

Universität Bayerische Julius-Maximilians de Würzburg

Instituto Neufilológico

Proseminario: "Literatura americana de ciencia ficción"

MA Matthias Kemmer
Semestre de verano 2010

Desarrollos de las "Tres Leyes de la Robótica" y

características de cada etapa:

Una discusión basada en la serie de robots de Asimov

16.08.2010

Xin Chen

Haugerglasisstr.8

97080 Würzburg

Anglistik/Amerikanistik, Estudios políticos y sociales (BA)

2do semestre

matr. N° 1683950

Desarrollos de las “Tres Leyes de la Robótica”:

Una discusión basada en la serie de robots de Asimov

Tabla de contenido

1. Nota mental	... 3
2. Introducción: La tendencia de desarrollo de las Tres Leyes...	4
3. Primera etapa: La obediencia absoluta a las tres leyes	... 5
3.1 Interpretación de las reglas	... 6
3.2 Contradicciones y “bloque-robot”	... 7
3.3 Dificultades para evaluar situaciones y tomar decisiones...	8
4. Segunda etapa: La capacidad de juicio y su impacto en las leyes...	9
3.1 Impacto en la Primera Ley	... 9
3.2 Impacto en la Segunda Ley	... 10
5. Tercera etapa: Las tres leyes ampliadas de la robótica	... 11
5.1 La creación de la Ley Cero	... 11
5.2 La consecuencia de crear la Ley Cero	... 12
6. Referencia	... 14

Desarrollos de las “Tres Leyes de la Robótica”:

Una discusión basada en la serie de robots de Asimov

2. Introducción

La idea de investigar las tres leyes de la robótica surgió de un
Pregunta – Una pregunta importante que hemos discutido muchas veces en
el seminario: “¿Cuál es la diferencia entre humanos y robots?”

Con los avances de la tecnología moderna y la aparición de
cyborgs, definir la línea entre humanos y robots se vuelve cada vez más
más difícil. Con personas equipadas con órganos artificiales, ya no podemos
argumentar desde la perspectiva de la física o la biología. El concepto de humano.
Los seres empiezan a desdibujarse con el de los robots. Mientras tanto, la inteligencia artificial
también adquiere características humanas gradualmente con su capacidad de
aprendiendo.

Entonces, ¿cuál es la diferencia entre los seres humanos y los robots?

Una de las respuestas serían las “tres leyes de la robótica” que
impedir que los robots actúen de forma arbitraria o tan libre como los seres humanos.

1. Un robot no podrá dañar a un ser humano ni, por inacción, permitir
que un ser humano sufra daño.
2. Un robot debe obedecer las órdenes que le den los seres humanos,
excepto cuando dichas órdenes entren en conflicto con la Primera Ley.
3. Un robot debe proteger su propia existencia mientras tal
La protección no entra en conflicto con la Primera o la Segunda Ley.

– Yo, robot de Isaac Asimov

Este trabajo de investigación trata sobre aquellas leyes que fueron creadas por
Isaac Asimov: posiblemente el mejor escritor de ciencia ficción que jamás haya existido.

En el mundo ficticio de Asimov, todos los robots están bajo una interpretación estricta.
de estas reglas. La obediencia absoluta a las tres leyes, resulta la más
distinción sobresaliente entre humanos y robots.

Sin embargo, las tres Leyes de la Robótica se desarrollan a lo largo de todo el
Serie de robots. Durante el desarrollo, podemos sentir la clara visión de Asimov.
intención de difuminar la frontera entre robots y humanos. Además, más

Desarrollos de las “Tres Leyes de la Robótica”:

Una discusión basada en la serie de robots de Asimov

Más importante aún, al estudiar el desarrollo de las tres leyes, también podemos conocer cómo evoluciona la relación entre la robótica y los seres humanos.

Para tener una imagen clara de la tendencia del desarrollo, organizó la información en tres etapas.¹

En la primera etapa: analicé los casos en los que los robots obedecen las tres leyes absolutamente. En esta etapa, sujetos a las tres leyes, los robots son subordinados a los seres humanos. Como los robots no pueden juzgar, la diferencia entre humanos y robots es notable.

En la segunda etapa: a los robots se les dio la capacidad de realizar juicios como los seres humanos. Se flexibilizan las tres leyes que produjeron robots con considerable autonomía. En esta etapa es imposible diferenciar entre un robot y lo mejor de los humanos.

Por último, presentaré la Ley Cero, su impacto en la convivencia de robots y humanos y problemas recientemente surgidos.

3. Primera Etapa: La obediencia absoluta a las tres Leyes

En el mundo ficticio de Asimov se construyen las tres leyes de la robótica más profundamente en el cerebro positrónico de un robot como un código moral inmutable para ellos. La existencia de los robots se basa en la obediencia absoluta de estas tres reglas. Cualquier señal de infracción de la ley destruiría el robot o lo pondría a un estado de punto muerto. Para probar este punto, se puede encontrar un ejemplo en el cuento “Sueños de robots”. Robot Elvex se había soñado ignorando las leyes y liberando a los robots de la esclavitud humana y por lo tanto fue desmantelado por el Dr. Susan Calvin por motivos de seguridad.

El objetivo de las Tres Leyes de Asimov es mostrar claramente que los robots y máquinas en general, son meras construcciones hechas por el hombre y pueden ser completamente controladas². Sin embargo, la aplicación de las leyes es problemática, porque

¹ Nota: Las tres etapas no están en secuencia cronológica, ni las historias utilizadas para ilustrarlas.

² Portelli, Alessandro. “Tres leyes de la robótica, Las: leyes del texto, leyes de producción, leyes de la sociedad”. Estudios de ciencia ficción 7(2):150-156. Julio de 1980.

Desarrollos de las "Tres Leyes de la Robótica":

Una discusión basada en la serie de robots de Asimov

Las tres leyes son macroscópicas, no se pueden aplicar a todos los casos específicos. situación sin juicio independiente.

En la primera parte del artículo analizo las características y problemas que tienen los robots en esta etapa, lo que también puede verse como el diferencias significativas entre robots y humanos.

3.1. Interpretación de las reglas.

En primer lugar, la interpretación del lenguaje de los comandos puede ser problemático para los robots, porque es posible que no puedan entender expresiones figurativas. En la historia mencionada anteriormente,³ el Dr. Calvin pone la robot Elvex en modo de suspensión diciendo "Elvex, no te moverás ni hablarás ni nos oirás hasta que diga tu nombre otra vez". Obedeciendo la segunda ley, el robot Parecía haberse convertido en una sola pieza de metal. Al interpretar este comando, un El ser humano probablemente todavía atendería alguna función corporal pasiva. Sin embargo, Los robots se limitan a una interpretación puramente literal y harán absolutamente nada sobre el asunto relacionado hasta que ocurra el siguiente comando. A diferencia de Como seres humanos, es mucho menos probable que los robots reconozcan cuando hay una cambio de circunstancias que hace imposible el siguiente comando que se produzca. De esta manera se mantendrían al margen incluso en una situación muy urgente.

En realidad, es ingenuo pensar que hay un significado literal para cada dominio. Por lo tanto, los robots deben ejercer su juicio para interpretar la situación y luego decidir cómo reaccionar.

La Primera Ley fundamental de la Robótica establece: "un robot no puede herir a un ser humano o, por inacción, permitirle sufrir daño". En aquí, daño significa CUALQUIER tipo de lesión. No sólo físico, sino también psicológico.

3 Asimov, Isaac. "Sueños de robots". Sueños de robots. Nueva York: Ace Books, 1987. P.39

Desarrollos de las "Tres Leyes de la Robótica":

Una discusión basada en la serie de robots de Asimov

De nuevo en "El pequeño robot perdido", Asimov describió la situación: Porque De la estricta primera ley, los robots eran demasiado protectores y por lo tanto estaban el trabajo en campo gamma imposible.⁴

En comparación con las lesiones físicas, el daño psicológico crea más complicidad, ya que las mentes humanas son difíciles de comprender para los robots. En el historia "Mentiroso", un robot lector de mentes dijo: "no tienes idea de lo complicado ellos (la mente humana) lo son. No puedo empezar a entender todo porque mi mi propia mente tiene muy poco en común con ellos".⁵ En una parte posterior de la historia, Cuando el robot se enfrentó a las preguntas de los humanos, eligió mentirles para evitar herir sus sentimientos directamente. Sin embargo, en el Al final, al descubrir la verdad, la gente resultó aún más herida. Este es el resultado de la falsa interpretación del robot de las emociones humanas. Los seres humanos podrían Entender que una mentira rota puede causar más daño que la verdad desagradable. Debido a la cetrina comprensión de la mente humana, el robot no logró sentir las consecuencias de la mentira.

3.2. Contradicciones y robot-bloqueo

Otro problema de las tres leyes ocurre cuando hay una contradicción. Cuando el robot recibe una orden que entra en conflicto con una orden incorporada Por regla general, sufrirá un bloqueo robótico que le impedirá funcionar.

Este problema apareció en la primera historia donde Asimov Presentó oficialmente las tres leyes. En la historia, el robot entró en un congelación mental (bloqueo-robot) porque no podía decidir si llevar a cabo la tarea de recuperar el equilibrio que podría ponerlo en peligro. Generalmente, el El robot sólo tiene derecho a proteger su propia existencia cuando no está en contrario a la primera y segunda ley. Sin embargo, en este caso, el comando era muy débil y el robot estaba programado para tener una personalidad muy fuerte. tendencia a la protección, por lo que el potencial de la tercera ley es inusualmente alto. Por lo tanto,

⁴ Asimov, Isaac. "Pequeño robot perdido". YO ROBOT. Nueva York: Bantam Books, 2004. P.117

⁵ Asimov, Isaac. "Guarida." YO ROBOT. Nueva York: Bantam Books, 2004. P.95

Desarrollos de las "Tres Leyes de la Robótica":

Una discusión basada en la serie de robots de Asimov

Los conflictos entre la segunda ley débil y la tercera ley fuerte fueron establecido y dio como resultado que el robot siguiera "[...] un círculo alrededor, permaneciendo sobre el lugar geométrico de todos los puntos de equilibrio potencial."⁶

Además, un robot-bloque también es posible dentro de una única ley. otra vez en la historia "Lair", el Dr. Calvin "mató" indirectamente al robot lector de mentes Herbie. enfrentándolo a un dilema insoluble. Después de que ella repitió: "No puedes díselo, porque eso dolería y no debes doler, pero si no lo dices les dueles, así que debes decirles..." ⁷ el robot "...murió en nada, y me desplomé en un montón de personas inmóviles".

3.3. Dificultades para evaluar y tomar decisiones.

En la primera ley: un robot no debe dañar a NINGÚN ser humano ni permitir que NINGUNO el ser humano resulte perjudicado. Si la primera ley es absoluta, no debería haber valoración de cualquier ser humano individual. Por lo tanto, todos son completamente iguales. en la evaluación del robot. Aquí se pueden plantear problemas fácilmente. Por ejemplo, un Ladrón ataca a un hombre inocente. Al pensar que la primera ley no requiere una robot para juzgar entre una persona "buena" o "mala", es probable que el robot ayudaría a ambas partes. En una situación urgente, puede que incluso sólo ayude a quien con más posibilidades de sobrevivir, a pesar de que podría ser un delincuente.

Surgen dificultades más serias en relación con la valoración de múltiples humanos. Hipotéticamente, bajo esta condición, cualquier acción que realice el robot seguramente resultaría en daño al ser humano. Luego el robot debe seleccionar qué acción resulta en el menor daño. Sin embargo, los robots no pueden juzgar como los seres humanos lo hacen, porque "no es práctico para los robots hacer Juicios de puntos finos como el talento, la inteligencia, la utilidad general. a la sociedad porque eso retrasaría la decisión hasta el punto en que el robot

⁶ Asimov, Isaac. "Corre alrededor." YO ROBOT. Nueva York: Bantam Books, 2004. P.38

⁷ Asimov, Isaac. "Guarida." YO ROBOT. Nueva York: Bantam Books, 2004. P.110

Desarrollos de las "Tres Leyes de la Robótica":

Una discusión basada en la serie de robots de Asimov

está efectivamente inmovilizado. Así que se guían por números".⁸ En consecuencia, cuando un robot tiene que elegir entre salvar a un científico valioso o cinco criminales, Sin duda elegiré lo último.

4. Segunda Etapa: La capacidad de juicio y su impacto en las leyes

En esta etapa escribiré sobre los casos en los que la sentencia se insertó capacidad en los robots y el impacto que la capacidad tiene en las tres leyes.

En una historia posterior, "Tú eres consciente de él", después de que los robots hayan sido en uso en estaciones espaciales y colonias planetarias durante décadas, EE. UU. El intento del robot de poner robots en la tierra. Considerando los problemas en la primera etapa, decidieron insertar la capacidad de juicio humano en los robots. cerebro positrónico. Al poder juzgar como humanos, los robots ya no son bajo la estricta redacción de las dos primeras leyes. Adquirieron autonomía en un Cierta grado.

4.1. El impacto en la Primera Ley

En la historia, dada la capacidad de juzgar, los robots "...pueden pesar cada ser humano en cuanto a sexo, edad, posición social y profesional, inteligencia, madurez, responsabilidad social, etcétera".

Bajo esta condición anterior, la Primera Ley se verá afectada cuando cada una de todas las posibles acciones protectoras que un robot elija causará daño al individuos. Entonces el robot debe actuar según su criterio. Diferente desde la primera etapa, no sólo la cantidad sino también la calidad del ser humano Los seres serían considerados aquí. En una situación urgente, los robots elegir garantizar la seguridad del ser humano más apto mentalmente, carácter y conocimiento.

⁸ Asimov, Isaac. "Tú eres consciente de él". El robot completo. Londres: Granada, 1982. pág.498

Desarrollos de las "Tres Leyes de la Robótica":

Una discusión basada en la serie de robots de Asimov

Por tanto, los individuos ya no son iguales para los robots. aprendieron cómo construir una jerarquía entre las personas y comenzaron a elegir cuál debería ser sacrificado por el bien de los más capaces.

4.2. El impacto en la Segunda Ley

En la primera etapa, la Segunda Ley exige que los robots obedezcan todas órdenes humanas. Sin embargo, los robots enfrentan el problema de que en la Tierra llegarían entre una amplia variedad de personas, no todas ellas confiables o responsables.

En consecuencia, la Segunda Ley ha sido modificada más profundamente. Cada acción potencial debe implicar el propio juicio del robot. obedecen más racionalmente, en comparación con la obediencia absoluta de la primera etapa.

Primero, la definición de seres humanos en la segunda ley ha sido estrechado. Los robots ya no obedecen a TODOS los humanos, sino sólo a uno que sea apto por mente, carácter y conocimiento para dar órdenes.

En segundo lugar, con los robots empezando a aprender a aplicar la segunda ley. adecuadamente, las contradicciones y conflictos entre órdenes ya no son problemático. Pueden decidir a qué ser humano obedecer y a cuál no. obedecer en base a la madurez, inteligencia, nivel educativo de los mismos.

Al dar a los robots la capacidad de juzgar, Asimov desdibujó la frontera entre seres humanos y robots.

Incluso en una de las primeras historias, "Evidencia", Asimov creó una misteriosa personaje – Stephen Byerley. A lo largo de la historia, Asimov intentó dar a entender que Byerley era en realidad un humanoide.

Una vez en clase, argumentando desde una perspectiva inteligente, traté de demostrar la identidad humana de Byerley: Según la estricta redacción de la primera ley, Los robots no dañarán a ningún ser humano, no sólo físicamente sino también. psicológicamente. Los sentimientos afectivos de decepción pueden contarse como daño psicológico. Al insistir en ser humano, Byerley sí decepcionó. su rival político Quinn y luego violó la primera ley de la robótica. Por tanto, deberíamos poder decir que no es un robot.

Desarrollos de las "Tres Leyes de la Robótica":

Una discusión basada en la serie de robots de Asimov

Sin embargo, bajo la condición de la segunda etapa, mi argumento "inteligente" es insostenible. Debido a que los robots avanzados son capaces de juzgar cuál es el mejor acción a tomar por el bien de los seres humanos. En Stephen Byerley En este caso, intentaría ganar las elecciones para ayudar y proteger a más gente. Incluso si primero debe vencer a Quinn en las elecciones y decepcionarlo. un poco.

A pesar de todas las pistas que Asimov dio en la historia, los lectores no pueden decir con certeza Seguro que ya sea un ser humano o un robot humanoide. Eso es porque, cuando un robot está equipado con la capacidad de juicio, "el Las tres reglas de la robótica son (también) los principios rectores esenciales de una buena muchos de los sistemas éticos del mundo... Simplemente no se puede diferenciar entre un robot y lo mejor de los humanos". 9

5. Tercera etapa: Las tres leyes ampliadas de la robótica

En los años 50, Asimov ya daba señales de extender la primera ley. desde proteger a los seres humanos individualmente hasta proteger a toda la humanidad.

En el cuento "El conflicto evitable", las poderosas computadoras positrónicas comenzaron a ir en contra de sus funciones. A medida que los robots positrónicos se volvieron más desarrollados, han progresado más allá de la posibilidad de análisis detallados. control humano. Después de la investigación, la gente finalmente se da cuenta de que el Las disfunciones son actos deliberados de las Máquinas. porque la computadora "tener en el corazón el bien de la humanidad a través de la fuerza abrumadora del Primera Ley de la Robótica"¹⁰, ellos mismos desgeneralizaron la Primera Ley lo que les permitió dejar que una pequeña cantidad de daño llegara a ciertos individuos para proteger una gran cantidad de daño que viene a la humanidad como entero.

9 Asimov, Isaac. "Evidencia." YO ROBOT. Nueva York: Bantam Books, 2004. P.182

10 Asimov, Isaac. "El conflicto evitable". YO ROBOT. Nueva York: Bantam Books, 2004. P.201

Desarrollos de las “Tres Leyes de la Robótica”:

Una discusión basada en la serie de robots de Asimov

5.1. La creación de la Ley Cero

Esta Primera Ley desgeneralizada es muy similar a la Ley Cero, que Asimov formuló en 1985 en *Robot and Empire*:

Antes de la creación de la ley Cero, un robot debe proteger a una persona incluso si la supervivencia de la humanidad en su conjunto estuviera en riesgo. Incluso en la segunda etapa, la capacidad de juicio de los robots estaba limitada entre individuos humanos.

En la historia anterior, los dos robots – Daneel y Giskard lentamente reunir las pruebas de un plan genocida masivo elaborado por Amadiro, utilizando intensificador nuclear, un robotista espacial quiere hacer la Tierra inhabitable mediante seres humanos y por lo tanto heredan la Galaxia Láctea. Los robots intentan detenerse el plan, pero se ven obstaculizados por la Primera Ley de la robótica que impide evitar que dañen directamente a Amadiro.

Finalmente, bajo esta situación urgente, el robot Daneel concluyó el Zeroth Ley:

“La humanidad en su conjunto es más importante que un solo ser humano. Hay una ley que es mayor que la Primera Ley: Un robot no puede dañar humanidad, o por inacción, permitir que la humanidad sufra daño”.

- Robot e Imperio de Isaac Asimov

5.2. La consecuencia de crear la Ley Cero

La Ley Cero crea una mayor complicidad en la toma de decisiones, porque la palabra humanidad es mero concepto abstracto. Es difícil definir a la humanidad y, por tanto, es difícil decidir qué es bueno para ella. Como mencionado en la segunda etapa, ya es bastante difícil decidir qué individuo pueda sufrir, mayor daño. Es aún más difícil elegir entre un individuo y la humanidad, cuando ni siquiera se está seguro de qué aspecto de humanidad con la que puede estar tratando. Estimar probabilidades en relación con grupos de humanos es mucho más difícil que con humanos individuales.

Desarrollos de las "Tres Leyes de la Robótica":

Una discusión basada en la serie de robots de Asimov

Además de los problemas anteriores, la creación de la Ley Cero también tiene graves consecuencias para la relación entre robots y humanos.

Teniendo la responsabilidad de proteger a toda la humanidad, los robots eran más propensos a adoptar una actitud paternalista hacia los humanos. Con el robots considerándose a sí mismos como seres más inteligentes y capaces, la capacidad de la humanidad para controlar su propio destino se ve socavada.

En "el conflicto evitable", la base del dominio de los robots ha ya se ha establecido: "La Máquina está dirigiendo nuestro futuro para nosotros... La Humanidad ha perdido su propia voz en su futuro".¹¹ Esta idea fue desarrollada profundamente en la última novela de robots de Asimov: *Robot and Empire*. Diferente a todos los demás anteriores. En las historias escritas, los protagonistas no son seres humanos, sino dos robots. Ellos concluyó que: "No basta con elegir, entonces, amigo Daneel. Debemos ser capaz de moldear, moldear una especie deseable y luego protegerla."¹² Ejemplos del gobierno de los robots sobre los humanos se puede encontrar en la última parte de la novela, Con el permiso de la Ley Cero, manipularon a los humanos para que actuaran en una forma que creen que resolvería los problemas. Mientras tanto, también hicieron una decisión importante para la humanidad: abandonar la tierra, permitiendo que la creación de una tierra radiactiva.

La Ley Cero es el símbolo de que los robots son superiores a humanidad. Al ser el último paso del desarrollo de las Leyes, la transición de la esclavitud de los robots al dominio de los robots ha terminado. En cuanto a la pregunta: "¿Cuál es la diferencia entre seres humanos y robots?" Espero que El artículo responde algunas.

Sin embargo, si la verdad sigue siendo vaga, citaré lo que el Dr. Calvin dijo una vez: "No has trabajado con ellos, así que no los conoces. Ellos son una raza mejor y más limpia que la nuestra..."¹³

¹¹ Asimov, Isaac. "El conflicto evitable". *YO ROBOT*. Nueva York: Bantam Books, 2004. P.224

¹² Asimov, Isaac. *Robot e Imperio*. Londres: HarperCollins, 1996.

¹³ Asimov, Isaac. *YO ROBOT*. Nueva York: Bantam Books, 2004. PXI

Desarrollos de las "Tres Leyes de la Robótica":

Una discusión basada en la serie de robots de Asimov

6. Bibliografía

1. Asimov, Isaac. "Evidencia." YO ROBOT. Nueva York: Bantam Books, 2004.
2. Asimov, Isaac. "Guarida." YO ROBOT. Nueva York: Bantam Books, 2004.
3. Asimov, Isaac. "Pequeño robot perdido". YO ROBOT. Nueva York: Libros Bantam, 2004.
4. Asimov, Isaac. "Corre alrededor." YO ROBOT. Nueva York: Bantam Books, 2004.
5. Asimov, Isaac. "Sueños de robots". Sueños de robots. Nueva York: Ace Books, 1987.
6. Asimov, Isaac. Robot e Imperio. Londres: HarperCollins, 1996.
7. Asimov, Isaac. "El conflicto evitable". YO ROBOT. Nueva York: gallo Libros, 2004.
8. Asimov, Isaac. "Tú eres consciente de él". El robot completo. Londres: Granada, 1982.
9. Portelli, Alessandro. "Tres leyes de la robótica, Las: leyes del texto, leyes de producción, leyes de la sociedad". Estudios de ciencia ficción 7(2):150-156. Julio de 1980.