

Una versión editada de esto apareció en The AISB Quarterly Issue 133 (primavera de 2012), página 14-15.

La creación de los principios de robótica de EPSRC

Joanna J. Bryson

Grupo de Modelos Artificiales de Inteligencia Natural (AmonI) Departamento
de Informática, Universidad de Bath <http://www.cs.bath.ac.uk/> jjb

5 de diciembre de 2011

A finales de 2010, el EPSRC me invitó inesperadamente a asistir a una reunión en New Forest sobre el tema de la ética y los robots. He estado escribiendo artículos ocasionales sobre ese tema desde 1996, en respuesta a mi experiencia de estar en Cog (un proyecto de robot humanoide) y ver cuán fácil e incluso insistentemente la gente atribuía la obligación moral hacia un completamente no funcional (en 1993) pero vagamente robot humanoide. Desde 1998 también mantengo una página web sobre el tema. Sin embargo, esta era la primera vez que un organismo gubernamental se me acercaba y, por supuesto, dije que sí.

La reunión fue un offsite de tres días presidida (expertamente) por la periodista Vivienne Parry. Además de mí, otros participantes incluyeron a Margaret Boden, Darwin Caldwell, Kerstin Dautenhahn, Paula Duxbury, Lilian Edwards, Ann Grand, Hazel Grian, Sarah Kember, Stephen Kemp, Paul Newman, Geoff Pegman, Andrew Rose, Tom Rodden, Tom Sorell, Mick Wallis, Shearer West, Blay Whitby y Alan Winfield, así como la hábil asistencia de Ian Baldwin, Denise Dabbs y Paul O'Dowd. Los participantes provenían principalmente de la robótica, pero también de las humanidades, el derecho y las ciencias sociales. Estábamos empleados principalmente en la academia, pero también en la industria y los consejos de investigación (incluido el entonces director de la AHRC, Shearer West).

Me sorprendió que el EPSRC pagara por tanta gente durante tanto tiempo en un hotel tan agradable sobre este tema, pero dejaron muy claro el objeto de su preocupación desde el principio. El EPSRC ve la robótica como una tecnología crítica para el Reino Unido, y no quiere que enfrente el mismo destino que otras tecnologías "futuristas", en términos de disgusto público que roza la histeria que ya no puede ser abordado por ninguna cantidad de medidas medidas. evaluación científica. El EPSRC quiere que la ética de los robots sea correcta desde el principio, para garantizar tanto la seguridad como la aceptación de las tecnologías robóticas.

Casi de inmediato se hizo evidente que habían logrado seleccionar un grupo de expertos muy pragmático y socialmente preocupado. El grupo tomó una línea muy fuerte sobre cuál podría ser el papel moral y ético de la robótica, y uno que no diría que es el dominante en las reuniones típicas de AISB.

El último día completo, la profesora de Derecho de Internet Lilian Edwards y yo estábamos en un

pequeña reunión de grupo en una sesión destinada a diseñar entregables como resultados de la reunión. Decidimos hacer un conjunto "real" de leyes para robots.

Lillian estaba interesada en que siguieran claramente las leyes de Asimov pero las corrigieran, mientras que yo estaba interesado en incluir varias que ya había desarrollado mientras escribía Una propuesta para la Liga de constructores de agentes humanoides (HAL) (Bryson, 2000). Al final nos decidimos por cinco, los tres primeros de los cuales reflejan y refractan Asimov a las preocupaciones del grupo. Luego, el grupo en su totalidad perfeccionó no solo nuestras "leyes", sino también sus versiones en lenguaje ordinario, y desarrolló una lista adicional de conceptos para comunicarles a ustedes, nuestros colegas.

La versión completa de estos documentos ahora se puede encontrar buscando en Google los Principios de robótica de EPSRC, y se han convertido en políticas de EPSRC desde abril de 2011. Los principios básicos se reproducen en el recuadro. A continuación se presentan siete ideas de alto nivel que el grupo quiere comunicarles a ustedes, nuestros colegas. Para obtener explicaciones detalladas, consulte el sitio web, pero he dado los aspectos más destacados aquí.

1. Creemos que los robots tienen el potencial de brindar un inmenso impacto positivo a la sociedad. Queremos fomentar la investigación robótica responsable. No somos un grupo de luditas que "no entienden" el verdadero potencial de la IA. Somos profesionales preocupados que realmente quieren hacer que la IA funcione y que los robots sean reales.
2. Las malas prácticas nos perjudican a todos. No podemos ignorar la situación si algunos de nuestros colegas hacen cosas que nos hacen quedar mal a todos.
3. Abordar las preocupaciones públicas obvias nos ayudará a todos a progresar.
4. Es importante demostrar que nosotros, como roboticistas, estamos comprometidos con los mejores estándares de práctica posibles.
5. Para comprender el contexto y las consecuencias de nuestra investigación, debemos trabajar con expertos de otras disciplinas, incluidas: ciencias sociales, derecho, filosofía y artes. A todos nos llamó la atención lo mucho que aprendimos de este equipo de trabajo multidisciplinario.
6. Deberíamos considerar la ética de la transparencia: ¿existen límites a lo que debería estar disponible abiertamente? Todos en la reunión estaban comprometidos con las soluciones y los enfoques del tipo de software de código abierto, pero nos dimos cuenta de que con los robots y la IA, en general, tenemos la obligación de asegurarnos de que cada "script kiddy" no pueda piratear un sistema que tiene información o memoria sobre la vida privada de los humanos.
7. Cuando vemos relatos erróneos en la prensa, nos comprometemos a tomarnos el tiempo para contactar a los periodistas informantes. La mayoría de los reporteros científicos realmente no quieren quedar en ridículo al informar sobre un "experto" que resulta ser autopromocionado o sensacionalista. Una palabra tranquila o un correo electrónico a menudo pueden amortiguar la histeria generada por declaraciones irresponsables.

Me gustaría agradecer al EPSRC y también a nuestros colegas que abogaron por esta reunión. Dos de estos últimos fueron Alan Winfield y Tom Rodden. Personalmente, me siento extremadamente orgulloso y feliz por mi profesión y mi nación de que el Reino Unido ahora tenga un conjunto oficial de principios de robótica que abordan asuntos tan importantes. Pero somos solo un país, y todavía queda mucho trabajo e incidencia por hacer para garantizar que la robótica inteligente se utilice adecuadamente en nuestra sociedad.

¡Caja!

1. Los robots son herramientas de usos múltiples. Los robots no deben estar diseñados única o principalmente para matar o dañar a los seres humanos, excepto en interés de la seguridad nacional. Si bien reconocimos que incluso los peces muertos pueden ser utilizados como armas por personas creativas, nos preocupaba prohibir la creación y el uso de robots autónomos como armas. Aunque reconocimos pragmáticamente que esto ya está sucediendo en el contexto militar, no queremos que se usen en otros contextos.
2. Los humanos, no los robots, son los agentes responsables. Los robots deben diseñarse y operarse en la medida de lo posible para cumplir con las leyes existentes y los derechos y libertades fundamentales, incluida la privacidad. Estábamos muy preocupados de que cualquier discusión sobre la "ética de los robots" pudiera llevar a individuos, empresas o gobiernos a derogar su propia responsabilidad como constructores, compradores e implementadores de robots. Sentimos que las consecuencias de esta preocupación superan con creces cualquier "ventaja" al placer de crear algo que la sociedad se digne consciente y responsable. Esta era la ley que sabíamos que la mayoría de las veces defendería a algunos de nuestros colegas en AISB, en consecuencia (con David Gunkel) Estoy organizando un simposio en AISB 2012 para examinar si esta es una regla razonable. El simposio se llama "La cuestión de las máquinas: IA, ética y responsabilidad moral".
3. Los robots son productos. Deben diseñarse utilizando procesos que aseguren su seguridad y protección. Este principio nos recuerda nuevamente que la responsabilidad recae en nosotros, como creadores de robots, no en los robots mismos, para garantizar que los robots no causen daños.
4. Los robots son artefactos manufacturados. No deben diseñarse de manera engañosa para explotar a los usuarios vulnerables; en cambio, su naturaleza de máquina debería ser transparente. Esta fue la regla más difícil de acordar en la redacción. La idea es que todo aquel que tenga un robot sepa que no está "vivo" ni "sufrido", pero el engaño de la vida y el compromiso emocional es precisamente el objetivo de muchos robots terapéuticos o de juguete. Decidimos que siempre que la persona responsable que hace la compra de un robot tenga acceso incluso indirecto (por ejemplo, documentación de Internet) a información sobre cómo funciona su "mente", eso proporcionaría una población suficientemente informada para evitar que las personas sean explotadas.

5. Se debe atribuir la responsabilidad legal de un robot. Siempre debería ser posible averiguar quién es dueño de un robot, al igual que siempre es posible saber quién es dueño de un automóvil. Esto nuevamente nos recuerda que cualquier cosa que haga un robot, alguna institución humana o humana (por ejemplo, una empresa) es responsable de sus acciones.

Ver también (nota para el editor, combine las páginas web y las referencias en una lista para un total de cinco elementos):

- El sitio web de Principios de Robótica de EPSRC.
- La cuestión de la máquina: IA, ética y responsabilidad moral.
- Ética: IA, Robots y Sociedad.
- (Wilks, 2010)

Referencias

Bryson, JJ (2000). Una propuesta para la Humanoid Agent-builders League (HAL). En Barnden, J., editor, Simposio AISB'00 sobre inteligencia artificial, ética y derechos (cuasi) humanos, páginas 1–6.

Wilks, Y., editor (2010). Compromisos cercanos con compañeros artificiales: cuestiones clave sociales, psicológicas, éticas y de diseño. John Benjamins, Amsterdam.