots sont utilisés pour remplacer les humains ou une partie de l'action humaine, alors les attributions de responsabilité autrefois attribuées à l'agent humain ou aux actions humaines sont alors soit déplacées dans un système plus large, soit peut-être ignorées.

Les robots seront utilisés dans un système d'agents et de comportements humains. la recherche révèle qu'il peut y avoir de fortes cultures locales et des valeurs concernant les lignes de responsabilité et l'imputabilité.

Pour un exemple de la façon dont les responsabilités et les responsabilités peuvent être déplacées, si un robot prend en charge certains des rôles d'un assistant de santé dans un cadre d'attribution, alors les responsabilités peuvent alors être déplacées de diverses manières possibles vers différents acteurs au sein du système de gestion des soins de santé. un manque de conscience chez un employé, par exemple, peut devenir une difficulté à comprendre ou à faire fonctionner des machines.

Tracer et comprendre ces chaînes de responsabilité et d'imputabilité peuvent être complexes.

La règle 2 parle de respecter les lois existantes et les droits et libertés fondamentaux, y compris la vie privée. Cependant, en plus, dans certains contextes, il y aura des protocoles et pratiques plus spécifiques et locaux auxquels il sera souhaitable que les robots se conforment.

Par exemple, au sein du NHS, il existe des normes de soins qui visent à fournir des soins centrés sur la personne et à traiter les patients avec dignité.

déterminer l'impact de l'utilisation de robots sur des valeurs aussi riches et contextuelles peut être très important, mais peut être une question plus difficile que celle du simple respect de la loi et des responsabilités prévues par la loi. Cela a de l'importance dans la conception et l'utilisation des robots, sauf s'il est supposé que cela est largement compris.

Règle 3 : les robots sont des produits. Ils doivent être conçus à l'aide de processus qui garantissent leur sûreté et leur sécurité.

Est-ce que cela fait simplement référence à la sécurité physique ?

Cette règle stipule que les produits doivent être sûrs. Cependant, bien que les règles d'éthique soient souvent formulées en termes d'évitement des dommages, il n'est pas nécessaire d'aspirer à concevoir des produits simplement « sûrs ».

Là où les robots remplacent ou étendent l'action humaine, la tâche qu'ils entreprennent peut ne pas être entièrement transparente.

e. Inversement, les robots peuvent être conçus pour remplacer certains de ces aspects de la tâche. Cependant, ceci peut être enfreint.

Des problèmes de sécurité surgissent alors parce que découvrir comment l'utilisation de robots peut perturber, voire améliorer, certains aspects éventuellement cachés de tâches que les robots prennent en charge, peut impliquer des analyses et des recherches considérables. Il est possible qu'un tel travail important ne soit pas reconnu ni reconnu. être considéré dans un domaine étendu. Cela se traduit par une prise en compte de la question de la façon dont la fabrication de risques et de la fabrication de la fabrication et de ce qui est en train de faire de la manière. Des problèmes de sécurité devraient être considérés comme une question de sécurité.

L'attention à la sécurité comprendra bien entendu l'examen des problèmes de sécurité de l'utilisation des robots dans un système plus large.

confusion. Parfois, la déshydratation est aggravée par des difficultés à atteindre et à gérer les boissons. Supposons qu'un système robotique puisse être conçu pour aider ces patients à boire.

Cependant, un tel système pourrait potentiellement avoir des conséquences négatives importantes dans un contexte particulier. Une fois de plus, cela entraîne un "blocage de lit", car les patients doivent alors trouver un logement dans des établissements capables de répondre à leurs besoins.

Lorsque la règle 3 parle de « sécurité », il est clair que cela inclut ou n'inclut pas des considérations sur la façon dont le robot pourrait fonctionner dans un système de travail plus large ?

Contribution de Roeland de Bruin et Madeleine de Cock Buning à la
Atelier AISB sur les principes de la robotique, 4 avril 2016, Sheffield UK

1. Introduction

Cinq ans se sont écoulés depuis la publication des principes de robotique de l'EPSRC développés par un panel d'éminents experts britanniques en robotique et en intelligence artificielle lors d'une retraite financée par l'EPSRC/AHRC.¹ Les principes, qui visaient à « réglementer les robots dans le monde réel », étaient énoncées sous la forme de cinq « règles » et sept « messages de haut niveau ». Les principes ont en effet eu un impact significatif sur la robotique britannique recherche, et continuent de provoquer des débats substantiels. Étant donné que le public s'inquiète actuellement de la le développement des technologies robotiques s'intensifie nous jugeons utile de revoir les principes pour considérer leur pertinence continue selon les critères suivants.

Nos contributions se concentrent sur le deuxième principe :

Principe 2 : Les humains, et non les robots, sont des agents responsables. Les robots doivent être conçus ; exploité dans la mesure du possible pour se conformer aux lois existantes et aux libertés et droits fondamentaux, y compris la vie privée.

En fait ce deuxième principe des principes de la robotique de l'EPSRC est double. D'une part le traite de la responsabilité - y compris la responsabilité - pour les actions du robot, d'autre part, la implique des méthodes de conception de machines qui peuvent aider à la conformité aux lois et droits et libertés fondamentaux, y compris la vie privée.

Étant donné que la responsabilité et la conception constituent l'épine dorsale de l'introduction de la technologie robotique, comme pour exemple intégré dans les voitures intelligentes autonomes de notre société, nous testerons ce double principe en se concentrant sur le développement et le déploiement actuels de voitures intelligentes autonomes. Que cela deuxième principe EPSRC peut être considéré comme évolutif, sera testé par rapport à trois critères :

1. Validité - est le principe correct car les déclarations sur la nature des robots, les développeurs de robots, et la relation entre les robots et les gens, ou est-elle ontologiquement erronée, inexacte, out daté ou trompeur.
2. Suffisance/généralité — le principe est-il suffisant et assez large pour couvrir tous les problèmes importants qui pourraient survenir dans la réglementation de la robotique dans le monde réel ou sont préoccupations importantes ignorées.
3. Utilité - est le principe d'utilisation pratique pour les développeurs de robots, les utilisateurs ou les législateurs, dans déterminer des stratégies pour les meilleures pratiques en robotique, ou des normes ou cadres juridiques, ou sont ils ont limité leur utilisation par manque de spécificité ou en autorisant des exceptions critiques.

¹<https://www.epsr.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>

2. Etat de l'art

2.1 État des voitures intelligentes autonomes (AIC)

Avant de mettre le principe à l'épreuve, nous présenterons brièvement l'état de l'art des AIC.

Actuellement, les voitures grand public sont de plus en plus équipées de technologies qui assistent certains aspects de la conduite. Des exemples de telles technologies comprennent l'assistance au maintien de voie, le freinage d'urgence, aide au stationnement et régulateur de vitesse adaptatif. Dans un avenir proche, des niveaux plus élevés d'automatisation des voitures deviennent disponibles, menant éventuellement à l'introduction de véhicules entièrement autonomes.

De plus, certaines voitures sont déjà équipées de certaines formes d'automatisation. Il y a même prototypes disponibles qui peuvent conduire sans opérateur humain. Google est actuellement pionnier en soi au volant de la technologie automobile, et a mis un prototype AIC entièrement fonctionnel à l'essai sur route à Bay Area, en Californie, début 2015.² Toujours dans l'Union européenne, les constructeurs automobiles se concentrent sur le développement de la technologie AIC.³ Scania teste le "Platooning": un train routier de camions autonomes qui suivaient de manière autonome un camion contrôlé par un humain à la tête du convoi a été déployé sur les routes néerlandaises.⁴ Volvo prévoyait de déployer 100 voitures qui devraient être en mesure de prendre en charge tous les aspects de la conduite en Suède d'ici 2017⁵ et en Allemagne, une partie de l'autoroute A9 entre Munich et Berlin est réservé aux tests approfondis de véhicules autonomes dans les années à venir.⁶

Une définition des voitures intelligentes autonomes se compose de trois éléments. L'autonomie est liée au niveau de l'intervention humaine nécessaire à l'exploitation, qui peut être vue comme un spectre : un moindre besoin l'intervention humaine implique un niveau plus élevé d'autonomie. L'intelligence se rapporte à la manière dont un système peut percevoir son environnement et est capable d'adapter son comportement à des environnements changeants. Cela comprend la capacité d'apprendre, de traiter des informations complexes et de résoudre des problèmes⁷. véhicules motorisés, utilisés pour le transport de marchandises et/ou de personnes et pour la réalisation de services.

² Wikipedia, "Google driverless car", disponible sur Internet à l'adresse http://en.wikipedia.org/wiki/Google_driverless_car (dernier accès le 17 mars 2015), en référence à Matt O'Brian, "Google's 'goofy' new voiture autonome un signe des choses à venir", 22-12-2014, disponible sur Internet à l'adresse http://www.mercurynews.com/business/ci_27190285/googles-goofy-new-self-driving-car_sign-things (dernière consultation le 28 janvier 2016).

³ Voir par exemple <https://www.media.volvocars.com/us/en-us/media/pressreleases/145619/volvo-car-groups-first-self-driving-autopilot-cars-test-on-public-roads-autour->> (Volvo), <http://www.pcmag.com/article2/0,2817,2387524,00.asp>, (Volkswagen) <http://www.bbc.com/news/technology-25653253> (BMW) (dernier accès le 28 janvier 2016).

⁴ Voir <http://www.scania.nl/about-scania/media/platooning/> (consulté le 20 mars 2015).

⁵ Alecander Stoklosa, "Volvo Has a "Production-Viable" Autonomous Car, Will Put It on the Road by 2017", disponible sur Internet à <http://blog.caranddriver.com/volvo-has-a-production-viable-la-voiture-autonome-la-mettra-sur-la-route-d-ici-2017/>. (dernière consultation le 20 mars 2015).

⁶ Stephen Edelstein, "Germany plans autonome car test program on high speed autobahn", 28 janvier 2015, disponible sur Internet à l'adresse http://www.motorauthority.com/news/1096521_germany-plans-autonomous-car-test-program-sur-l'autoroute-a-grande-vitesse (dernier accès le 20 mars 2015).

⁷ Voir Madeleine de Cock Buning, Lucky Belder & Roeland W. de Bruin, Document de travail : « Mapping the Legal Framework for the introduction into Society of Robots as Autonomous Intelligent Systems », p. 3-4, disponible sur Internet à http://www.caaai.eu/wp-content/uploads/2012/08/Mapping-L_N-fw-for-AIS.pdf (dernière consultation le 28 janvier 2016) et les références à Samir Chopra et Laurence F. White, A Legal Theory for Autonomous Intelligent Agents (Ann Arbor : University of Michigan Press 2011) à la p. 10 (autonomie) et Collin R. Davies, « Une étape évolutive dans les droits de propriété intellectuelle – Intelligence artificielle et propriété intellectuelle », 27 Computer Law & Security Review 2011, p. 601-619 (renseignement); et par les mêmes auteurs le chapitre « Mapping the Legal Framework for the Introduction into Society of Robots as Autonomous Intelligent Systems », in Sam Muller et al (eds.), The Law of the Future and the Future of Law, series 2012 (De Cock Buning, Belder & De Bruin 2012), p. 195-210.

Les AIC peuvent contribuer à trouver des solutions aux défis auxquels notre société est actuellement confrontée. Route la sécurité augmentera considérablement lorsque « l'erreur humaine » sera supprimée comme facteur de causalité les accidents. Les AIC pourraient réduire considérablement les risques d'accidents de voiture puisque 93 % des accidents de la circulation sont causés par une défaillance humaine,⁸ entraînant 1,3 million de décès et 50 millions de blessés graves dans le monde chaque année.⁹ En plus de contribuer à la sécurité routière, les AIC peuvent conduire à une utilisation plus efficace du réseau routier, réduire les émissions de CO2 et contribuer à l'amélioration de la mobilité des personnes handicapées.¹⁰ La mise en place des AIC pourrait ainsi apporter des réponses pour réduire les risques aujourd'hui manifestes qui résultent de l'évolution technologique.

innovation au cours des dernières décennies.¹¹

Cependant, tout le monde n'est pas optimiste quant à un avenir sans conducteur. Il est indiqué que même si les AIC pourraient être bénéfique pour la sécurité routière, d'autres risques découleront de l'introduction des véhicules autonomes. Les AIC vont être vulnérable au piratage par exemple. En outre, les modèles commerciaux et l'emploi dans les taxis et les marchés des transports vont changer de manière significative tandis que les conducteurs pourraient éventuellement devenir obsolètes après l'autonomisation de la conduite.¹² En outre, les risques d'accidents pourraient augmenter lorsque des voitures autonomes et non autonomes coexistent sur les mêmes routes.¹³

2.2 Etat du droit

Une certitude suffisante quant au statut juridique est essentielle pour la croissance et l'acceptation sociétale des consommateurs technologie. L'incertitude provoque le contraire. Dans ce cas, la machine pourrait-elle être la réponse à la machine? Ci-dessous, nous discuterons brièvement des questions de responsabilité qui remettent actuellement en cause l'introduction et déploiement dans la société des AIC et aborder les solutions possibles de technologie de la preuve pour certains d'entre eux. ces défis qui pourraient impliquer la protection de la vie privée dès la conception.

Responsabilité

Réglementation en vigueur dans l'UE concernant la responsabilité et la responsabilité pour les dommages pouvant être causés par Les CIA posent des défis en termes d'innovation dans le domaine des CIA et d'acceptation sociétale de celle-ci. Sur d'une part, les producteurs d'AIC craignent qu'en vertu de la directive sur la responsabilité du fait des produits (PLD), ils puissent être facilement

⁸ Bryant Walker Smith, « Human error as a cause for vehicle crashes », 18 novembre 2013, disponible sur Internet à <<http://cyberlaw.stanford.edu/blog/2013/12/human-error-cause-vehicle-crashes>> (dernière consultation le 28 janvier 2016).

⁹ OCDE, « Panorama de l'OCDE 2013 : Statistiques économiques, environnementales et sociales », 2013, disponible sur Internet à l'adresse <<http://www.oecd-ilibrary.org/sites/factbook-2013-en/06/02/03/index.html?contentType=&itemId=/content/chapter/factbook-2013-50-en&containerItemid=/content/serial/18147364&accessItemid=&mimeType=text/html>> (dernier accès le 28 janvier 2016), également cité dans Gillian Yeomans, « Autonomous Vehicles – transfert de contrôle : opportunités et risques pour l'assurance », disponible sur Internet à l'adresse <<https://www.lloyds.com/~media/lloyds/reports/emerging%20risk%20reports/autonomous%20vehicles%20final.pdf>> (dernière consultation le 28 janvier 2016)(Yeomans 2014) à la p. 5.

^{dix} Voir par exemple Yeomans 2014, p. 5. Aussi Anne Pawsey, « Autonomous Road Vehicles », septembre 2013, p. 1. Disponible sur Internet à <<http://www.parliament.uk/briefing-papers/post-pn-443.pdf>>, (POSTnote 2013); Robolaw 2014, p. 42.

¹¹ La pollution, le changement climatique, l'exclusion sociétale des «parties les plus faibles» et les risques élevés d'accidents sur les routes (européennes) peuvent tous être considérés comme le résultat des processus de modernisation et d'individualisation qui ont eu lieu au cours du siècle dernier. Ces effets secondaires doivent maintenant à leur tour être traités.

Voir pour l'identification et une étude sur le concept de société du risque par Ulrich Beck, son livre Risk Society, Towards a New Modernity, Londres : Sage Publications 1992.

¹² Voir par exemple Scott Le Vine & John Polak, « Automated Cars : A smooth ride ahead ? », février 2014, p. 14, disponible sur Internet via <<http://www.theitc.org.uk/docs/114.pdf>> (dernier accès le 28 janvier 2016).

¹³ Voir Wayne Cunningham et Antuan Goodwin, « Six raisons d'aimer ou de détester les voitures autonomes », 8 mai 2013, disponible sur Internet à l'adresse <<http://www.cnet.com/news/six-reasons-to-love-or-loathe-autonomous-cars/>> (dernière consultation le 28 janvier 2016).

tenu responsable des dommages causés par des AIC défectueux, ce qui aurait un effet dissuasif sur innovation.¹⁴ Alors que, d'autre part, le cadre actuel en matière de responsabilité du fait des produits ne

fournir une boîte à outils simple permettant aux consommateurs de tenir les fabricants d'AIC responsables des défauts de leurs produits à tous. Une charge de la preuve assez lourde incombe aux consommateurs pour établir qu'il y a effectivement eu un défaut de l'AIC, ainsi que sur le lien de causalité entre le défaut et le dommage survenu.

Apporter des preuves sera plus complexe lorsque l'autonomie et l'intelligence des voitures augmenteront, car les victimes devront procéder à une analyse (technologique) approfondie, entre autres, du logiciel (d'origine), du mises à jour et les données opérationnelles dont un AIC est équipé, afin d'établir la cause précise d'un accident. Dans le même temps, les fabricants ont amplement la possibilité de se défendre contre réclamations en responsabilité. Face aux AIC, le PLD ne protège pas de manière optimale les intérêts des consommateurs en leur fournissant des moyens faciles d'obtenir une indemnisation pour les dommages qu'ils ont subis en raison de AIC défectueux des fabricants.

La marge d'amélioration de la législation actuelle est en outre constituée par les différents régimes européens non harmonisés en matière de responsabilité des véhicules à moteur. À ce jour, 28 cadres différents sont en place dans l'Union européenne. Par exemple, la loi française Badinter¹⁵ impose un régime strict de responsabilité sans faute dans afin d'apprécier si le conducteur ou le gardien d'une voiture doit indemniser ou non les dommages des victimes (autres que le conducteur)¹⁶ d'accidents dans lesquels des véhicules à moteur sont impliqués. La responsabilité ne peut être exonérée que si le conducteur (ou le gardien) peut prouver une faute inexcusable de la part de la victime¹⁷.

"Wegenverkeerswet" attribue une responsabilité (semi-stricte) au propriétaire ou au détenteur (note : plutôt qu'au conducteur ou un gardien) d'un véhicule à moteur impliqué dans un accident ayant causé des dommages à des usagers de la route non motorisés.¹⁸ Au moins 50 % des dommages subis doivent être indemnisés, sauf cas de force majeure pouvant être prouvée.¹⁹ Au Royaume-Uni, des règles de négligence sont appliquées pour établir si le conducteur d'un véhicule à moteur peut être tenu responsable. Dans de tels cas, il n'existe pas de régime de responsabilité stricte²⁰ au Royaume-Uni, bien que le niveau de diligence exigé des conducteurs de véhicules à moteur soit plutôt élevé. Jurisprudence explique qu'un conducteur qui perd connaissance sans qu'il y ait eu faute de sa part agit néanmoins avec négligence²¹, de même que le conducteur dont les freins échouent alors que cette défaillance n'était pas prévisible²².

Cependant, les victimes d'accidents causés par des véhicules à moteur doivent prouver que les conducteurs étaient à

¹⁴ Voir Erica Palmerini, Federico Azzari, Fiorella Battaglia et al, D 6.2, « Guidelines on Regulating Robotics », 22 septembre 2014, (RoboLaw 2014), p. 60.

¹⁵ Loi "tendant à l'amélioration de la situation des victimes d'accidents de la circulation et à l'accélération des procédures" d'indemnisation".

¹⁶ Voir A. Tunc, « The 'Loi Badinter' – Ten Years of Experience », 3 Maastricht Revue de droit européen et comparé, 1996 (Tunc 1996), p. 330. L'article 3 se lit comme suit : « Les victimes hormis les conducteurs [...] sont indemnisées des dommages résultant de leur personne qu'elles ont subis, sans que puissent leur être opposées à leur propre faute ».

¹⁷ Voir aussi Tunc 1997, p. 335.

¹⁸ L'indemnisation des dommages subis par les victimes à l'intérieur d'un véhicule à moteur est régie par les règles générales de responsabilité prévues par Article 6:162 du Code civil néerlandais.

¹⁹ Marloes de Vos ea, Cour suprême des Pays-Bas 2 juin 1995, NJ 1997/700-702, et Said Hyati ea, 5 décembre 1997 NJ 1998/400-402. La notion de « Betriebsgefahr » est empruntée à l'allemand Straßenverkehrsgesetz.

²⁰ Ou la responsabilité présumée comme on l'appelle dans la loi écossaise.

²¹ Roberts c. Ramsbottom [1980] 1 WLR 823, également cité dans. Cees van Dam, European Tort Law, Oxford : Oxford University Press 2006 (Van Dam 2006), p. 364, note de bas de page 52.

²² Henderson c. HE Jenkins & Sons and Evans [1970] AC 282, cité dans Van Dam 2006, à la p. 364, note de bas de page 53. Van Dam en outre prend note de l'arrêt Worsley v Hollins [1991] RTR 252 (CA), dans lequel les juges ont estimé que la plainte de la victime pour négligence avait échoué parce que le défendeur pouvait prouver que bien que ses systèmes de freinage aient échoué, causant ainsi des dommages, son minibus avait récemment été révisé et passé son CT.

faute, c'est-à-dire : ils avaient agi avec négligence.²³ Les différences significatives dans la façon dont la responsabilité véhicules est adressée dans tous les États membres, n'est pas bénéfique pour le développement, l'assurance et déploiement des AIC en Europe. En tout état de cause, les régimes nationaux attribuant la responsabilité aux conducteurs d'automobiles les véhicules doivent être mis à jour afin de pouvoir aborder la responsabilité des véhicules sans conducteur humain.

Confidentialité

Alors que l'avènement de la technologie AIC est prometteur en termes de sécurité accrue sur les routes, résultant en moins de dommages à couvrir, les compagnies d'assurance observent également que lorsqu'un accident se produit causé par une technologie autonome, il « aurait besoin d'une expertise approfondie en matière d'analyse logicielle et matérielle afin de savoir comment et pourquoi cela s'est produit ».²⁴ L'une des options pour évaluer la cause d'un accident, et donc d'aider à répondre à la question de la responsabilité, pourrait être d'équiper véhicules avec des boîtes noires, ou avec des solutions télématiques connectant les AIC à une infrastructure dédiée, et/ou à des serveurs distants²⁵. Les objectifs de ces types de technologies sont, entre autres, d'enregistrer les déplacements des voitures autonomes et les choix opérationnels qui sont faits par la voiture elle-même ou par le conducteur contrôlant son mouvement, ainsi que des données concernant les événements et les objets dans le proximité d'un véhicule autonome. La technologie de la boîte noire enregistre et stocke les données recueillies dans un véhicule et offre un potentiel d'évaluation ultérieure. La technologie télématique peut avoir des applications. Les données pourraient non seulement être utilisées pour évaluer les erreurs et les causes des dommages après survenue d'accidents, elle pourrait même avoir un effet préventif. Communication de véhicule à véhicule (V2V) et la communication véhicule-infrastructure pourraient être utilisées pour la prévention des accidents en temps réel et servir « la sécurité, la mobilité et les avantages environnementaux » en général.²⁶ Bien que la boîte noire technologies et solutions télématiques telles que V2V et V2I (ci-après dénommées « traçage technologie") peut être prometteuse en termes de prévention des accidents et de répartition des dommages causés par Accidents AIC, ceux-ci imposent également des risques en termes de droit à la vie privée (à l'information) des personnes à l'intérieur et à proximité des voitures équipées de ces technologies.

La confidentialité des informations des citoyens est strictement réglementée dans l'Union européenne par la directive sur la protection des données (DPD)²⁷ et le sera encore plus après l'entrée en vigueur du règlement général sur la protection des données (RGPD)²⁸.

Le cadre actuel et à venir prescrit pour

exemple que déjà lors de la phase de conception des AIC équipés de la technologie de traçage, un

²³ Il existe une règle d'obligation légale qui – dans une certaine mesure – établit une responsabilité stricte pour les conducteurs de véhicules à moteur s'approchant d'un passage à niveau sur la route : "Le conducteur de tout véhicule s'approchant d'un passage à niveau doit, à moins qu'il ne puisse voir qu'il n'y a pas de passage, avancer à une vitesse telle qu'on puisse, au besoin, s'arrêter avant d'atteindre ce passage à niveau », tel que cité dans Van Dam 2006, à la p. 365, note de bas de page 57, renvoyant au Reg. 3 du Pedestrian Crossing Places (Traffic) Regulations 1941, remplacé par le Zebra Pedestrian Crossing Regulations 1971, SI 1971, No. 1524. Une défense qu'un conducteur a à cet égard est la force majeure.

²⁴ Yeomans 2014, p. 18.

²⁵ Yeomans 2014, p.18. Voir également James M. Anderson, Nidhi Kalra, Karlyn D. Stanley et al., Autonomous Vehicle Technology – A Guide for Policymakers, RAND Transportation, Space and Technology Program 2014, (RAND report), p. 94-95.

²⁶ rapport RAND, à la p. 81.

²⁷ Directive 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995 relative à la protection des personnes handicapées concernant le traitement des données à caractère personnel et à la libre circulation de ces données, Journal officiel L 281, 23/11/1995 P. 0031 - 0050

²⁸ Proposition de RÈGLEMENT DU PARLEMENT EUROPÉEN ET DU CONSEIL relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (règlement général sur la protection des données) COM/2012/011 final - 2012/0011 (COD). Veuillez noter que le trilogue entre la Commission européenne, le Conseil de l'Europe et le Parlement européen s'est conclu sur le texte final du RGPD, ce texte n'a cependant pas encore été officiellement publié.

une évaluation d'impact doit être effectuée. En outre, « des mesures techniques et organisationnelles appropriées des mesures de protection des données personnelles contre la destruction accidentelle ou illicite ou la perte accidentelle, modification, divulgation ou accès non autorisés, en particulier lorsque le traitement implique la transmission de données sur un réseau, et contre toutes les autres formes de traitement illicites » doit être mise en œuvre, et doit être activé par défaut.³⁰ Les coûts de sécurité et de mise en œuvre de pointe doivent être pris en compte pour la

mise en œuvre des mesures. En outre, ceux-ci « doivent assurer un niveau de sécurité adapté à la risques représentés par le traitement et la nature des données à protéger ».

Un autre défi encore plus récent est formé par la récente décision de la Cour européenne de justice déclarer le Safe Harbor Framework, qui est à la base de nombreux échanges de données personnelles entre l'UE et les États-Unis d'Amérique, invalide. Il est probable que la technologie de traçage incorporés dans les AIC constitueront la transmission internationale de données (personnelles), à travers le frontières de l'Union européenne, et éventuellement importer ces données aux États-Unis, par exemple via le cloud computing. La CJUE a statué que les États-Unis n'offraient pas un niveau de protection adéquat pour données personnelles, car il est devenu clair après les révélations d'Edward Snowden, que les autorités américaines telles comme l'Agence de sécurité nationale ont un accès facile aux données personnelles traitées par les entreprises et institutions américaines.³¹ Le tribunal a jugé que les pouvoirs des autorités européennes de surveillance sont sapés par les pratiques américaines, qui peuvent ne pas être rendues possibles par une décision de la Commission européenne. Cette décision implique que l'exportation de données personnelles vers les États-Unis n'est plus possible sur la base de la cadre de la sphère de sécurité. Bien que les États-Unis et la Commission européenne négocient actuellement un traité alternatif³², dans l'intervalle, l'échange de données personnelles entre l'UE et les États-Unis n'est pas autorisé sur la base des règles encore invalides de la sphère de sécurité.

3. Testez le principe

Dans cette partie, nous testerons si le deuxième principe EPSRC peut être considéré comme une preuve future contre les critères validité, suffisance/généralité et utilité

3.1 Validité

Etant donné l'état actuel de la technologie et de la loi, la première partie du principe Humains, non robots, sont des agents responsables s'est en effet avérée toujours d'actualité. C'est une affirmation correcte sur le la nature des robots, les développeurs de robots et la relation entre les robots et les personnes. Il peut y avoir toujours soit un être humain soit une personne morale tenue pour responsable et redevable des actions des AIC. La création spécifique d'une entité juridique distincte pour les AIC semble actuellement farfelue compte tenu de la situation actuelle. statut technologique et juridique des AIC, cela ne contribuerait en outre pas à résoudre le problème de la responsabilité

²⁹ Art. 17(1) DPD ; voir également art. 5(1)(b) et section 2 (art. 30 et suivants) sur la sécurité des données dans le RGPD.

³⁰ Art. 23 RGPD.

³¹ Affaire C-362/14, Maximilian Schrems/Facebook [2015].

³² Voir pour les dernières nouvelles sur l'"accord-cadre UE-États-Unis" (accord entre les États-Unis d'Amérique et l'Union européenne sur la protection des informations personnelles relatives à la prévention, la recherche, la détection et la poursuite des infractions pénales) : [http : //ec.europa.eu/justice/newsroom/data-protection/news/150908_en.htm](http://ec.europa.eu/justice/newsroom/data-protection/news/150908_en.htm) (dernier accès le 9 mars 2016).

défis rencontrés comme décrit sous 2.2. Il en va de même pour la deuxième partie du deuxième principe selon laquelle stipule que les robots doivent être conçus ; exploité dans la mesure du possible pour se conformer aux lois et les droits et libertés fondamentaux, y compris la vie privée, se sont avérés toujours valables. Avec un œil sur la technologie de la preuve (à) intégrer dans les AIC, cette idée fondamentale s'est avérée même plus vraie qu'on aurait pu l'imaginer sur sa conception.. Comme nous l'avons vu sous 2.2 en fait les défauts du régime de responsabilité actuel peut être partiellement résolu par des systèmes intelligents de collecte et de sauvegarde de preuves intégrés à l'AIC. Ces systèmes de collecte et de sauvegarde des preuves devraient être conçus de manière à que les données personnelles collectées soient protégées au maximum : privacy by design et privacy by default doit être intégré dans les AIC (technologie de traçage) à tout moment.

3.2 Suffisance/généralité

Dans le même temps, le principe reste encore suffisant et suffisamment large pour couvrir tous les aspects importants problèmes qui pourraient survenir dans la réglementation des AIC dans le monde réel. Les humains, et non les robots, sont agents responsables Les robots doivent être conçus ; exploité dans la mesure du possible pour se conformer aux lois existantes et aux droits et libertés fondamentaux, y compris la vie privée. Aucune préoccupation majeure ne semble être négligé. Bien que certains auteurs semblent soutenir qu'une entité juridique devrait être créée pour les machines intelligentes autonomes, faisant des robots l'agent responsable,³³ cela n'a pas été convaincant pour beaucoup³⁴ et certainement pas pour nous.,

Les défis posés par l'introduction dans la société des voitures intelligentes autonomes et leur la responsabilité pour dommages en elle-même ne semble pas exiger une personnalité juridique distincte. Cela ajouterait simplement un acteur de plus pour l'attribution de la responsabilité. Dans le même temps, cela nécessiterait une refonte substantielle du système de responsabilité tel qu'il est actuellement appliqué au monde réel, alors que la technologie est encore en développement étape comportant le risque de sous ou de surréglementation.

3.3 Utilité

Tant que les moyens juridiques actuels ne sont pas épuisés, notamment en visant une harmonisation plus poussée des régimes de responsabilité législatifs de l'UE en combinaison avec une technologie de la preuve efficace, il n'y a pas preuve qui sous-tendrait un changement de paradigme complet par l'introduction des AIC en tant que responsables agents en eux-mêmes. Étant donné que l'AIC peut en effet être conçu et exploité pour se conformer aux lois existantes l'utilité de ce principe reste évidente. Cependant, les technologies de boîte noire et les solutions télématiques tels que V2V et V2I peuvent être prometteurs en termes de prévention des accidents et de répartition des dommages causés par les accidents AIC, car ceux-ci présentent également des risques en termes de droit à la confidentialité (des informations) des personnes à l'intérieur et à proximité des voitures équipées de ces technologies, les systèmes devraient

³³ Voir par exemple James Boyle, « Endowed by Their Creator?: The Future of Constitutional Personhood », The Future of the Constitution, 9 mars 2011, p. 6, également disponible sur Internet à http://www.brookings.edu/~media/research/files/papers/2011/3/09-personhood-boyle/0309_personhood_boyle.pdf (dernière consultation le 9 mars 2016 Voir également JP Günther, F. Münch, S. Beck, S. Löffler, C. Leroux, & R. Labruto, "Issues of Privacy and Electronic Personhood in Robotics. 21st IEEE International Symposium on Robot and Human Interactive Communication", Paris 2012, tel que cité dans Christophe Leroux, Roberto Labruto, Chiara Boscarato et autres, « Suggestion for a Green Paper on legal issues in robotics », décembre 2012, disponible sur Internet à http://www.eu-robotics.net/cms/upload/PDF/euRobotics_Deliverable_D.3.2.1_Annex_Suggestion_GreenPaper_ELS_IssuesInRobotics.pdf (dernier accès le 28 janvier 2016) ; et Robolaw 2014, p. 24.

³⁴ Voir par exemple Peter Asaro, « Robots and Responsibility from a Legal Perspective », via <http://www.peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf> , dernière consultation le 9 mars 2016 ; et Lawrence Solum, « Legal Personhood for Artificial Intelligences », North Carolina Law Review, (avril 1992), pp. 1231-1287.

inclure le respect de la vie privée dès la conception pour protéger ces droits fondamentaux, comme le prévoient les traités internationaux et européens.³⁵

Il est crucial que ces exigences du droit et de la technologie soient satisfaites avant la contestation de la l'introduction et le déploiement des CIA dans la société peuvent être atteints.

4. Conclusion

Nous pouvons conclure avec diligence que le principe 2 des principes de robotique de l'EPSRC tels que développés par Des experts britanniques en robotique et en IA à la retraite financée par l'EPSRC/AHRC se sont avérés être à l'épreuve du temps lorsque nous nous sommes appliqués à l'état actuel de l'art du droit et de la technologie entourant les AIC.

Les humains, et non l'AICS, sont des agents responsables. Les AIC doivent être conçus ; exploité dans la mesure où possible de se conformer aux lois existantes et aux droits et libertés fondamentaux, y compris la confidentialité par deisgn. Donnant ainsi la preuve du fait que la réponse de la machine est au moins partiellement dans le machine elle-même.

³⁵ Voir par exemple l'art. 7 & 8 de la Charte des droits fondamentaux de l'Union européenne et l'article 8 de la Convention européenne des droits de l'homme.

Traitement équitable des données et robotique

Burkhard Schafer, Edinburgh Law School
Lilian Edwards, Université de Strathclyde

Cette intervention combine le principe 4, Les robots sont des artefacts manufacturés. Ils ne doivent pas être conçus de manière trompeuse pour exploiter les utilisateurs vulnérables ; au lieu de cela, leur nature de machine doit être transparente avec le principe 2, Les robots doivent être conçus ; exploités dans la mesure du possible pour respecter les lois et les droits et libertés fondamentaux en vigueur, y compris la vie privée. Cela suggère une élaboration/spécification des principes que l'intersection entre eux nécessite.

pour ceux qui s'inquiètent de cette formulation, les obligations envers le propriétaire d'un robot doivent être

Les robots posent des défis uniques pour les pratiques équitables de traitement des données, défis qui sont au moins en partie causés par leur capacité à « tromper », si par inadvertance, les personnes avec lesquelles ils interagissent.

Bien avant la technologie moderne, les humains ont développé des techniques de protection de la vie privée, des rideaux aux fenêtres en passant par le voile, en apprenant quand chuchoter pour laver son odeur. ceux-ci protègent non seulement les informations, mais aussi le partage et l'échange d'informations (chuchotement, insonorisation de votre studio)

La loi, avec son système de règles et d'exceptions, a fréquemment reconnu formellement ces mesures de protection à faible technologie.

attente de vie privée» et «de sécurité dans nos maisons et logements», mais aussi des espaces dans lesquels

La technologie robotique menace de rendre ces solutions low-tech au problème de confidentialité de plus en plus redondantes. mintooourhome.

Personne n'est l'hérode ses domestiques.

la maison pourrait anticiper ce qu'elle pourrait voir exactement, elle comprendrait l'environnement normatif (à la fois social et juridique) qui l'empêche de collecter et surtout de partager des données sur son employeur. La compréhension de l'environnement normatif ainsi que celle des capacités sensorielles permettraient alors une évaluation et une gestion rationnelles des risques. (Je fais confiance à mon majordome pour mes sous-vêtements sales, mais pas pour ma chemise trempée de sang.

La

robotiquemenacecesstratégiesdéfensivespasseulementparcequ'ellepeututiliserdescapteurshorsdu

Une partie de la conception éthique devrait donc également être d'indiquer les capacités sensorielles des robots de manière à faciliter l'émergence de défenses « intuitives » du type que nous utilisons avec d'autres humains, et de s'abstenir, dans la mesure du possible, de susciter des inférences trompeuses, et d'inclure la « facilité du mécanisme défensif » dans l'évaluation de l'intrusivité lorsqu'un choix entre différents capteurs peut être fait.

Nous protégeons non seulement les données nous concernant, mais également nos idées commerciales, nos découvertes scientifiques ou technologiques ou nos compétences.

La loi sur la propriété intellectuelle est donc une autre contrainte légale qui doit être observée sous cet en-tête, et une notion plus large de « pratiques équitables de transmission des données » qui va au-delà de la loi sur la protection des données peut être nécessaire. protection/indemnisation ?

Cela, potentiellement, soulève aussi une question qui mène plus radicalement au-delà des Principes. et les devoirs éventuels qui leur sont dus/aux robots.

Par exemple, pour permettre la conception sûre d'un robot peut impliquer une obligation pour les tiers de divulguer ou de partager certaines informations avec le robot qui, par le passé, étaient légalement privilégiées. L'approche américaine selon laquelle il s'agirait d'une copie à des fins fonctionnelles plutôt qu'expressives

– pas la parole – et donc ne devrait pas être une violation du droit d'auteur – cependant, une fois que les machines coordonnent leur action en partageant ces données, même cette absence d'argumentation peut atteindre ses limites.

Les citoyens peuvent choisir d'utiliser une technologie pour empêcher les capteurs de les remarquer (par exemple, camoufler la peinture faciale - <https://cvdazzle.com>), mais cela peut signifier qu'ils acceptent un risque plus élevé que le robot leur tombe dessus. Si des tiers sont impliqués, cela peut créer des problèmes juridiques encore plus complexes. Manipuler intentionnellement le processus d'apprentissage, est-ce que cela entre dans le territoire de la Loi sur les abus informatiques ?

La loi de base sur la négligence et sa distinction entre acte et omission et la façon dont la négligence y fait face en établissant des devoirs envers les voisins, feront partie de la réponse juridique après qu'un accident se soit produit.

1. S'appuyer sur le devoir éthique/social des tiers de ne pas manipuler l'acquisition des connaissances de la machine².
S'appuyer uniquement sur une obligation légale plus étroite de s'abstenir de certaines manipulations de données potentiellement dangereuses³.

Notrelyatallunenvironnementcoopératiflorsqu'onpenseàlasécuritéetàlaconformitéauxloisdurobot

Pour bien comprendre pourquoi ce problème se pose dans le cadre d'une discussion sur la « transparence des capteurs » : Si nous acceptons l'obligation éthique discutée ci-dessus, c'est-à-dire que les robots devraient normalement divulguer comment et avec cela ce qu'ils peuvent détecter, alors ils s'ouvrent inévitablement à la manipulation.

Si nous acceptons 2, alors nous devons affronter le fait qu'aux deux extrémités du spectre, la loi est claire : il est difficile d'avoir des devoirs envers les humains, par contre je peux avoir le devoir envers le voisin de ne pas incendier sa grange, mais ai-je le devoir d'absenter les opérateurs de robots (ou les concepteurs) de ne pas confondre leur formation de robot ? Est-il raisonnablement prévisible que le robot serait confondu ? Après tout, je n'ai même pas le devoir de ne pas mentir à des étrangers, par exemple lorsque je donne des instructions, à moins d'une relation de conseils professionnels. Mais encore une fois, égarer un enfant serait une proposition différente.

Les discussions de l'un d'entre eux ont porté sur les problèmes causés par les humains qui retiennent/déforment/manipulent des données dont le robot a besoin pour fonctionner en toute sécurité et dont ils sont moralement ou légalement redevables.

Mais nous sommes également confrontés à des choix de conception éthiques lorsque des personnes coopèrent et fournissent volontairement des informations qu'elles ne sont pas légalement tenues de fournir, mais qu'elles choisissent/n'empêchent pas par sens du devoir civique.

This could mean that not only robots need to be identifiable as robots,

their sensors as sensors, but robot generated output also must be identifiable as machine, not human generated.

Par conséquent, le thème principal de cette intervention est en fin de compte celui de la transparence algorithmique : les devoirs juridiques et éthiques influencent quand et comment les robots doivent divulguer leurs capacités sensorielles. Une fois qu'ils le font, leur environnement a le choix : coopérer ou ne pas coopérer. de (certaines) données pour disposer de machines plus sécurisées.

Bien qu'il s'agisse principalement de questions juridiques, pour la question de la conception éthique (et conforme aux lois), les développeurs doivent également être en mesure d'anticiper le type d'interaction à attendre et le type d'informations auxquelles ils auront un accès légal.

Les robots peuvent-ils être des agents moraux responsables ? Et pourquoi devrions-nous nous en soucier ?

Amanda Sharkey,

Département d'informatique et de robotique de Sheffield, Université de Sheffield

Principe 2. Les êtres humains, et non les robots, sont les agents responsables.

À première vue, cette déclaration ou ce principe semble convaincant. Il est logique d'insister sur le fait que les humains et non les robots sont des agents responsables. Il est également sensé de suggérer que les robots soient conçus et exploités dans le respect des lois et des droits et libertés fondamentaux en vigueur : il est difficile d'imaginer que quelqu'un suggère le contraire.

Mais, après examen, il devient évident que la déclaration ne donne aucune justification pour dire que les humains et non les robots sont des agents responsables, elle ne fournit aucune indication sur où et quand les robots doivent être utilisés, ni sur les conséquences qui découlent de la supposition que les robots ne sont pas des agents responsables. (a) Pour presumer que les robots et pas les humains sont les agents responsables ? (b) Suffit-il de concevoir les robots pour respecter les lois et les droits et libertés fondamentaux ? et (c) Si les robots ne sont pas des agents responsables, cela devrait-il limiter leurs rôles conférés et les situations dans lesquelles ils sont déployés ?

(a) Quelles sont les raisons de supposer que les humains et non les robots sont des agents responsables ?

Outre la responsabilité légale, il est possible d'identifier deux raisons à cette hypothèse. La seconde concerne la nécessité pour la société d'accepter la responsabilité des artefacts que l'homme a produits.

(i) Machines biologiques versus machines mécaniques : Tenir un agent responsable de ses actions équivaut à le considérer comme un agent moral. Patricia Churchland (2011) discute de la base de la moralité chez les êtres vivants et affirme que la base de l'attention portée aux autres réside dans la neurochimie de l'attachement et des liens chez les mammifères. En plus de l'attachement et de l'empathie pour les autres, les humains et les autres mammifères développent des relations sociales plus complexes et sont capables de comprendre et de prédire les actions des autres. En conséquence, les humains ont un sens intrinsèque de la justice.

Bekoff et Pierce (2009) fournissent de nombreux exemples de preuves d'un sens moral de la justice chez les mammifères. Par exemple, les singes capucins

travailler pour les friandises semblait offensé et refuserait de coopérer davantage s'il voyait qu'un autre singe recevait une récompense plus désirable pour le même travail (Brosnan et de Waal, 2003).

Par contre, les robots ne se soucient pas de leur propre préservation ou de l'évitement de la douleur, sans parler de la douleur des autres. Un corps vivant est une entité autopoïétique intégrée (Maturana et Varela, 1980) dans la mesure où une machine fabriquée par l'homme ne l'est pas.

Responsabilité sociétale : de nombreux auteurs seraient d'accord avec l'implication de l'affirmation selon laquelle les robots ne sont pas des agents moraux à part entière.

car les comportements et les sorties des robots et des systèmes informatiques dépendent nécessairement des concepteurs et des utilisateurs. Des arguments connexes sur le manque d'indépendance vis-à-vis des concepteurs humains ont été avancés dans le passé sur la manière dont les robots, contrairement aux machines vivantes, ne peuvent jamais être considérés comme pleinement incarnés, car ils ont toujours nécessité une intervention et une implication humaines dans leur développement (Sharkey et Ziemke, 2001). L'intervention humaine. ed. Johnson

(2006) fait une distinction utile entre agents moraux et entités morales, et place les robots et les artefacts informatiques dans la seconde catégorie.

- (b) Suffit-il de concevoir des robots pour se conformer aux lois et aux droits fondamentaux existants et libertés, y compris la vie privée ?

Un problème majeur avec la suggestion selon laquelle les robots devraient être conçus pour se conformer aux lois et aux libertés et droits fondamentaux existants, et la raison pour laquelle ce n'est pas suffisant, est que les lois et les droits de l'homme existants n'ont pas été formulés en tenant compte des développements technologiques tels que la robotique. d'apparaître comme des amis et des compagnons, et qui en résultent sont les bienvenus dans nos foyers et notre environnement intime. (Sharkey et Sharkey, 2012; Sparrow et Sparrow 2006), mais la loi sur les droits de la personne ne prévoit aucune protection explicite contre une telle situation.

ont été soulevés à propos de laisser des enfants aux « soins » de robots dans la mesure où leurs attachements aux humains sont compromis (Sharkey et Sharkey 2010) mais encore une fois, il n'y a pas de législation ou de droits qui empêchent explicitement une telle possibilité, autre que celle associée à la négligence envers les enfants. Les humains placent des robots en position de pouvoir sur les humains.

Lorsque les humains prennent des décisions sur la façon d'agir dans des situations sociales, ils doivent faire plus que suivre des règles ou des lois. Ils prennent des décisions basées sur une compréhension morale de ce qui est inapproprié ou approprié pour eux. Les bonnes décisions morales dans les situations sociales. Par contre, un robot ne répondrait pas émotionnellement et pourrait être programmé, par l'intermédiaire d'un « gouverneur éthique », pour évaluer les actions avant de les exécuter, et pour n'effectuer que celles jugées moralement acceptables.

Divers auteurs ont plaidé contre l'idée de pouvoir programmer des robots pour qu'ils prennent des décisions morales. image, compréhension des intentions qui sous-tendent les actions des gens, et compréhension des valeurs et anticipation de la direction dans laquelle les événements se déroulent » (2013, [A /HRC/23/47](#)).

Dans un article intéressant sur les exigences pour créer des robots avec ce qu'ils appellent la « compétence morale », Malle et Scheutz (2014) soutiennent que, entre autres, les robots auraient besoin d'un réseau de normes morales, afin de savoir ce qui est ou n'est pas moralement acceptable. Ils suggèrent qu'il pourrait être nécessaire d'élever les robots dans des environnements humains, puisque c'est peut-être "la seule façon d'exposer la richesse des situations morales humaines et des interactions communicatives" (Malle et Scheutz, 2014).

(RiedlandHarrison, 2016), et les obligeant à rétroconcevoir les valeurs humaines qu'ils représenter.

Il est certes difficile d'exclure la possibilité qu'à l'avenir un robot puisse être formé ou élevé à la morale, mais il existe un certain nombre de raisons d'être sceptique quant à la probabilité de réussite. Les raisons du scepticisme incluent le manque de base biologique de la moralité du robot comme discuté plus tôt. Comme déjà discuté, un robot individuel ne se soucie même pas de son propre corps, et encore moins de celui d'un humain - il ne souffrirait pas si l'une de ses roues était retirée par exemple. généralisable, compréhension des différences entre le bien et le mal.

comportement, comme les robots programmés par Winfield et al (2014) pour agir pour empêcher d'autres robots de tomber dans un piège. Les robots en question pourraient être légitimement loués ou blâmés pour leurs actions.

(c) Si les robots ne sont pas des agents responsables, cela devrait-il limiter les rôles sociaux qui leur sont confiés et les situations dans lesquelles ils sont redéployés ?

La déclaration originale selon laquelle les robots ne sont pas des agents responsables ne précise pas ce que cela implique pour le déploiement des robots. ations, mais aussi parce que les êtres humains devraient avoir le droit d'avoir des décisions de vie ou de mort à leur sujet prises par d'autres êtres humains.

Cet argument peut, et je dirais qu'il devrait, être étendu à d'autres types de décisions où les robots pourraient restreindre les libertés des humains. ourisqué. Un robot n'aurait pas à prendre une décision similaire concernant ses jeunes charges. Le point est que toutes ces décisions sont susceptibles d'impliquer des jugements moraux et des évaluations de situations sociales, et pour des raisons déjà évoquées, le robot est peu susceptible de pouvoir faire de bons choix. Il faut veiller à maintenir le contrôle humain, l'implication et la responsabilité dans les décisions qui affecteront la vie des humains.

Résumé : Il est facile d'être d'accord avec le principe de l'EPSRC selon lequel les robots ne sont pas des agents responsables, mais même cette brève considération se révèle insuffisante pour orienter les actions futures.

Les robots programmés pour respecter la loi et pour respecter les droits et la liberté des individus ne comprendront pas les situations sociales et ne seront pas en mesure de prendre systématiquement les bonnes décisions morales concernant les situations sociales humaines. pris pour éviter ou minimiser la prise de décision automatique et algorithmique dans toutes les situations dans lesquelles le jugement humain est requis.

Les références

Arkin, R. (2009). Governing lethal behavior in autonomous robots. Chapman-Hall review. Ordinateurs et éducation, 58(3), 978–988.

Bekoff, M. et Pierce, J. (2009) Wild Justice : The Moral Lives of Animals. The University of Chicago Press, Londres.

Brosnan, S. F. and de Waal, F. B. (2003). Monkeys reject unequal pay. Nature, 425, 297–99

Churchland, P. (2011) *Braintrust: WhatNeurosciencetellsusaboutMorality*. Princeton University Press, Oxford.

Heyns,C.(2013).Rapportdurapporteurspéciauxurlesexécutionsextrajudiciaires,sommairesouarbitraires,A/HRC/23/47

Johnson,DG(2006).ComputerSystems:MoralEntitiesbutnotMoralAgents. *Éthique et technologies de l'information*, 8(4):195–204

Johnson, DG et Miller, KW (2008) Ne fabriquent pas d'agents moraux artificiels. *Éthiqueetinformation Technologie* (2008)10:123–133

Malle, BF, & Scheutz, M. (2014).

Maturana, HR & Varela, FJ (1980). *AutopoiesisandCognition-TheRealizationoftheLiving*. Dordrecht, Pays-Bas : D. Reidel Publishing

Riedl, MO, et Harrison, B. (2016) Utiliser des histoires pour enseigner les valeurs humaines aux agents artificiels.

Sharkey, AJC, & Sharkey, NE (2012). *Éthique et technologies de l'information*, 14(1),27–40.

Sharkey, NE, & Sharkey, AJC (2010). *InteractionStudies*, 11(2),161–190.

Sharkey, NE & Ziemke, T. (2001) .

Sparrow, R., & Sparrow, L. (2006).Inthehandsofmachines?Thefutureofagedcare. *Mindand Machine*, 16 141–161.

Winfield,AF,Blum,C.,etLiu,W.(2014)Towardsanethicalrobot:Internalmodels,consequencesandethicalactionselection.InM.Mistry,A.Leonardis,M.Witkowski,&C.Melhuish(Eds)
Avancéesdanslessystèmesrobotiquesautonomes : Actesdela15econférenceannuelle,TAROS2014 (pp85–96).Birmingham,Royaume-Uni,1–3septembre

Réflexions sur la confidentialité, la sécurité et la tromperie

Tom Sorell, Université de Warwick
Heather Draper, Université de Birmingham

Les cinq principes de Robotique formulés lors de la retraite AHRC-
EPSRC en 2010 ne sont pas le dernier mot de la robotique, mais l'un des premiers mots.

Confidentialité

Le principe 2 exige que les robots soient utilisés conformément aux lois en vigueur et aux libertés et droits fondamentaux, y compris le droit à la vie privée.

et 8), n'est pas fondamental dans les anciens traités sur les droits de l'homme, tels que le Pacte international relatif aux droits civils et

Les robots de soins en particulier et les robots sociaux en général sont souvent conçus pour collecter des informations sur les êtres humains avec lesquels ils interagissent.

les personnes dont les informations peuvent en principe accorder leur consentement à leur collecte et à leur stockage. Parce que l'

Cependant, le consentement ne règle pas nécessairement toutes les questions relatives à la bonne utilisation des informations personnelles.

Les implications de la seconde sont plus difficiles à prévoir et à consentir à l'avance que les
Il est même discutable qu'il n'y a pas de consentement correctement informé à la surveillance et au suivi continu d'un robot de soins vivant précisément parce qu'il n'est pas possible de prédire ou même d'imaginer à l'avance ce que serait l'expérience de vivre avec un robot.

Au lieu de déterminer les limites de la vie privée uniquement à partir de ce à quoi un utilisateur consent avec de bonnes informations, on peut aussi se fier à des arguments concernant les limites de la vie privée basés sur le

A robot conçu pour maintenir l'autonomie d'une personne âgée peut en partie être jugé par le qu'elle la personne âgée a autant latitude qu'un adulte non assisté pour décider de tous les aspects de sa vie, y compris la divulgation d'informations.

La tension entre l'autonomie (et la vie privée) et la sécurité

L'une des façons dont l'adulte standard exerce son autonomie est d'être son propre juge des risques à prendre.

Si d'autres sont mis en danger ou si on compte sur eux pour sauver dangereusement le preneur de risques autonome, alors il peut y avoir un argument contre la prise de risque à partir de l'autonomie en limitant les charges qu'elle impose aux autres.

Dans le cas des personnes âgées assistées, le briefing de conception du robot combinera généralement sécurité et l'autonomie.

Dans les cas les plus intéressants sous l'angle de la théorie morale, l'utilisateur est prêt à prendre un risque relativement faible – disons le risque de faire une chute mineure – pour continuer à mener sa vie quotidienne de la même manière qu'il le faisait lorsqu'il était plus jeune. tomber. C'est parce qu'il veut maintenir

Cependant, si l'objectif primordial est d'assurer la sécurité de l'utilisateur, il se peut qu'il n'ait pas ce droit. -techtelecare, une alarme dépendante est portée par un utilisateur, et lui permet de décider ou d'appeler de l'aide. La valeur d'autonomie supporte cette norme.

Les dérogations de l'utilisateur pourraient également être intégrées dans la conception de robots compagnons où les choix de style de vie d'un utilisateur, s'ils étaient signalés à des amis et à des relations, pourraient provoquer des interventions coercitives de la part de ces personnes.

personne; sinon une sorte d'âgeisme restreint l'autonomie des personnes assistées et il sont plus traités par les concepteurs

Personne non assistée. Et cette limitation est difficile à défendre sans âgisme.

Latensionentreautonomieetréadaptation

Les robots de soins et certains robots non sociaux sont conçus pour aider les personnes âgées à retrouver les capacités qu'ils ont perdues et pas seulement à exercer celles qu'ils ont de manière autonome. Au domicile de l'utilisateur, une obligation de coopération peut exister .

Que se passe-t-il si l'assistance d'un robot pour une personne âgée est fournie à condition qu'elle accepte d'opérer avec une réadaptation qui pourrait être proposée à l'avenir ? Dans ce cas, le refus autonome peut être annulé par une entreprise autonome de coopération.

quel'utilisateurconclut,quipréciselesresponsabilitésquel'utilisateurassumeenéchangepourlafournituredurobot.Untel
Bien entendu, les responsabilités en vertu du contrat n'excluent pas les droits.

Tromperie

Enfin, nous arrivons au principe 5. Il exige la transparence dans la conception des robots et interdit la tromperie des personnes vulnérables. La tromperie est la création intentionnelle de fausses croyances. La tromperie est généralement erronée parce que le trompeur veut manipuler la personne trompée pour faire quelque chose qui sert ses intérêts. n'a pas l'intention de tromper, mais sa conception enfantine est trompeuse ? Cela fonctionne-t-il en faisant croire à un enfant qu'un autre enfant l'aide ? un cas d'auto-tromperie, et la modélisation du robot éducatif sur un enfant humanoïde n'est pas non plus un cas de tromperie, non plus.

Il peut être nécessaire de traiter une représentation d'enfant.

Il ne s'agit pas d'un cas de tromperie de la part du fabricant de Paro, ni d'un cas d'auto-tromperie non plus.

Il n'est pas crucial de tirer du réconfort de Paro que l'on pense que c'est un vrai sceau ou que l'on pense qu'il est vivant - le fait que son comportement simule celui d'un apétissement est suffisant. être accessible par raisonnement parallèle pour les robots enseignants humanoïdes.

Commentaire pour l'atelier de l'ASIB sur les principes de la robotique

Emily C. Collins

L'Université de Sheffield, Sheffield, Royaume-Uni.

1. Introduction

Le principe suivant vise à réguler les robots dans le monde réel : Non. 4. Les robots sont des artefacts manufacturés. Ils ne doivent pas être conçus de manière trompeuse pour exploiter les utilisateurs vulnérables ; au lieu de cela, leur nature de machine devrait être transparente.

Ce commentaire proposera une critique de ce principe selon les critères suivants : a. Validité. Les

principes sont-ils corrects en tant qu'énoncés sur la nature des robots (par exemple, qu'ils sont des outils et des produits), les développeurs de robots et la relation entre les robots et les personnes (par exemple, que les robots doivent avoir une conception transparente), ou sont-ils ontologiquement défectueux ? , inexacts, obsolètes ou trompeuses.

La critique décomposera le principe en ce que je considère être ses deux principales déclarations : 1. Les robots

ne doivent pas être conçus de manière trompeuse pour exploiter les utilisateurs vulnérables.

2. La nature de la machine doit être transparente.

Je soutiendrai que les deux énoncés constitutifs de ce principe sont fondamentalement défectueux en raison de la nature indéfinie des termes critiques : « trompeur », « vulnérable » et « nature de la machine », et qu'en tant que tel, le principe dans son ensemble est trompeur.



Fig. 1. Panneau de gauche : « Ekso », abréviation d'exosquelette, est un robot portable qui aide les patients paralysés à marcher. Panneau de droite : Deux mammifères robots MIRO, un exemple de robot « social ».

Aux fins de ce commentaire, un robot est défini comme un fait manufacturé, en particulier un outil avec lequel un utilisateur humain peut augmenter un état existant, par exemple en fournissant à un individu qui ne peut pas marcher la capacité de marcher au moyen d'une aide mécanique, ou en offrant à l'utilisateur une forme avancée de divertissement, comme avec un robot compagnon (Figure 1). Ce commentaire portera en particulier sur le biomimétisme [1], les robots sociaux, et leur rôle en tant qu'outils de pari terme des utilisateurs. Un robot social est ici défini comme un appareil doté d'une certaine autonomie et d'une certaine présence physique, capable d'interagir socialement avec les gens, et en tant que tel, on peut s'attendre à ce qu'il suscite une sorte de réponse émotionnelle de la part de son utilisateur [2]. Voici notre premier problème, avant même d'aborder le principe : pour définir « robot », il faut au moins définir l'application du robot et l'étendue de ses capacités. Il existe des types de robots mutuellement exclusifs, qui ont la capacité de potentiellement tromper les utilisateurs de diverses manières distinctes en fonction de la manière dont chaque robot interagit. Les robots industriels, mobiles, de service, éducatifs, spatiaux, sociaux, pour n'en citer que quelques-uns, ont des morphologies différentes et répondent à des attentes différentes de leurs utilisateurs. Aucun des principes de la robotique ne commence par une définition de « robot », et j'ai donc défini la mienne.

2 Les robots ne doivent pas être conçus de manière trompeuse pour exploiter les utilisateurs vulnérables

Tout d'abord, commençons par nous demander ce qui est « trompeur » ? Dans ce contexte, c'est le robot qui est étiqueté comme trompeur, donc une meilleure question pourrait être de savoir comment le robot est trompeur de telle sorte qu'il irait à l'encontre de ce principe ?

Des robots sont développés pour ressembler à des êtres vivants. Ce que l'on sait de la dynamique homme-animal est utilisé pour intégrer des comportements et une morphologie de type animal dans la conception de robots sociaux. Par définition, le biomimétisme signifie la conception par la nature, via l'imitation des modèles, des systèmes et des éléments de la nature, dans le but de résoudre les problèmes humains. Les robots sont des outils, des produits à utiliser, dont le but est de résoudre des problèmes humains. Ici, les principes de conception mêmes qui sous-tendent la nature d'un robot biomimétique et social, et ce dont les développeurs de robots ont besoin, sont motivés par ce que l'on pourrait sans doute appeler "être trompeur": tenter d'imiter les êtres vivants pour le bien du robot et de son utilisateur.

Des robots ressemblant à des animaux, tels que Paro [3] et les "FurReal Friends Lulu Cuddlin Kitty", produits par Hasbro (Figure 2), sont utilisés par les thérapeutes d'une manière similaire à la thérapie assistée par l'animal (TAA) [4], dans lequel un animal peut être amené dans une séance de thérapie existante pour aider à la facilitation sociale (comme avec la thérapie de groupe), ou utilisé de manière individuelle pour aider à concentrer un client ou un patient pendant la thérapie. Ces robots ont un but précis, ressembler à des animaux et aider le thérapeute. Leur existence cependant, bien que basée sur un traitement impliquant un être vivant, n'est pas destinée à remplacer les animaux. Les animaux en AAT sont considérés comme des co-thérapeutes. Ils reçoivent le respect qu'une créature vivante devrait avoir et sont retirés des séances où le mal peut arriver



Fig. 2. Panneau de gauche : Le robot thérapeutique « Paro ». Panneau de droite : Les "FurReal Friends Lulu Cuddlin Kitty".

eux, ou dans lesquels ils sont eux-mêmes perturbateurs [5]. Cet exemple démontre qu'un robot conçu dans un esprit trompeur - pour ressembler à un animal - avec l'intention d'être utilisé par des populations vulnérables - les personnes en thérapie - sont n'est pas destiné à être exploité tel que défini en essayant de convaincre un utilisateur que le robot ressemblant à un animal est vraiment vivant. Ils sont plutôt utilisés pour déclencher des souvenirs d'autres êtres vivants. Il est difficile de faire passer cette idée, mais la nuance est importante. Ces robots ne sont pas construits pour être des animaux convaincants. Ils sont conçus pour être des outils robotiques convaincants et pour réaliser que les idées de la nature sont emprunté.

Deuxièmement, que signifie « exploiter les utilisateurs vulnérables » ? Qu'est-ce qu'un vulnérable utilisateur ? La vulnérabilité est-elle un état d'être unique ? Et si oui, à quel moment pourrait être considéré, ou ne plus être considéré comme vulnérable ? En effet, qui pourrait décider à quel moment un individu est devenu suffisamment vulnérable pour avoir son état de Un robot d'art qui leur a été enlevé ? ¹

En médecine, il existe une définition normalisée des groupes vulnérables, qui existent des domaines de vulnérabilité définis (par exemple, [7]). La manière dont un individu vulnérable est exploité par des robots trompeurs dépend de la vulnérabilité de l'individu ment. Par exemple, les domaines médicaux incluent la vulnérabilité économique. Considérez l'exploitation émotionnelle de la peur, créée par un média populiste qui propage la croyance que le travail d'une personne peut être menacé par des robots qui sont faussement dépeints comme plus avancés qu'eux. Cependant nous pouvons supposer que le principe ne fait pas référence à une telle vulnérabilité en tant qu'économique (bien qu'en vérité nous ne pouvons pas supposer que ; une partie du problème avec ces principes de la robotique est qu'ils ne sont pas du tout définis de cette manière, mais pour le Aux fins de ce commentaire, supposons que la vulnérabilité à laquelle il est fait référence est physique plutôt que conceptuel). Supposons donc peut-être que par « vulnérable », le principe fait référence à des groupes. Supposons également qu'un utilisateur général saura quand un robot est un robot à moins que ce robot ne soit si exceptionnellement réaliste

¹ Pour un exemple de cela dans la fiction, voir la première publication d'Issac Asimov, Robbie [6]. La peur des robots exploitant les personnes vulnérables est ancienne au sein du communauté de la robotique, mais il faut démêler la fiction de la réalité pour apprécier ce problème.

pour passer pour vivant. Pour passer pour vivant, le robot devrait, et cette liste n'est en aucun cas exhaustive, bouger, réagir, cligner des yeux, respirer et vocaliser de manière synchrone tout en étant morphologiquement exact. Une telle technologie n'existe pas. Ainsi, compte tenu de l'état de l'art qui existe actuellement, comme les robots sociaux qui font l'objet de ce commentaire, la question se pose du fait que ce sont précisément les plus vulnérables au sein d'une population qui ont le plus à gagner de leur utilisation. Les deux groupes les plus vulnérables sont généralement considérés comme les personnes âgées et les mineurs, et au sein de ces groupes les personnes souffrant de troubles cognitifs.

Aux fins de ce commentaire, concentrons-nous sur les groupes vulnérables au sein de la population âgée. Le robot Paro susmentionné est un robot interactif avancé conçu pour fournir un soutien physique et émotionnel aux malades et aux personnes âgées, non pas par lui-même, mais avec l'aide d'un praticien clinicien formé à la thérapie assistée par robot (RAT). Chez les personnes souffrant de démence et d'autres conditions de déclin cognitif, la capacité émotionnelle ne décline pas de manière univoque avec la cognition [8]. Cela permet une application significative de la thérapie psychologique et émotionnelle par un thérapeute avec des dispositifs statiques tels que Paro, qui est conçu pour ressembler à un être vivant, pour être tenu et pour être choyé [9]. Ici, la tromperie ressemble à celle observée avec la thérapie de poupée.

Dans les interventions de thérapie par poupées, des poupées qui ressemblent à des bébés réalistes sont utilisées par les soignants de la maladie d'Alzheimer pour tenter de soulager l'anxiété et apporter de la joie aux personnes atteintes de démence. Ceci est réalisé via l'introduction d'une activité utile et gratifiante, mais physiquement inoffensive : à savoir prendre soin de la poupée (par exemple, [10]). Bien que controversées [11], de telles thérapies qui introduisent des outils de point focal réalistes dans le processus de soins ont été saluées pour améliorer la qualité de vie (QoL) des patients, et ces études incluent celles qui ont exploré l'impact de l'utilisation de robots ressemblant à des animaux en thérapie. aussi [12].

La qualité de vie est une mesure complexe englobant les aspects émotionnels, sociaux et physiques de la vie d'un individu. Il existe sur un continuum, en dehors du domaine des « dichotomies soit/ou » où x est considéré comme mauvais et y bon. Si un outil robotique est utilisé auprès d'une population vulnérable qui a des capacités mentales pouvant être exploitées pour soulager la souffrance des individus au sein de cette population, la question de savoir si cet outil doit exister ou non devient vague et trop complexe pour y répondre. avec une seule déclaration. Le débat se résume à savoir dans quelle mesure nous devrions tromper les personnes vulnérables et à quel moment cela devient une exploitation dans un sens négatif. Lorsque cette considération est mise en regard des améliorations de la qualité de vie des individus souffrant de maladies neurodégénératives incurables, il devient clair que ce quatrième principe est insuffisant. Il est fondamentalement défectueux parce que ses termes composants ne sont pas définis. Sans savoir ce que l'on entend vraiment par exploiter les personnes vulnérables, tout le principe est trompeur.

Si la chose exploitée est le déclin cognitif lui-même et que le robot tire profit de la nature vulnérable de l'individu, mais dans le but délibéré d'améliorer la qualité de vie de cet individu, n'est-ce pas positif ? Lorsqu'il ya

n'y a-t-il pas d'autre alternative pour accéder aux restes du quotient émotionnel d'une personne atteinte de démence, quelqu'un qui pourrait autrement être terriblement déclenché par un animal vivant autrement réconfortant, où est le vrai mal ? Le mal réside-t-il dans l'esprit de ceux qui ne souffrent pas et ne sont pas témoins de ce qu'ils considèrent eux-mêmes comme un état triste ? Et si tel est le cas, ne faut-il pas d'autant plus chercher à se projeter dans l'esprit des plus vulnérables, et apprécier cette situation pour ce qu'elle est ? Une tentative de fournir des soins en utilisant tous les outils disponibles, édictée avec bonne volonté et supervisée par des soignants qui connaissent toute l'étendue des dommages causés par les maladies neurodégénératives, tant aux patients qu'à leurs proches qui les regardent.

3 La nature de la machine doit être transparente

Considérons la population saine qui observe les robots. Comme indiqué précédemment dans ce commentaire, je crois qu'il n'existe pas de technologie robotique telle qu'elle soit parfaitement trompeuse. Même les robots les plus sophistiqués sont clairement des robots. Un utilisateur peut croire que l'IA d'un robot est plus avancée qu'elle ne l'est à première vue, mais, au moins de manière anecdotique grâce à mes propres expériences en laboratoire, je pense que toute période de temps avec un robot est suffisante pour qu'un utilisateur établisse une relation approximative une approximation suffisante de ses limites pour que toute surestimation initiale des capacités du robot soit rapidement remplacée par la réalité. Quant aux populations suffisamment vulnérables pour être trompées en leur faisant croire qu'un robot est plus avancé ou plus "vivant" qu'il ne l'est, je pense que ce n'est pas le robot qui devrait être conçu différemment, mais que les utilisateurs humains ou les praticiens cliniques de la RAT qui doivent être formés pour utiliser leur outil, leur produit robot, de la manière la plus efficace et la plus positive.

4 Résumé

Un robot qui est si parfait qu'il trompe complètement un utilisateur en lui faisant croire qu'il est autre chose qu'une machine, est quelque chose que je n'imagine pas exister de sitôt. Pour les individus suffisamment vulnérables pour être convaincus qu'un robot, qui est de toute évidence une machine, est en fait vivant, ma recommandation est de considérer aussi objectivement et aussi largement que possible tous les avantages positifs qui peuvent découler d'une telle situation. Considérer ce que signifie réellement exploiter les personnes vulnérables, et peut-être reformuler un scénario avec des résultats ostensiblement positifs pour l'utilisateur vulnérable, sans utiliser le terme « exploiter », mais plutôt avec le mot « aider » :

Les robots sont des artefacts manufacturés, mais ce sont des outils qui nous aident et qui peuvent être conçus en utilisant des principes dont nous savons qu'ils fonctionnent, y compris des principes biomimétiques. Les robots conçus de manière trompeuse, destinés à soulager la souffrance d'utilisateurs vulnérables, devraient faire connaître leur nature de machine à la

soignants de ces usagers vulnérables. Puisse-t-il être de la responsabilité des soignants d'améliorer la qualité de vie de leurs patients par tous les moyens sûrs nécessaires.

Les références

1. TJ Prescott, MJ Pearson, B. Mitchinson, JCW Sullivan et AG Pipe, "Fouetter avec des robots des vibrisses de rat à la technologie biomimétique pour le toucher actif", IEEE Robotics and Automation Magazine, vol. 16, non. 3, p. 42–50, 2009.
2. EC Collins, A. Millings et TJ Prescott, « Attachment to assistive technology : A new conceptualisation », dans Actes de la 12e conférence européenne de l'AAATE (Association pour l'avancement des technologies d'assistance en Europe), 2013.
3. T. Shibata, "" Robot d'engagement mental (PARO) ". [En ligne], <http://www.paro.jp>.
4. MR Banks, LM Willoughby et WA Banks, "Thérapie assistée par les animaux et solitude dans les maisons de soins infirmiers : utilisation de chiens robotisés par rapport à des chiens vivants", Journal de l'American Medical Directors Association, vol. 9, non. 3, p. 173–177, 2008.
5. S. Brooks, « Psychothérapie assistée par les animaux et psychothérapie facilitée par les chevaux », Travailler avec des jeunes traumatisés en protection de l'enfance, pp. 196–218, 2006.
6. I. Asimov, « Étrange camarade de jeu », Super Science Stories, pp. 67-77, 1940.
7. BM Association et al., « Safeguarding Vulnerable Adults—a Tool Kit for General praticiens », 2011.
8. C. Magai, C. Cohen, D. Gomberg, C. Malatesta et C. Culver, « Expression émotionnelle au cours de la démence de stade intermédiaire à avancé », International Psychogeriatrics, vol. 8, non. 03, p. 383–395, 1996.
9. EC Collins, TJ Prescott et B. Mitchinson, « Le dire avec la lumière : une étude pilote de la communication affective à l'aide du robot miro », dans Biomimetic and Biohybrid Systems, pp. 243–255, Springer, 2015.
10. M. Ehrenfeld, « Utilisation de poupées thérapeutiques avec des patients psychogériatriques », Thérapie par le jeu avec des adultes, pp. 291–297, 2003.
11. G. Mitchell, « Utilisation de la thérapie par poupée pour les personnes atteintes de démence : un aperçu : Gary Mitchell présente les arguments pour et contre cette intervention controversée, mais populaire », Nursing Elderly People, vol. 26, non. 4, p. 24–26, 2014.
12. M. Heerink, J. Albo-Canals, M. Valenti-Soler, P. Martinez-Martin, J. Zondag, C. Smits et S. Anisuzzaman, « Explorer les besoins et les robots alternatifs pour animaux de compagnie pour la thérapie assistée par robot adultes atteints de démence », dans Social Robotics, pp. 104-115, Springer, 2013.

Pourquoi mon robot se comporte-t-il ainsi ?

Concevoir la transparence pour l'inspection en temps réel des robots autonomes

Andréas Théodorou¹ et Robert H. Wortham² et Joanna J. Bryson³

Abstrait. Les principes de robotique de l'EPSRC dictent la mise en œuvre de la transparence dans les systèmes robotiques, cependant, la recherche qui s'y rapporte n'en est qu'à ses balbutiements. L'article actuel présente au lecteur à la nécessité d'avoir des agents intelligents transparents à l'inspection. Nous fournissons une définition robuste de la transparence, en tant que mécanisme pour exposer la prise de décision du robot, en considérant et en élargissant sur d'autres définitions importantes trouvées dans la littérature. Le document conclut en abordant les décisions de conception potentielles que les développeurs doivent prendre en compte. prendre en compte lors de la conception de systèmes transparents.

1. INTRODUCTION

La transparence, à notre avis, est un élément clé relatif à l'éthique implications du développement et de l'utilisation de l'intelligence artificielle, un sujet d'intérêt et de débat de plus en plus public. Nous utilisons fréquemment techniques inspirées de la philosophie, des mathématiques et de la biologie pour construisant des agents artificiels interactifs et intelligents, mais nous les traitons comme boîtes noires sans comprendre comment le temps réel sous-jacent la prise de décision fonctionne.

La nature de la boîte noire des systèmes intelligents, comme dans les applications sensibles au contexte, rend l'interaction limitée et souvent non informative pour l'utilisateur final [14]. Limiter les interactions peut avoir un effet négatif les performances du système ou même compromettre la fonctionnalité du système. Imaginez un système robotique autonome conçu pour fournir l'accompagnement sanitaire des personnes âgées. Cependant, les personnes âgées peuvent avoir peur et se méfier du système. Ils peuvent ne pas permettre au robot de interagir avec eux. Dans un tel scénario, des vies humaines sont en danger, car elles peut ne pas recevoir le traitement médical requis à temps, car un humain surveillant le système doit détecter le manque d'interaction et intervenir. Inversement, si l'utilisateur humain accorde trop de confiance à un robot, il pourrait entraîner une mauvaise utilisation, une dépendance excessive et une désuétude du système [13]. Dans notre exemple du robot de soins de santé, si l'agent fonctionne mal et que ses patients ne sont pas conscients de son dysfonctionnement, les patients peuvent continuer utiliser le robot, au risque de leur propre santé. Les robots dans les deux scénarios brisent le premier principe de robotique d'EPSRC en mettant l'homme vit en danger [1].

Pour éviter de telles situations, un bon calibrage de la confiance entre les les opérateurs humains et leurs robots sont extrêmement importants, voire essentiels, en particulier dans les scénarios à haut risque tels que l'utilisation de robots dans l'armée ou à des fins médicales [9]. Le calibrage de la confiance se produit lorsque l'utilisateur final a un modèle mental du système et s'appuie sur le

système dans les limites des capacités du système et est conscient de ses limites [6].

Nous croyons que l'application de la transparence n'est pas seulement bénéfique pour les utilisateurs finaux, mais aussi pour les développeurs d'agents intelligents. Temps réel le débogage du mécanisme de prise de décision d'un robot pourrait aider les développeurs à corriger les bogues, à prévenir les problèmes et à expliquer les écarts potentiels dans les performances d'un robot. Nous envisageons que par la mise en œuvre correcte de la transparence, les développeurs pourraient concevoir, tester et déboguer leur agents en temps réel - de la même manière que les développeurs de logiciels travailler avec le développement et le débogage de logiciels traditionnels.

Malgré ces avantages possibles de la transparence dans les systèmes intelligents, il existe peu de recherches sur les agents transparents et même moins de mise en œuvre d'agents transparents. De plus, il y a des incohérences dans les définitions de la transparence et les critères de un robot pour être considéré comme un système transparent. Dans cet article, nous allons présenter les définitions incohérentes trouvées dans la littérature et tenter pour les compléter avec les nôtres. De plus, dans la troisième partie de cet article, nous discuterons des décisions de conception dont un développeur a besoin à considérer lors de la conception de systèmes robotiques transparents.

Nous utilisons spécifiquement le terme agent intelligent pour désigner la combinaison à la fois du logiciel et du matériel d'un robot autonome.

système, travailler ensemble en tant qu'acteur, vivre et changer le monde [3]. Dans cet article, les mots robot et agent sont utilisés de manière interchangeable.

2 DÉFINIR LA TRANSPARENCE

Malgré l'utilisation prédominante du mot-clé transparence dans le EPSRC Principles of Robotics, la recherche visant à rendre les systèmes transparents en est encore à ses balbutiements. Au fil des ans, très peu de publications se sont concentrées sur la nécessité de systèmes transparents et même moins ont tenté de répondre à ce besoin. Chaque étude donne son propre définition du mot-clé, sans exclure les autres. À ce jour, le concept de transparence s'est limité aux explications des comportements anormaux, à la fiabilité du système et aux tentatives de définition des fondements analytiques d'un système intelligent.

2.1 Le principe de transparence de l'EPSRC

Les principes de robotique de l'EPSRC considèrent la transparence comme l'un de ses principes clés, en définissant la transparence en robotique comme suit : « Les robots sont objets manufacturés. Ils ne doivent pas être conçus de manière trompeuse moyen d'exploiter les utilisateurs vulnérables ; au lieu de cela, leur nature de machine devrait être transparent. »

La définition de la transparence de l'EPSRC met l'accent sur le maintien de la utilisateur final conscient de la fabrication, mécanique et donc artificielle

¹ Université de Bath, Royaume-Uni, e-mail : a.theodorou@bath.ac.uk

² Université de Bath, Royaume-Uni, e-mail : rhwortham@bath.ac.uk

³ Université de Bath, Royaume-Uni, e-mail : jjbryson@bath.ac.uk

nature du robot. Cependant, la formulation utilisée permet de considérer même des informations indirectes, telles que la documentation technique en ligne, comme une méthodologie suffisante pour faire respecter la transparence[4]. Cela place le fardeau de la responsabilité avec l'utilisateur final. L'utilisateur devra trouver, lire et comprendre la documentation ou toute autre information fournie par le fabricant. Certains groupes d'utilisateurs, comme les personnes âgées ou utilisateurs non spécialisés, peuvent avoir des difficultés à comprendre les termes que l'on retrouve souvent dans les manuels techniques.

2.2 La transparence comme mécanisme de signalement fiabilité

L'une des premières publications définissait la transparence en termes de communication d'informations à l'utilisateur final, concernant la tendance du système aux erreurs dans un contexte donné [6]. Alors que l'interprétation du Dzindolet n'est qu'une partie de notre définition d'un système transparent, la l'étude présente des résultats intéressants sur l'importance de la transparence systèmes. L'étude a montré que fournir des commentaires supplémentaires aux utilisateurs concernant les défaillances du système, il a aidé les participants à faire confiance à le système. Les utilisateurs savaient que le système n'était pas fiable à 100%, mais ils ont pu calibrer leur confiance dans le système autonome dans l'expérience, car ils ont pris conscience du moment où ils pouvaient compter dessus et quand ne pas le faire. L'utilisation militaire des systèmes robotiques devient de plus en plus populaire, en particulier sous la forme de systèmes sans pilote. véhicules aériens (UAV) et la transparence des systèmes de combat est essentielle. Imaginez si un agent identifie un bâtiment civil comme terroriste et décide de prendre des mesures à son encontre. Qui est responsable? Le robot pour ne pas être fiable ? Ou le surveillant humain, qui a placé sa confiance dans les capteurs du système et le mécanisme de prise de décision ? Tandis que le Le principe EPSRC de la robotique considère l'opérateur humain comme responsable, les dommages causés sont irréversibles. Robots travaillant de manière autonome pour détecter et neutraliser les cibles, il faut avoir un comportement transparent [17]. Les humains devraient pouvoir calibrer leur confiance dans le système et en cas de combat, de scénarios médicaux ou autres où si un robot agit peu fiable peut blesser ou tuer des humains, la transparence comme mécanisme pour signaler que la fiabilité du système est fondamentale.

2.3 La transparence comme mécanisme pour exposer comportement inattendu

Des études ultérieures par Kim Hinds [11] et Stumpf et. al [14], concentré sur la fourniture de mécanismes de rétroaction aux utilisateurs concernant les comportement d'un agent intelligent. Dans leurs études, l'utilisateur était alerté uniquement lorsque l'agent considère que son comportement est anormal. Kim et l'étude de Hinds, fait intéressant, a montré qu'en augmentant l'autonomie l'importance de la transparence a également augmenté à mesure que la responsabilité déplacé de l'utilisateur vers le robot. Leurs résultats sont en accord avec [10] la recherche, qui, ensemble, démontrent que les humains sont plus susceptibles blâmer un robot pour les échecs que d'autres artefacts manufacturés et collègues de travail.

Être capable d'alerter l'utilisateur lorsque le robot se comporte de manière inattendue est essentiel pour atteindre la transparence. Dans les situations à haut risque, il pourrait aider à sauver des vies humaines ou des ressources précieuses en alertant un surveillant humain du système pour prendre le contrôle ou calibrer sa confiance respectivement. Cependant, dans l'implémentation de Kim et Hinds, le robot n'alertait l'utilisateur que lorsqu'il détectait qu'il se comportait de manière une manière inattendue. A notre avis, cette implémentation tente de corriger un boîte noire en utilisant une autre. Il n'y a aucune garantie que le robot ait un comportement inattendu sans qu'il soit au courant de son comportement atypique. La transparence doit être appliquée en temps réel en tant que mécanisme toujours actif

anisme, permettant à l'utilisateur de décider si le comportement de l'agent est considéré comme attendu ou inattendu.

2.4 La transparence comme mécanisme pour exposer prise de décision

Nous croyons que les mécanismes de transparence devraient être intégrés au système, fournissant des informations en temps réel sur son fonctionnement, ainsi que ainsi que la fourniture de documents supplémentaires conformément au principe actuel du PE SRC. L'agent intelligent, c'est-à-dire un robot, doit contenir les mécanismes nécessaires pour fournir des informations significatives à l'utilisateur final. Pour considérer un robot transparent à l'inspection, le l'utilisateur final doit avoir la possibilité de demander des interprétations précises de les capacités du robot, ses objectifs, sa progression par rapport auxdits objectifs, entrées sensorielles - conscience de la situation, sa fiabilité et son imprévu comportement, comme les messages d'erreur. Les informations fournies par le robot doit être présenté dans un format compréhensible par l'homme. Un agent transparent, avec un mécanisme de prise de décision inspectable, pourrait également être débogué de la même manière que dans quels logiciels traditionnels non intelligents sont généralement débogués. Le développeur peut voir quelles actions l'agent effectue, pourquoi et comment il passe d'une action à l'autre. Ceci est similaire à la manière dans lequel les environnements de développement intégrés (IDE) populaires fournissent des options pour suivre différents flux de code avec des points de débogage, et ont des capacités telles que "Step-up" et "Step-in" sur des blocs de code.

3 CONCEVOIR DES SYSTÈMES TRANSPARENTS

Dans cette section de cet article, nous discuterons des différentes décisions auxquels les développeurs peuvent être confrontés lors de la conception d'un système transparent. Jusqu'à maintenant, des recherches de premier plan dans le domaine de la conception de systèmes transparents axés sur la présentation de la transparence uniquement dans le contexte de la collaboration homme-robot (HRC). Ainsi, ils se sont concentrés sur la conception de systèmes transparents capables d'établir la confiance entre les participants humains et le robot.[12]. Nous pensons que la transparence doit être présente même dans des environnements non collaboratifs, tels que des compétitions homme-robot [11] ou même lorsque des robots sont utilisés par les militaires. Dans notre vue, les développeurs devraient s'efforcer de développer des agents intelligents, qui peut communiquer efficacement des informations à l'utilisateur final humain, et séquentiellement lui permettre de développer un modèle mental du système et son comportement.

3.1 Convivialité

Afin d'assurer la transparence, des affichages supplémentaires ou d'autres méthodes de communication avec l'utilisateur final doivent être soigneusement conçus, car ils intégreront des informations potentiellement complexes. Agent les développeurs doivent tenir compte à la fois de la pertinence réelle et du niveau de abstraction des informations qu'ils exposent et comment ils vont présenter ces informations.

3.1.1 Pertinence des informations

Différents utilisateurs peuvent réagir différemment aux informations exposées par le robot. [16] démontre que les utilisateurs finaux sans formation technique ne comprennent ni ne retiennent les informations provenant d'inputs techniques tels que les capteurs. Ceci est contraire au développeur de l'agent, qui a besoin d'accéder à ces informations pendant le développement et les tests du robot pour calibrer efficacement les capteurs et résoudre les problèmes détectés. Cependant, dans la même étude, Tullio démontre que

les utilisateurs sont capables de comprendre au moins les concepts de base de l'apprentissage automatique, quelle que soit leur formation non technique et leurs antécédents professionnels.

Les recherches de Tullio établissent un bon point de départ pour comprendre quelles informations peuvent être pertinentes pour l'utilisateur pour l'aider à comprendre les systèmes intelligents. Néanmoins, des travaux supplémentaires sont nécessaires dans d'autres domaines d'application pour établir à la fois des tendances spécifiques au domaine et spécifiques à l'utilisateur concernant les informations à prendre en compte importance.

3.1.2 Extraction des informations

Les développeurs de systèmes transparents devront s'interroger non seulement lesquels, mais aussi la quantité d'informations qu'ils exposeront à l'utilisateur en établissant un niveau de complexité avec lequel les utilisateurs peuvent interagir avec les informations relatives à la transparence. Ceci est particulièrement important dans les systèmes multi-robots.

Les systèmes multi-robots permettent l'utilisation de plusieurs, généralement de petite taille robots, où un objectif est partagé entre plusieurs robots, chacun avec son propre apport sensoriel, sa fiabilité et sa progression vers l'exécution de ses tâche assignée pour l'ensemble du système à accomplir. Les développements récents de l'intelligence des essaims inspirée par la biologie permettent l'utilisation de grands quantités de minuscules robots travaillant ensemble dans un tel système multi-robot [15]. L'armée envisage déjà le développement de essaims de petits soldats robotiques autonomes. La mise en œuvre de la transparence dans un tel système n'est pas une tâche triviale. Le développeur doit faire des choix rationnels quant au moment où des informations de bas ou de haut niveau doivent être exposées. En exposant toutes les informations à tout moment, par tous les types d'utilisateurs, le système peut devenir inutilisable car l'utilisateur être surchargé d'informations. Nous pensons que différents utilisateurs nécessitent différents niveaux d'abstraction de l'information pour éviter l'infobésité. Des niveaux d'abstraction plus élevés pourraient se concentrer sur la présentation seulement un aperçu du système. Au lieu d'avoir la progression d'un système vers un objectif, en montrant les actions en cours, le système est prise par rapport à la réalisation dudit objectif, il pourrait simplement présenter une barre d'achèvement. De plus, dans un système multi-robot, les informations de niveau inférieur pourraient également inclure l'objectif, le capteur, l'objectif-processus et l'ensemble. comportement des agents individuels de manière détaillée. A l'inverse, un Une vue d'ensemble de haut niveau pourrait afficher tous les robots comme une seule entité, indiquant les moyennes de chaque machine. Agents intelligents avec une conception basée sur une architecture cognitive, comme le Behavior Oriented Design (BOD) [2], ne pourrait présenter que des éléments de plan de haut niveau si une vue d'ensemble système est nécessaire. Dans le cas d'un agent conçu avec BOD, les utilisateurs peut préférer voir et être informé de l'état des disques ou Compétences mais pas Actions individuelles. D'autres utilisateurs peuvent souhaiter pour ne voir que des parties du plan en détail et d'autres parties à un niveau élevé aperçu.

Une bonne mise en œuvre de la transparence devrait fournir à l'utilisateur avec de telles options, offrant aux individus ou aux groupes d'utilisateurs potentiels avec des configurations flexibles et prédéfinies afin de répondre à un large éventail des besoins des utilisateurs potentiels. Nous émettons l'hypothèse que le niveau d'abstraction dont un individu a besoin dépend d'un certain nombre de facteurs, y compris, mais sans s'y limiter, le contexte démographique de l'utilisateur.

1. Utilisateur : Nous avons déjà discuté de la manière dont les différents utilisateurs ont tendance à réagir différemment aux informations concernant l'état actuel d'un robot. De même, on peut s'attendre à ce que divers utilisateurs répondent de manière similaire aux différents niveaux d'abstraction basés sur leur utilisation du système. Les utilisateurs finaux, en particulier les non-spécialistes, préféreront une vue d'ensemble de haut niveau des informations disponibles, tandis que nous nous attendons à ce que les développeurs s'attendent à avoir accès à un niveau d'information inférieur.

2. Type de système robotique : Comme indiqué dans nos exemples ci-dessus, un système multi-robot est le plus susceptible de nécessiter un niveau d'abstraction plus élevé, pour éviter l'infobésité de l'utilisateur final. Un système avec un seul agent nécessiterait beaucoup moins d'abstraction, car moins de données sont affichées à son utilisateur.
3. Objectif du système robotique : L'objectif prévu du système doit être pris en compte lors de la conception d'un agent transparent. Par exemple, un robot militaire est beaucoup plus susceptible d'être utilisé avec un utilisateur professionnel dans ou sur la boucle et en raison de son opération à haut risque, il est beaucoup plus nécessaire d'afficher et de capturer autant d'informations sur le comportement de l'agent que possible. De l'autre main, une réceptionniste robotique ou un assistant personnel est plus susceptible d'être utilisé par des utilisateurs non techniques, qui peuvent préférer une version simplifiée aperçu du comportement du robot.

3.1.3 Présentation des informations

Les développeurs doivent réfléchir à la façon de présenter à l'utilisateur l'un des des informations supplémentaires concernant le comportement de l'agent qu'ils exposer. Des études antérieures utilisaient une représentation visuelle ou audio de la information. A notre connaissance, il n'existe pas d'études antérieures comparant les différentes approches. Les systèmes robotiques autonomes peuvent prendre des dizaines de décisions différentes par seconde. Si l'agent utilise un plan réactif, tel qu'un plan POSH [5], l'agent peut effectuer des milliers d'appels par minute vers les différents éléments du plan. Cette quantité d'informations est difficile à gérer avec systèmes orientés audio. De plus, la visualisation des informations, c'est-à-dire en fournissant une représentation graphique du plan de l'agent où les différents éléments du plan clignotent au fur et à mesure qu'ils sont appelés, devraient système explicite et facile à suivre par les utilisateurs moins techniques. Enfin, une visualisation graphique comme moyen de fournir des informations liées à la transparence présente des avantages supplémentaires dans le débogage de l'application. Le développeur doit être en mesure de suivre une trace des différents éléments du plan appelés, en visualisant l'entrée sensorielle qui a déclenché jusqu'à ce qu'un élément spécifique soit utilisé.

3.2 Utilité du système

Jusqu'à présent, dans cet article, nous avons développé l'importance et la conception choix concernant la mise en œuvre de la transparence. Cependant, nous Je pense que le développeur doit également déterminer si la mise en œuvre la transparence peut en fait nuire à l'utilité d'un système. [18] argumente que l'utilité d'un agent est mesurée par le degré auquel il est de confiance. Accroître la transparence peut réduire son utilité. Cela pourrait, par exemple, avoir un effet négatif pour un robot de compagnie ou un robot de soins, conçu pour aider les enfants. Dans de tels cas, le système est conçu selon les principes EPSRC de la robotique, car il exploite le ressenti de ses utilisateurs pour augmenter son utilité et ses performances sur sa tâche assignée.

Une autre décision de conception importante qui affecte le système est la transparence physique du système. L'apparence physique d'un l'agent peut augmenter sa facilité d'utilisation [7], mais aussi il peut contraster avec transparence en masquant sa nature mécanique. De retour dans notre exemple de robot de compagnie, un robot humanoïde ou ressemblant à un animal peut être préféré à un agent où ses mécanismes et ses composants internes sont exposés, révélant son caractère manufacturé [8]. Discuter des compromis entre utilité et transparence est loin au-delà de la portée de cet article. Cependant, les développeurs doivent être conscients de cela lorsqu'ils conçoivent et développent des robots.

4. CONCLUSION

Nous croyons fermement que la mise en œuvre et l'utilisation d'intelligence Les systèmes qui sont de nature transparente peuvent aider le public à comprendre l'IA en supprimant le mystère effrayant autour de pourquoi est-ce qu'il a comme ça. La transparence permettra de comprendre un agent comportement émergent. Dans cet article, nous avons redéfini la transparence comme un mécanisme permanent capable de signaler le comportement, la fiabilité, sens et objectifs, car ces informations pourraient nous aider à comprendre comportement du système autonome.

Des travaux supplémentaires sont nécessaires pour tester et établir de bonnes pratiques concernant la mise en œuvre de la transparence au sein de la communauté robotique. Compte tenu des avantages des systèmes transparents, nous suggèrent la promotion de ce principe clé par les conseils de recherche, tels comme EPSRC, et d'autres communautés universitaires.

REMERCIEMENTS

Nous voudrions remercier Swen Gaudl (Université de Bath) pour ses plus Des idées précieuses.

LES RÉFÉRENCES

- [1] Margaret Boden, Joanna Bryson, Darwin Caldwell, Kerstin Dauten hahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorell, Mick Wallis, Blay Whitby, et Alan Winfield. Principes de la robotique. Engi neering and Physical Sciences Research Council (EPSRC) du Royaume-Uni, avril 2011. publication Web.
- [2] Joanna Bryson, « La conception orientée comportement de l'intelligence modulaire des agents », dans System, volume 2592, 61–76, (2002).
- [3] Joanna J. Bryson, "Les robots devraient être des esclaves", dans Close Engagements avec des compagnons artificiels : problèmes sociaux, psychologiques, éthiques et de conception clés, éd., Yorick Wilks, 63–74, John Benjamins, Amsterdam, (mars 2010).
- [4] Joanna J. Bryson, Darwin Caldwell, Kerstin Dautenhahn, Paula Duxbury, Lilian Edwards, Hazel Grian, Sarah Kember, Stephen Kemp, Paul Newman, Geo Peg, Andrew Rose, Tom Rodden, Tom Sorell, Mick Wallis, Shearer West, Alan Winfield et Ian Baldwin, « The making of les principes epsrc de la robotique », 133(133), 14–15, (2012).
- [5] Joanna J. Bryson, Tristan J. Caulfield et Jan Drugowitsch, « Intégration de la sélection d'actions réalistes dans les environnements de simulation d'agents basés sur le cycle », dans Actes de l'agent 2005 : Processus sociaux génératifs, Modèles et mécanismes, éd., Michael North, David L. Sallach et Charles Macal, pp. 67–81, Chicago, (octobre 2005). Nationale d'Argonne Laboratoire.
- [6] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce et Hall P. Beck, "Le rôle de la confiance dans la dépendance à l'automatisation", Journal international d'études informatiques humaines, 58 (6), 697–718, (2003).
- [7] Kerstin Fischer, 'Comment les gens parlent avec les robots : Concevoir le dialogue pour réduire l'incertitude de l'utilisateur », AI Magazine, 32(4), 31–38, (2011).
- [8] Jennifer Goetz, Sara Kiesler et Aaron Powers, « Faire correspondre l'apparence et le comportement du robot aux tâches pour améliorer la coopération homme-robot », Actes - IEEE International Workshop on Robot and Human Interactive Communication, 55–60, (2003).
- [9] Victoria Groom et Clifford Nass, 'Les robots peuvent-ils être coéquipiers ?', Inter Action Studies, 8(3), 483–500, (2007).
- [10] Peter H. Kahn, Rachel L. Severson, Takayuki Kanda, Hiroshi Ishiguro, Brian T. Gill, Jolina H. Ruckert, Solace Shen, Heather E. Gary, Aimee L. Reichert et Nathan G. Freier, 'Les gens détiennent-ils un humanoïde robot moralement responsable du mal qu'il cause ? », Actes de la septième conférence internationale annuelle ACM/IEEE sur l'homme-robot Interaction - HRI '12, (février 2016), 33, (2012).
- [11] Taemie Kim et Pamela Hinds, 'Qui dois-je blâmer ? effets de l'autonomie et de la transparence sur les attributions dans l'interaction homme-robot », Actes - IEEE International Workshop on Robot and Human Interactive Communication, 80–85, (2006).
- [12] Joseph B Lyons, « Être transparent sur la transparence : un modèle pour interaction homme-robot », Trust and Autonomous Systems: Papers from le Symposium du printemps AAAI 2013, 48–53, (2013).
- [13] R Parasuraman et V Riley, 'Humans and automation: Use, misuse, désuétude, abus », Human Factors, 39(2), 230–253, (1997).
- [14] Simone Stumpf, Weng-keen Wong, Margaret Burnett et Todd Kulesza, 'Rendre les systèmes intelligents compréhensibles et contrôlables par les utilisateurs finaux », 10–11, (2010).
- [15] Ying Tan et Zhong-yang Zheng, « Research advance in swarm robotics', Defense Technology, 9(1), 18–39, (3 2013).
- [16] Joe Tullio, Anind K. Dey, Jason Chalecki et James Fogarty, « How ça marche : une étude de terrain d'utilisateurs non techniques interagissant avec un système intelligent », conférence SIGCHI sur les facteurs humains en informatique systèmes (CHI'07), 31–40, (2009).
- [17] Lu Wang, Greg a Jamieson et Justin G Hollands, 'Trust and dependence sur un système automatisé d'identification au combat », Facteurs humains, 51(3), 281–291, (2009).
- [18] Robert Wortham, Andreas Theodorou et Joanna J. Bryson, « The iron triangle : transparence-confiance-utilité". soumis, 2016.

Transparence, confiance et utilité du robot

Robert H. Wortham¹, Andreas Theodorou² et Joanna J. Bryson³

Abstrait. À mesure que le raisonnement des robots devient plus complexe, le débogage devient de plus en plus difficile en se basant uniquement sur un comportement observable, même pour les concepteurs de robots et les spécialistes techniques. De même, les utilisateurs non spécialistes ont du mal à créer des modèles mentaux utiles de raisonnement robotique uniquement à partir du comportement observé. Les principes de robotique de l'EPSRC exigent que nos artefacts soient transparents, mais qu'est-ce que cela signifie dans la pratique, et comment la transparence affecte-t-elle à la fois la confiance et l'utilité ? Nous étudions cette relation dans la littérature et la trouvons complexe, en particulier dans les environnements non industriels où la transparence peut avoir un plus large éventail d'effets sur la confiance et l'utilité en fonction de l'application et de l'objectif du robot. Nous décrivons notre programme de recherche pour étayer notre affirmation selon laquelle il est néanmoins possible de créer des agents transparents qui sont émotionnellement engageants malgré leur nature de machine transparente.

1. INTRODUCTION

Les principes de robotique de l'EPSRC incluent une référence spécifique à la transparence : « Les robots sont des artefacts manufacturés. Ils ne doivent pas être conçus de manière trompeuse pour exploiter les utilisateurs vulnérables ; au lieu de cela, leur nature de machine devrait être transparente. voir [1]. Cela apparaît à première vue comme une simple affirmation normative, s'appuyant sur l'idée communément admise selon laquelle les agents ne doivent pas être trompeurs, puisque la tromperie conduit généralement à l'exploitation. Cet article examine si en fait la transparence est vraiment une idée aussi simple, et aussi si rendre certains types d'agents transparents réduit leur utilité. En examinant cette question, nous devons également aborder la relation entre la transparence et la confiance.

Dans cet article, nous utilisons les termes robot et agent de manière interchangeable et par ces termes, nous entendons un artefact intelligent incarné et autonome.

Que signifie faire confiance à un robot ? Nous pourrions d'abord dire simplement que si une IA est plus transparente, alors nous sommes capables de lui faire davantage confiance, et donc son utilité augmente. Nous pourrions également soutenir que la confiance n'est requise que lorsqu'un agent n'est pas totalement transparent, et donc qu'une transparence accrue réduit le besoin de confiance [4]. Si l'utilité d'un artefact est mesurée par le degré auquel on lui fait confiance, alors l'augmentation de la transparence peut réduire cette utilité. Cela pourrait, par exemple, être le cas pour un robot dont la fonction principale est de fournir de la compagnie.

Ainsi, nous commençons à voir qu'il existe une relation complexe entre les notions d'utilité, de transparence et de confiance. Cette relation dépendra de l'objectif de l'IA. Dans cet article, nous passons en revue la littérature relative à la transparence et à la confiance, et nous décrivons également les recherches pratiques en cours pour étudier la proposition selon laquelle il est en effet possible de construire un robot émotionnellement engageant mais transparent.

2 THÉORIE DE L'ESPRIT, CONFIANCE ET TRANSPARENCE

Si l'on peut supposer que la communication entre animaux, et particulièrement entre humains, doit être complexe, en réalité les systèmes de communication naturels tendent à exploiter des signaux relativement simples et minimaux, dont la signification découle de modèles extensifs [16]. En d'autres termes, l'évolution, ou une histoire phylogénétique partagée, fournit des a priori adéquats tels qu'un minimum de données est nécessaire pour communiquer le contexte. Bien que certains disent le contraire [8], il est généralement admis qu'une interaction efficace, qu'il s'agisse de coercition ou de coopération, repose sur le fait que chaque partie a une théorie de l'esprit (ToM) de l'autre [16, 14]. Les actions individuelles et les comportements composites sont ainsi interprétés dans un cadre ToM préexistant. Si cette ToM soit exacte n'a pas d'importance, à condition qu'elle soit prédictive en termes de comportement. Le modèle de transparence du robot ne définit pas la ToM employée par l'utilisateur humain, mais c'est le modèle de transparence que nous pouvons directement ajuster et c'est donc l'objet de cet article. Il est bien connu que le comportement observable peut communiquer les états mentaux internes de l'individu. Breazeal [2] a découvert que la communication non verbale implicite améliore la transparence par rapport à la seule communication non verbale délibérée. Ici implicite est défini comme véhiculant des informations inhérentes au comportement mais qui ne sont pas délibérément communiquées par le concepteur du robot. Les gens ont de fortes attentes quant à la façon dont les signaux non verbaux implicites et explicites correspondent aux états mentaux. Breazeal a également constaté que la transparence réduit les conflits lorsque des erreurs se produisent, en particulier lorsqu'une tâche conjointe est tentée.

La réduction des conflits implique que lorsqu'une erreur se produit lors de l'exécution d'une tâche, la récupération est toujours possible avec moins de responsabilité. Breazeal appelle ce conflit réduit la robustesse, et cette robustesse est une mesure efficace de l'utilité.

2.1 Anthropomorphisme et modèles mentaux des robots

Les humains ont une forte prédisposition à anthropomorphiser non seulement la nature, mais tout ce qui les entoure [5] — l'hypothèse du cerveau social [7] peut expliquer ce phénomène, mais les humains ne traitent pas les robots de la même manière que les humains, par exemple en ce qui concerne la position morale [dix]. Bien qu'il y ait un débat important sur l'ontologie des esprits robotiques par rapport aux esprits humains, ce qui est plus important sur le plan pratique, c'est la façon dont les esprits robotiques sont compris psychologiquement par les humains, c'est-à-dire quelle est l'ontologie perçue plutôt que réelle. Stubbs [15] considère qu'il est essentiel de former un modèle mental des robots afin de construire un terrain d'entente - que nous pourrions également interpréter comme la base de la confiance humaine. Stubbs [15] a également constaté que ce terrain d'entente peut être efficacement établi via un dialogue interactif avec le robot. Bien que cette étude ait principalement considéré des robots distants travaillant dans un environnement industriel ou exploratoire, plutôt que des robots opérant dans

¹ Université de Bath, Royaume-Uni, e-mail : rhwortham@bath.ac.uk

² Université de Bath, Royaume-Uni, e-mail : a.theodorou@bath.ac.uk

³ Université de Bath, Royaume-Uni, e-mail : jjbryson@bath.ac.uk

environnements domestiques, nous devons prendre note de l'importance du dialogue dans l'établissement de la confiance. En effet, Mueller [13] considère le dialogue comme un des trois principales caractéristiques des ordinateurs transparents, les autres être explication et apprentissage.

Meerbeek [12] étudie la relation entre la personnalité perçue d'un robot et le niveau auquel l'utilisateur se sent en contrôle pendant l'interaction. Pour être crédible, Meerbeek a trouvé que l'expression de la personnalité doit être liée à un modèle interne qui traite du comportement (par exemple, prise de décision) en fonction de la personnalité et émotion. Un comportement plus expressif et informel est associé à une perception plus élevée du contrôle de l'utilisateur.

Les humains non spécialistes ont soit peu de ToM pour les robots, soit un modèle basé sur la science-fiction contemporaine, et donc interpréter les comportements en utilisant une théorie des autres agents par défaut, qui suppose l'agent de partager des motivations humaines. Cela peut se comprendre en termes d'évolution par le besoin de nos ancêtres de catégoriser rapidement l'activité proximale comme neutre (le bruissement des feuilles dans le vent), amical (l'approche d'un membre de la tribu) ou hostile (le l'approche d'un prédateur ou d'un ennemi). Lorsque l'information sensorielle est incertaine, l'évolution d'un biais vers une hypothèse à la fois d'agentivité et d'hostilité est sélective pour la longévité individuelle dans un environnement où on est souvent la proie, pas le prédateur. Même dans nos environnements technologiques, nous rencontrons souvent de fausses agences, telles que la robotique numérotation des appels de vente, publications automatisées sur Twitter et génération automatique spams personnalisés.

Dans une étude menée en 2006 dans un hôpital communautaire du États-Unis, le personnel infirmier cherchait constamment les raisons pour lesquelles les robots ont agi comme ils l'ont fait. Ils se demanderaient à eux-mêmes et aux autres, "Qu'est-ce qui se passe ici? Est-ce que le robot est censé faire ça ou ai-je fait Quelque chose ne va pas?". Cette recherche affirme que de faibles niveaux de transparence ont amené les gens à remettre en question même les comportements normaux des robot, amenant parfois même les gens à penser à des comportements corrects comme des erreurs [11].

3 PROGRAMME DE RECHERCHE

Nous commençons un programme de recherche pratique pour étudier le triangle transparence, confiance, utilité. Utilisant initialement des non-humanoïdes robots, nous menons des expériences pour déterminer l'effet de diverses expressions de transparence sur la réponse émotionnelle des humains. Au cœur de nos expérimentations nous utilisons des techniques de planification réactives pour construire des agents autonomes. Nous avons développé le Planificateur réactif Instinct basé sur l'approche Behavior Oriented Design (BOD) de Bryson [3]. Le planificateur Instinct rapporte l'exécution et le statut de chaque élément du plan en temps réel, nous permettant de capturer implicitement le processus de raisonnement au sein du robot qui donne lieu à son comportement. Nos expériences étudieront et démontreront comment ces données de transparence du planificateur peuvent être utilisées pour rendre le comportement du robot plus compréhensible. Au départ, nous sommes principalement intéressé à rendre le comportement transparent pour le concepteur du robot, car les robots avec des plans complexes sont généralement très difficiles à concevoir et déboguer. Cependant, ces premières expériences peuvent également améliorer la transparence pour les observateurs non spécialistes.

Nous étudierons ensuite comment nous pouvons exploiter le mécanisme de transparence intégré à l'Instinct Planner pour produire un robot domestique plus efficace. La recherche examinera si la transparence fait que les gens se sentent plus ou moins liés à leur robot, et s'ils sont plus ou moins capables d'évaluer avec précision les besoins du robot, car il travaille pour atteindre ses objectifs.

Il est prévu que ces essais se déroulent dans un environnement domestique ou quasi domestique, comme une maison de retraite.

Nous devons recueillir les commentaires des observateurs/utilisateurs non spécialistes sur le niveau qualitatif d'intelligence du robot, et aussi sur la façon dont ils seraient à l'aise d'avoir un tel appareil dans leur environnement domestique. La recherche tentera d'évaluer les niveaux initiaux de peur, anxiété, méfiance vis-à-vis de l'IA et des robots en général, et des robots domestiques en particulier. Ayant établi un positionnement de référence, la transparence du robot doit être activé en fournissant un retour d'informations à l'utilisateur sur l'exécution en temps réel dans le planificateur réactif. Les méthodes nous envisageons actuellement sont :

- Présentation en temps réel des déclarations textuelles relatives au plan exécuté.
- Visualisation graphique en temps réel de l'exécution du plan.
- Déclarations audio (c'est-à-dire verbales) relatives à l'exécution du plan du robot.

Pour chacune de ces méthodes, les informations sur la transparence pourraient soit être présenté sur/ depuis un appareil distant, ou sur/ depuis le robot lui-même. Il y a donc six combinaisons possibles. Bien sûr, une fusion de transparence supplémentaire, telle que l'audio combiné avec le graphique, pourrait également être testé en fonction du succès ou de l'échec des résultats expérimentaux initiaux.

Comme la littérature indique que le dialogue est important pour établir confiance, cette recherche devrait envisager la possibilité d'accepter la saisie vocale, bien que limitée à de simples commandes, comme un signifie pour les utilisateurs de demander au robot ce qu'il fait, et d'avoir le robot répond de manière appropriée.

4. DISCUSSION

Le principe 1 de l'EPSRC affirme que les robots sont des outils. Au sein de l'industrie et environnements d'ingénierie, cela est assez clair, dans le sens où un humain utilise le robot pour accomplir une tâche technique. Le concepteur et l'utilisateur du robot partage l'objectif du robot : accomplir la tâche. Cependant, dans les environnements domestiques et de soins de santé, les robots peuvent ont plutôt une relation différente avec ceux avec qui ils interagissent. Ils peut être destiné à fournir une compagnie et une couverture simultanée suivi du bien-être des patients. Ce sont peut-être des outils pour le professionnel de la santé, mais pour le patient, ce sont des compagnons. Dans un tel un environnement dans lequel le service public peut être affecté négativement par transparence. Notre sens du compagnonnage est lié à la mesure d'agence que nous projetons sur le robot. Si nous sommes capables de comprendre le fonctionnement de l'intelligence semble-t-il intrinsèquement devenir moins intelligent au sens populaire, de sorte que nous projetons alors moins d'agence, et par conséquent, bénéficiez-vous moins du robot ? On pourrait comparer ceci avec la télévision. Nous savons qu'elle n'a pas d'agence, mais sa présence dans le coin de notre salon offre un compagnon comme des avantages. Cela a peut-être à voir avec la suspension consciente de l'incrédulité, ou peut-être avons-nous un détecteur d'agence inconscient qui est plus facilement trompé par la technologie.

Les notions de sens commun de l'intelligence sont confondues avec les idées de la psychologie populaire sur l'agentivité et aussi sur la vie. Des choses intelligentes sont vivants, dans le sens où ils ont leurs propres croyances, désirs et intentions que nous comprenons sont fondamentalement égoïstes ou égoïstes. Nous reconnaissons implicitement l'égoïsme comme une caractéristique fondamentale de toute vie [6]. Si un tel agent s'engage avec nous, il nous considère alors comme important dans la poursuite de ces objectifs égoïstes. De tels agents sont dignes de devenir nos compagnons car ils accordent une vraie valeur dans leur relation avec nous, ce qui augmente notre valeur dans la société. Inversement, les agents qui n'ont pas d'agence intéressée ne sont pas dignes de notre attention car ils ne véhiculent aucune valeur sociale. Peut-être donc, des agents artificiels dont le seul but est la compagnie et sont vraiment transparents à cet égard sont ainsi disqualifiés pour être de dignes compagnons. Dans certaines situations, la transparence du robot peut donc

être en contradiction avec l'utilité, et plus généralement il peut être orthogonal plutôt que bénéfique pour l'utilisation réussie du robot. Bien que nous puissions exposer des scénarios et continuer à discuter de l'interaction théorique et philosophique entre la transparence, la confiance et l'utilité, en tant que scientifiques, nous attendre le résultat de nos expériences.

5. CONCLUSION

Nous avons vu que débattre la transparence et la confiance est complexe, mais peut être en partie comprise en examinant comment les humains arrivent à se comprendre et ensuite à se faire confiance, et comment ils surmontent peurs évolutives afin de faire confiance aux autres agents, par des la communication non verbale. Des niveaux inacceptables d'anxiété, de peur et de la méfiance se traduira par une réponse émotionnelle et cognitive de rejet robots. Hancock [9] affirme que si nous ne pouvons pas faire confiance à nos robots, nous ne pas pouvoir en bénéficier efficacement. Cependant, étant donné que nous interagir joyeusement dans la société avec d'autres que nous ne connaissons pas complètement confiance, et de plus en plus nous interagissons avec les ordinateurs en sachant que leur recommandations peut-être erronées, nous devons conclure que Hancock est trop simplifier. Enfin, il peut y avoir des applications où la transparence est en contradiction avec l'utilité. Notre programme de recherche en cours est destiné à valider notre hypothèse selon laquelle nous pouvons effectivement créer des robots transparents qui sont néanmoins émotionnellement engageants et utiles outils dans un large éventail d'environnements domestiques et quasi-domestiques. En attendant, il reste encore beaucoup de travail à faire pour démêler la relation entre transparence, utilité et confiance.

LES RÉFÉRENCES

- [1] Margaret Boden, Joanna Bryson, Darwin Caldwell, Kerstin Dauten hahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorell, Mick Wallis, Blay Whitby, et Alan Winfield. Principes de la robotique. Engi neering and Physical Sciences Research Council (EPSRC) du Royaume-Uni, avril 2011. publication Web.
- [2] C. Breazeal, CD Kidd, AL Thomaz, G. Hoffman et M. Berlin, "Effets de la communication non verbale sur l'efficacité et la robustesse dans travail d'équipe humain-robot ", en 2005 Conférence internationale IEEE / RSJ on Intelligent Robots and Systems, pp. 708–713, Alberta, Canada, (2005). leee.
- [3] Joanna J. Bryson, « Intelligence by design : principes de modularité et coordination pour l'ingénierie des agents adaptatifs complexes », (2001).
- [4] Joanna J Bryson et Paul Rauwolf, 'Trust, Communication, and In égalité'. 2016.
- [5] Kerstin Dautenhahn, « Méthodologie et thèmes de l'interaction homme-robot : un domaine de recherche en pleine croissance », International Journal of Advanced Robotic Systems, 4(1 SPEC. ISS.), 103–108, (2007).
- [6] Richard Dawkins, « Organisation hiérarchique : un principe candidat pour éthologie », dans Growing Points in Ethology, eds., PPG Bateson et RA Hinde, 7–54, Cambridge University Press, Cambridge, (1976).
- [7] RIM Dunbar, 'L'hypothèse du cerveau social', Evolutionary Anthropol gie, 178–190, (1998).
- [8] Shaun Gallagher, « L'alternative narrative à la théorie de l'esprit », dans Rad ical Enactivism : Intentionality, Phenomenology, and Narrative, éd., R Menary, numéro Gallagher 2001, 223–229, John Benjamins, Amsterdam, (2006).
- [9] P. a. Hancock, DR Billings, KE Schaefer, JYC Chen, EJ de Visser, et R. Parasuraman, 'Une méta-analyse des facteurs affectant Trust in Human-Robot Interaction', Human Factors: The Journal of the Société des facteurs humains et de l'ergonomie, 53(5), 517–527, (2011).
- [10] Peter H. Kahn, Hiroshi Ishiguro, Batya Friedman et Takayuki Kanda, 'Qu'est-ce qu'un humain? - Vers des repères psychologiques dans le domaine de interaction homme-robot', Actes - IEEE International Workshop on Robot and Human Interactive Communication, 3, 364–371, (2006).
- [11] Taemie Kim et Pamela Hinds, 'Qui dois-je blâmer ? Effets de l'autonomie et de la transparence sur les attributions dans l'interaction homme-robot », Actes - IEEE International Workshop on Robot and Human Interactive Communication, 80–85, (2006).
- [12] Bernt Meerbeek, Jettie Hoonhout, Peter Bingley et Jacques Terken, 'Enquêteur sur la relation entre la personnalité d'un téléviseur robotique assistant et le niveau de contrôle de l'utilisateur », Actes - IEEE International Workshop on Robot and Human Interactive Communication, 404–410, (2006).
- [13] Erik T. Mueller, Ordinateurs transparents : Concevoir des Systèmes intelligents, Erik T. Mueller, San Bernardino, Californie, 2016.
- [14] Rebecca Saxe, Laura E Schulz et Yuhong V Jiang, 'Reading minds contre les règles suivantes : dissociation de la théorie de l'esprit et du contrôle exécutif dans le cerveau.', Social neuroscience, 1(3-4), 284–98, (jan 2006).
- [15] Kristen Stubbs, Pamela J Hinds et David Wettergreen, "Autonomy and Common Ground in Human-Robot Interaction: A Field Study", Systèmes intelligents IEEE, 22(2), 42–50, (2007).
- [16] Robert H Wortham et Joanna J Bryson, 'Communication', dans Hand book of Living Machines {in press.}, Oxford University Press, Oxford, (2016).