

Une version éditée de ceci est apparue dans The AISB Quarterly Issue 133 (Spring 2012) page 14-15.

La fabrication des principes EPSRC de la robotique

Joanna J. Bryson

Groupe des modèles artificiels d'intelligence naturelle
(AmonI) Département d'informatique, Université de
Bath <http://www.cs.bath.ac.uk/> jjb

5 décembre 2011

Fin 2010, l'EPSRC m'a inopinément invité à assister à une réunion dans la New Forest sur le thème de l'éthique et des robots. J'écris des articles occasionnels sur ce sujet depuis 1996, en réponse à mon expérience d'être sur Cog (un projet de robot humanoïde) et de voir avec quelle facilité et même avec insistance les gens attribuaient une obligation morale à un complètement non fonctionnel (en 1993) mais vaguement robot humanoïde. Depuis 1998, j'ai également maintenu une page Web sur le sujet. Néanmoins, c'était la première fois que j'étais approché par un organisme gouvernemental, et bien sûr j'ai dit oui.

La rencontre était un offsite de trois jours présidé (savamment) par la journaliste Vivienne Parry. Outre moi, les autres participants comprenaient Margaret Boden, Darwin Caldwell, Kerstin Dautenhahn, Paula Duxbury, Lilian Edwards, Ann Grand, Hazel Grian, Sarah Kember, Stephen Kemp, Paul Newman, Geoff Peg man, Andrew Rose, Tom Rodden, Tom Sorell, Mick Wallis, Shearer West, Blay Whitby et Alan Winfield, ainsi que l'aide compétente de Ian Baldwin, Denise Dabbs et Paul O'Dowd. Les participants venaient principalement de la robotique, mais aussi des sciences humaines, du droit et des sciences sociales. Nous travaillions principalement dans le milieu universitaire, mais aussi dans l'industrie et les conseils de recherche (y compris celui qui dirigeait alors l'AHRC, Shearer West).

J'ai été surpris que l'EPSRC fasse des éclaboussures pour tant de personnes pendant si longtemps dans un si bel hôtel sur ce sujet, mais ils ont fait très clairement l'objet de leur préoccupation assez tôt. L'EPSRC considère la robotique comme une technologie critique pour le Royaume-Uni et ne veut pas la voir subir le même sort que d'autres technologies "futuristes", en termes de dégoût du public à la limite de l'hystérie qui ne peut plus être résolu par aucune quantité mesurée. évaluation scientifique. L'EPSRC veut obtenir l'éthique du robot dès le début, pour assurer à la fois la sécurité et l'acceptation des technologies robotiques.

Il est devenu presque immédiatement évident qu'ils avaient réussi à sélectionner un groupe d'experts très pragmatiques et socialement concernés. Le groupe a adopté une position très ferme sur ce que pourrait être le rôle moral et éthique de la robotique, et je ne dirais pas qu'il s'agit du rôle dominant lors des rassemblements typiques de l'AISB.

Le dernier jour complet, Lilian Edwards, professeur de droit de l'Internet, et moi étions dans une

petite réunion en petits groupes lors d'une session destinée à concevoir des livrables en tant que résultats de la réunion. Nous avons décidé de créer un « vrai » ensemble de lois pour les robots.

Lillian tenait à ce qu'ils suivent clairement mais corrigent les lois d'Asimov, tandis que je tenais à en inclure plusieurs que j'avais déjà développées en écrivant A Proposal for the Humanoid Agent-builders League (HAL) (Bryson, 2000). Au final nous nous sommes arrêtés sur cinq, dont les trois premiers reflètent et réfractent Asimov aux préoccupations du groupe. Le groupe dans son ensemble a alors affiné non seulement nos « lois », mais aussi leurs versions en langage courant, et a élaboré une nouvelle liste de concepts à vous communiquer, chers collègues.

La version complète de ces documents peut maintenant être trouvée en recherchant sur Google les principes de robotique de l'EPSRC, et ils sont devenus la politique de l'EPSRC depuis avril 2011. Les principes de base sont reproduits dans l'encadré. Vous trouverez ci-dessous sept idées de haut niveau que le groupe souhaite vous communiquer, nos collègues. Pour des explications détaillées, veuillez consulter le site Web, mais j'ai donné les points saillants ici.

1. Nous pensons que les robots ont le potentiel d'avoir un immense impact positif sur la société. Nous voulons encourager la recherche responsable sur les robots. Nous ne sommes pas une bande de luddites qui "ne comprennent pas" le véritable potentiel de l'IA. Nous sommes des professionnels inquiets qui veulent vraiment que l'IA fonctionne et que les robots deviennent réels.
2. Les mauvaises pratiques nous blessent tous. Nous ne pouvons pas ignorer la situation si certains de nos collègues font des choses qui nous font tous mal paraître.
3. Répondre aux préoccupations évidentes du public nous aidera tous à progresser.
4. Il est important de démontrer que nous, en tant que roboticiens, nous engageons à les meilleures normes de pratique possibles.
5. Pour comprendre le contexte et les conséquences de notre recherche, nous devrions travailler avec des experts d'autres disciplines, notamment : les sciences sociales, le droit, la philosophie et les arts. Nous avons tous été frappés par tout ce que nous avons appris de cette équipe de travail multidisciplinaire.
6. Nous devrions considérer l'éthique de la transparence : y a-t-il des limites à ce qui devrait être librement accessible ? Tout le monde à la réunion était attaché aux solutions et aux approches de type logiciel open source, mais nous nous sommes rendu compte qu'avec les robots et l'IA plus généralement, nous avons l'obligation de nous assurer que chaque "script kiddie" ne puisse pas pirater un système qui possède des informations ou une mémoire sur la vie privée des humains.
7. Lorsque nous voyons des récits erronés dans la presse, nous nous engageons à prendre le temps de contacter les journalistes qui les rapportent. La plupart des journalistes scientifiques ne veulent vraiment pas qu'on leur donne l'air stupide en rapportant un « expert » qui s'avère faire de l'autopromotion ou faire du sensationnalisme. Un mot ou un e-mail discret peut souvent atténuer l'hystérie générée par des déclarations irresponsables.

Je tiens à remercier l'EPSRC ainsi que nos collègues qui ont plaidé pour cette réunion. Deux de ces derniers étaient Alan Winfield et Tom Rodden.

Personnellement, je suis extrêmement fier et heureux pour ma profession et ma nation que le Royaume-Uni dispose désormais d'un ensemble officiel de principes de robotique qui traitent de questions aussi importantes. Mais nous ne sommes qu'un seul pays, et il reste encore beaucoup de travail et de plaider à faire pour s'assurer que la robotique intelligente est utilisée de manière appropriée dans notre société.

Boîte!

1. Les robots sont des outils multi-usages. 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. Tout en reconnaissant que même les poissons morts peuvent être utilisés comme armes par des individus créatifs, nous étions soucieux d'interdire la création et l'utilisation de robots autonomes comme armes. Bien que nous ayons reconnu de manière pragmatique que cela se produit déjà dans le contexte militaire, nous ne voulons pas les voir utilisés dans d'autres contextes.
2. Les humains, et non les robots, sont des agents responsables. Les robots doivent être conçus et exploités dans la mesure du possible pour se conformer aux lois existantes et aux droits et libertés fondamentaux, y compris la vie privée. Nous craignons vivement que toute discussion sur « l'éthique des robots » ne conduise des individus, des entreprises ou des gouvernements à renoncer à leur propre responsabilité en tant que constructeurs, acheteurs et déployeurs de robots. Nous avons estimé que les conséquences de cette préoccupation l'emportaient largement sur tout « avantage » au plaisir de créer quelque chose que la société daigne sensible et responsable. C'était la loi dont nous savions qu'elle protégerait la plupart de nos collègues de l'ASIB - par conséquent (avec David Gunkel)
Je dirige un symposium à l'ASIB 2012 pour examiner si c'est une règle raisonnable. Le symposium s'intitule « La question de la machine : IA, éthique et responsabilité morale ».
3. Les robots sont des produits. Ils doivent être conçus à l'aide de processus garantissant leur sûreté et leur sécurité. Ce principe nous rappelle à nouveau qu'il nous incombe, en tant que créateurs de robots, et non aux robots eux-mêmes, de veiller à ce que les robots ne causent aucun dommage.
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. C'était la règle la plus difficile pour s'entendre sur la formulation. L'idée est que tous ceux qui possèdent un robot doivent savoir qu'il n'est pas "vivant" ou "souffrant", mais la tromperie de la vie et l'engagement émotionnel sont précisément l'objectif de nombreux robots de thérapie ou jouets. Nous avons décidé que tant que la personne responsable qui achète un robot a même un accès indirect (par exemple, la documentation Internet) à des informations sur le fonctionnement de son "esprit", cela fournirait une population suffisamment informée pour empêcher les personnes d'être exploitées.

5. La personne ayant la responsabilité légale d'un robot doit être désignée. Il devrait toujours être possible de savoir à qui appartient un robot, tout comme il est toujours possible de savoir à qui appartient une voiture. Cela nous rappelle à nouveau que quoi que fasse un robot, une institution humaine ou humaine (par exemple une entreprise) est responsable de ses actes.

Voir aussi (note à l'éditeur, veuillez combiner les pages Web et les références en une seule liste pour un total de cinq éléments) :

- Le site Web des principes de la robotique de l'EPSRC.
- La question de la machine : IA, éthique et responsabilité morale.
- Ethique : IA, Robots et Société.
- (Wilk, 2010)

Les références

En ligne Bryson, JJ (2000). Une proposition pour la Humanoid Agent-builders League (HAL).

Dans Barnden, J., éditeur, AISB'00 Symposium on Artificial Intelligence, Ethics and (Quasi-)Human Rights, pages 1–6.

Wilks, Y., éditeur (2010). Engagements rapprochés avec des compagnons artificiels : principaux problèmes sociaux, psychologiques, éthiques et de conception. John Benjamins, Amsterdam.