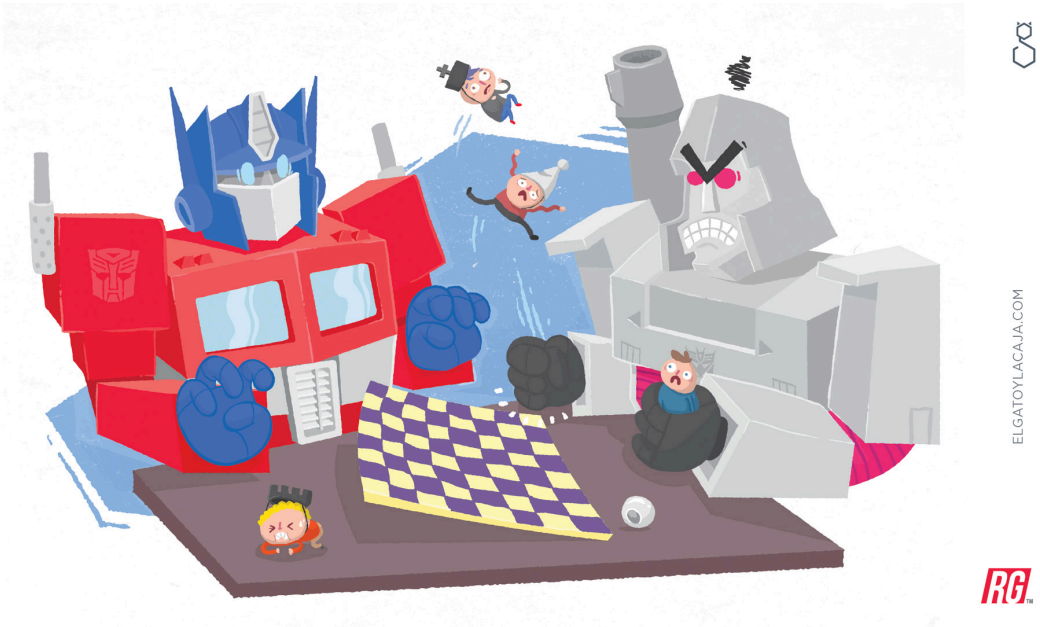


17/03/2016

Hacia una moral Transformer

TXT ALEJANDRO REPETTO IMG ELOY GALERATO

¿Cómo programás un auto robot? ¿Para qué un ingeniero necesita un filósofo?

A veces siento que mi vida se volvió una gran sesión de terapia de grupo que siempre arranca en 'Hola, soy Alejandro y me obsesiona hacer autos que se manejen solos'.

Desde que empezó la movida de los autos autocomandados (nombre horrible, si los hay), la gente ve más cerca un apocalipsis a lo Terminator y empieza a exigirle soluciones a los robots para cosas cuyo problema ni los humanos terminamos de entender.

El problema de los problemas me toca de frente cada vez que le grito a la licuadora por su incapacidad de detenerse en el punto justo entre un licuado con grumos y

uno demasiado acuoso, pero se pone más interesante con el surgimiento de un planteo de moral robótica que me parece de lo más absurdo, palabra que intento no usar apresuradamente y que voy a tratar de justificar.

El ensayo es más o menos así:

Supongamos que viene un micro con 10 bebés, 15 gatitos y, por qué no, un mimo, bajando una montaña, en una vía de doble circulación, con la montaña de un lado y el precipicio del otro. Supongamos también que de frente (subiendo) viene un auto que se conduce solo, llevando a un señor de unos 40 años. Justo al momento del cruce, por una falla en el micro, el auto tiene dos posibles modos de acción: A) obliga al micro a tirarse al precipicio y se salva; o B) se tira al precipicio – asesinando a su conductor/pasajero – y salva a los bebés, los gatitos y al mimo.

Si mantenemos la hipótesis de que en el micro va un mimo, la decisión es fácil. Ahora, si lo sacamos de la ecuación la cosa se torna más oscura. **¿Tiene el robot que matar a su dueño o a la persona que lleva, siguiendo la lógica del mal menor? ¿O debería asegurar la vida de sus propios pasajeros ante todo?**

Esta duda moral que se plantean los estudiosos (?) del tema tiene una falla enorme: juzgan al robot por cosas que los humanos no podemos siquiera discutir. O, en otros términos **¿por qué le vamos a poner 'moral robot' a la ejecución robot de la moral humana?** Sí, llegó ese momento terrible en que la ciencia y la ingeniería dejan de ser conceptos abstractos, lejanos, perfectos y mágicos y se enfrentan a un mundo real, con personas reales, gatitos reales y mimos reales, y, mal que nos pese, le vamos a tener que tirar un manotazo a la filosofía a ver si nos da una mano, cosa que, siendo uno ingeniero, lo hace sentir un cachito sucio. Pero bueh, qué sé yo, mal que nos pese, si escarbás y separás la paja del trigo (o el discurso meta discurso hiper recursivo paja del trigo), la filosofía está buenísima.

La pregunta que hago primero es: ante la misma situación, pero siendo el auto conducido por un humano, ¿qué haría el humano?

La respuesta es amplia. En primer lugar, porque **el humano ni siquiera podría evaluar quién va en el otro auto.** Luego, porque **el humano no podría realizar ni por cerca un cálculo de daño en una fracción de segundo, como para entender cuál es el mal menor**. Tercero, porque **el humano, por más**

entrenado que esté, muy probablemente siga su instinto de preservación y supervivencia (respuesta prácticamente automática, que no pasa por la razón, y que incluye untar al mimo, a los bebés y a los gatitos contra la ladera de una montaña).

Entonces, vuelvo al comienzo de la cuestión: ¿por qué le pedimos al robot que cumpla con una moral que ni siquiera nosotros podemos solucionar? Lo bueno es que, como si este problema fuese poco, agrego algo importante: en el caso de los robots nos podemos poner todos de acuerdo respecto a su accionar y decidir, en frío, pensando en el bien común, cómo debería actuar. Un ecuánime y calculador robot kantiano que valora la vida de mimos como si fuesen personas reales. ¿Distópico? Tal vez, pero el monorriel de la historia va directo en esa dirección.

Como la ingeniería me pesa, hablar de ciencia y de filosofía no me alcanza y necesito ensayar algún tipo de solución, que el auto no se maneja con teoría y discurso (Ingeniería wins, flawless victory). Podríamos hacer una ley que indique que el robot que conduce debe evaluar el daño y minimizar el riesgo humano. Podríamos hasta ponerle valores en los cuales prevalezcan la vida de los más jóvenes (bebés), luego gatitos (todos amamos los gatitos) y, al final de la cadena, los mimos. Inclusive, y esto es sólo una idea para abrir el debate, podríamos programar al robot para que si ve un mimo, le entre derecho a la altura del esternón.



Aká, sho, con mi atropellador de mimos

El problema sigue siendo el mi(s)mo, y es el enfrentamiento parejo entre el nosotros que piensa en términos de lo que haría y el nosotros que haría, posta. Entonces, la discusión pasa a tener que ver con nuestra sensación de control, de

libertad y hasta de intimidad sobre una decisión. ¿Es menos mía la decisión si la programo en un algoritmo para que el auto desacomode al mimo que si soy yo el que gira el volante para agarrarlo de lleno? ¿Voy a tener pesadillas con esa última línea de código que hizo que le pasara por encima al carrito de bebé para evitar la manifestación de ancianas? ¿Que estas preguntas sean difíciles e incómodas, implica que tenemos que renunciar a la idea de vehículos autocomandados? Total, tan buenos no pueden ser.

O capaz que sí, capaz que los vehículos autocomandados tienen infinitas ventajas, entre ellas:

– la reducción de la siniestralidad de hasta un 90% –90 de cada 100 accidentes que suceden tienen origen en la estupidez natural, cosa que la inteligencia artificial arreglaría–;

– la optimización del consumo energético –ahorro de frenado, optimización de velocidad crucero, etc.–;

– balanceo de carga en vías de circulación –si los robots se ponen de acuerdo, se avisan por dónde van y no taponean la Panamericana dirección Buenos Aires a las 8AM–;

–la no-necesidad de estacionamiento –el auto se maneja solo y, mientras no lo usamos puede dar servicio a otras personas (Uber Uber (?)) o irse de fiesta con sus otros amigos autos– se calcula que en un futuro de movilidad autónoma absoluta, con carsharing integrado se podría reducir la cantidad de autos al 20%.

Es acá donde, de nuevo, un ingeniero se pone profundamente incómodo (o profundamente pragmático, lo que pase primero). Es en este punto en el que la tecnología, una vez más, se hace carne y es el hombre que hace a la máquina el que, en una suerte de mancha venenosa moral, le transfiere sus prioridades y su noción de bien y mal a un sistema para que ejecute decisiones que nosotros no podemos tomar. Decisiones que, en una de esas, son hasta mejores que las que tomaríamos nosotros, limitados por toda esta urgencia de seguir existiendo, por más bebés, ancianas y mimos con los que haya que embadurnar el paragolpes para lograrlo.

Avanzar no solamente implica respuestas, inventos, iPads, televisores enormes que todo el mundo ve y Google Glasses que nadie, nadie, nadie usa y te hacen lucir más

bien tirando a boludo; sino que implica preguntas. A veces las preguntas son nuevas, a veces son tan viejas que duelen. ¿Qué está bien y qué está mal? ¿Cuánto mal vamos a soportar en función de cuánto bien? ¿Quién corno decide cuál es el vector y cuánto vale cada gatito?

Filosofía, no te quiero meter presión, pero somos un montón de científicos e ingenieros esperando una respuesta. Siempre quisiste que te tomáramos en serio. Bueno, es tu momento. No nos falles ahora.

PD: en esta nota ningún mimo fue maltratado.

BONUS TRACK





Referencias

http://arxiv.org/abs/1510.03346

http://www.sciencedaily.com/releases/2015/06/150615124719.htm

http://www.scientificamerican.com/article/self-driving-cars-could-cut-greenhouse-gas-pollution/

http://www.computerworld.com/article/2486635/emerging-technology/self-driving-cars-could-save-more-than-21-700-lives-450b-a-year.html

elgatoylacaja.com/hacia-una-moral-transformer

Sumate en
eglc.ar/bancar