

# A vueltas con la Inteligencia Artificial

Félix Ares

Fue a mediados de junio de 2016 cuando supimos que el dueño de un coche Tesla S, equipado con piloto automático, había sufrido un accidente mortal. Enseguida empezaron a oírse voces en contra de lo que va a ser una realidad en muy pocos años: los coches que conducen solos.

El accidente se produjo no con un coche que conduce solo, sino con uno que lleva un sistema de ayuda a la conducción en circunstancias muy concretas; únicamente en ciertas carreteras y en buenas condiciones climáticas. Lo que hace el sistema es mantener el carril en una buena carretera y evitar colisiones. Cuando la empresa Tesla vende sus vehículos con autopiloto, primero dan un curso a los dueños explicándoles que es un sistema experimental —una beta pública; esto es, técnicamente acabada, aunque aún en pruebas, a la búsqueda de posibles fallos de diseño—, y que no deben dejar de tener las manos en el volante y de mirar la carretera en ningún momento. Incluso tienen sensores para comprobar que las manos están donde deben.

Cuando escribo estas líneas, el accidente todavía está en investigación, pero parece ser que un camión con un remolque de color blanco se cruzó perpendicularmente a la carretera. Ni que decir tiene que el coche está preparado para detectar obstáculos tanto con cámaras visuales como con radar (en concreto, radar láser: LIDAR). Pero se dieron una serie de circunstancias fortuitas que llevaron a la catástrofe. La primera es que la iluminación era tal que el remolque se confundía con el cielo tanto para el piloto automático como para el ser humano, y lo más extraño es que el camión estaba tan alto que la señal de LIDAR pasó por debajo sin detectar el obstáculo.

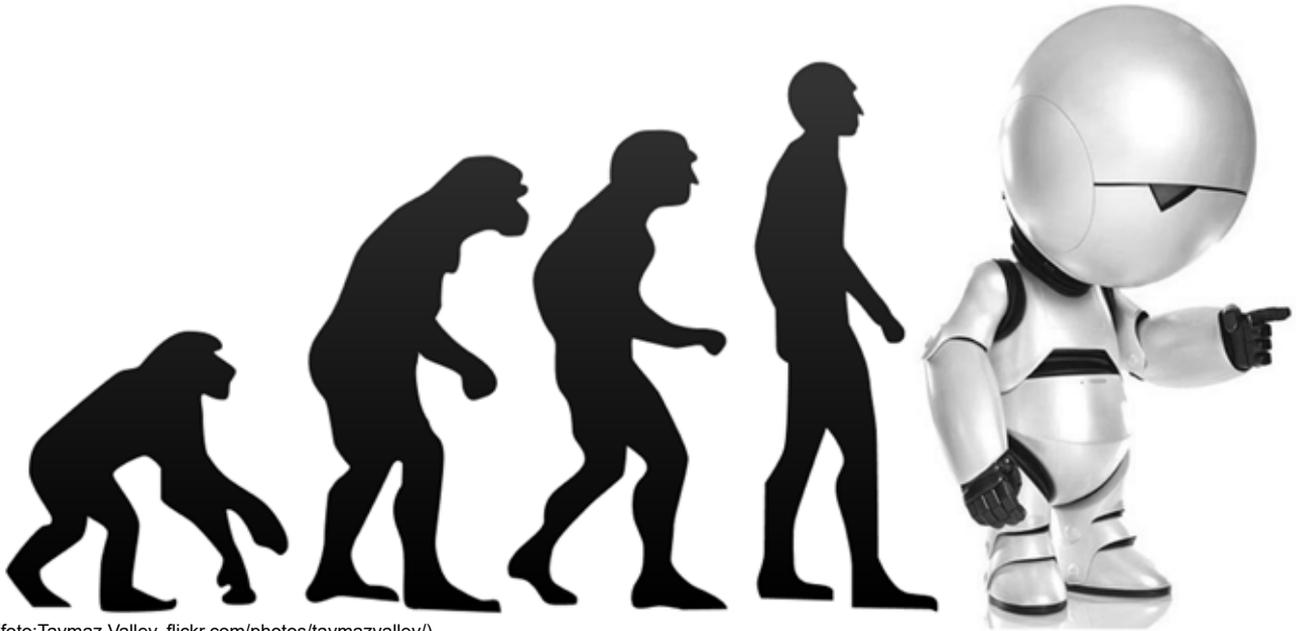
Y estas circunstancias tan singulares produjeron el desastre. Y, claro, las voces alarmistas se pusieron a despotricar. No solo los periódicos, también la gente de la calle. Escribí una de mis columnas semanales en el *Diario Vasco* sobre este tema, y pedí la opinión a varios conocidos. Una respuesta bastante habitual es que ellos no se subirían nunca a un coche sin conductor. Al margen de mi pequeña encuesta personal, sin ningún valor científico, está el hecho mucho más significativo de que las acciones de Tesla, que habían subido un 500% desde 2013 —y su tendencia era al alza—,

al día siguiente del accidente bajaron un 2,5%.

Intenté argumentar con algunos datos. Los coches con autopiloto activado llevaban recorridos cuando se produjo el accidente 209 millones de kilómetros. En Estados Unidos por término medio hay un accidente mortal cada 96 millones de kilómetros. Es decir, los datos no son preocupantes. Con un solo caso no es posible hacer estadísticas ni sacar conclusiones, pero las primeras impresiones son que el autopiloto es más seguro que el conductor humano. Insisto en que con un solo caso de accidente es imposible obtener más que una vaga impresión.

Mi siguiente argumento es muy simple: la grandeza de la ingeniería es que aprende de sus errores. Por poner un ejemplo, hoy volar es un sistema sumamente seguro de transporte, cuando los primeros vuelos comerciales no eran así. Esa seguridad se ha conseguido porque cada accidente se estudia y se tratan de encontrar soluciones para que no vuelva a ocurrir. No me cabe duda alguna de que los ingenieros tendrán en cuenta este accidente y lo incorporarán a los sistemas físicos o a los algoritmos de decisión para que no vuelva a ocurrir. Tal vez un cambio de la disposición del LIDAR, tal vez aumentar la sensibilidad al cambio de color en ciertas condiciones de iluminación, etc. No sé cuál será la solución, pero sí sé que se buscará y se encontrará. Y esa es la enorme ventaja de lo automático sobre lo humano. El sistema visual humano es como es, y si no ve un camión blanco contra un luminoso cielo de Florida, no lo verá nunca; mientras que los sistemas automáticos se corregirán, a los humanos no se nos puede corregir, al menos de momento.

En los artículos de prensa que pude leer *a posteriori* había un batiburrillo de cosas, donde se mezclaban los drones con soldados automáticos. Muchas veces confundían los drones pilotados remotamente con los drones autónomos. Y surgían preguntas de las que ya he hablado alguna vez. Una de ellas era que en algún momento los coches autónomos se encontrarían en una situación en la que tendrían que elegir entre atropellar, y probablemente matar, a —digamos— cinco peatones, o chocar contra una pared y matar al dueño del coche autónomo. Lógicamente, desde un punto de vista social, es mejor que muera uno a que mueran cinco; pero la respues-



(foto: Taymaz Valley, flickr.com/photos/taymazvalley/)

ta de los participantes en algunos foros me resultó difícil de digerir. Podríamos reducirlo a la siguiente frase: «yo jamás me desplazaré en un coche que pueda decidir matarme a mí para salvar a cinco». El egoísmo es evidente y pone sobre la mesa el tema de que los algoritmos de seguridad —y si ustedes quieren, los llamamos algoritmos éticos— deben ser consensuados socialmente y han de ser, si no imposible —que eso no existe—, al menos muy difíciles de cambiar. Es más, creo que habrá que hacer una legislación en la cual que un usuario cambie el algoritmo de su coche, para que este prefiera matar a cinco antes que a él, sea delito.

Pero, además, estamos pecando de simplicidad. La idea de los coches autónomos no solo es que implique la capacidad de conducir como un ser humano, sino que va mucho más allá. Lógicamente, no tardando mucho, los coches se comunicarán unos con otros. Pongamos un ejemplo: en un cruce de calles, aunque no haya visibilidad, los coches sabrán dónde están y cumplirán los protocolos para evitar los accidentes. Si a esto le añadimos sistemas de navegación por satélite con precisión de pocos centímetros y la quinta generación de telefonía móvil que permitirá la llamada *Internet de las cosas* (IoT, acrónimo de *Internet of Things*), tendremos sistemas de conducción autónomos que no dependerán únicamente de sus medios, sino que la carretera, los semáforos y otros coches, e incluso las farolas de la calle, les hablarán. Todo ello, bien gestionado, hará el tráfico infinitamente más seguro. Parte de estas soluciones se podrán adoptar si el conductor es humano, pero su lentitud en tomar decisiones

será menos eficiente que la decisión automática.

¿Cuándo tendríamos que autorizar los coches autónomos? Para mí, la respuesta es muy clara: cuando el número de accidentes que produzcan sea notablemente inferior al de los humanos. Con notablemente quiero decir que no basta un único ejemplo, como el del accidente del Tesla S en Florida. Tiene que haber muchos más datos.

Y una vez visto esto, muchas de las preguntas tienen una respuesta inmediata. ¿Cuándo habría que permitir soldados autónomos que sean capaces de disparar? Mi respuesta es que no debieran existir soldados, ni humanos ni autónomos, que disparen a matar, pero estamos en el mundo que estamos. La respuesta entonces es muy sencilla: cuando el índice de errores de los autónomos —matar a una persona que no era el objetivo— sea inferior al de los humanos.

¿Cuándo deberíamos permitir drones autónomos haciendo cosas como llevar paquetes dentro de las ciudades? Cuando el índice de accidentes sea menor que el que producen los humanos... y así sucesivamente. El momento en el que las máquinas, en una aplicación particular, cometan menos errores que los humanos, será el momento de dejar que ellas tomen o incrementen el control, si fuera conveniente.

Ha habido una cierta alarma, lanzada entre otros por el famoso astrofísico Stephen Hawking, de la que se sigue un mundo catastrófico, al más puro estilo de *Terminator*: la inevitable lucha entre la máquina y el ser humano.

Sinceramente, creo que es una visión catastrofista y sin fundamento. La razón esencial es que a casi nadie (hay al-

**«Yo jamás me desplazaré en un coche que pueda decidir matarme a mí para salvar a cinco».**

gunos a los que sí) le interesa una máquina con consciencia de sí misma y con problemas metafísicos. Imagínense el siguiente escenario: usted, con su móvil, pide un taxi autónomo al estilo de los de *Desafío total* (*Total Recall*). Llega el coche —obviamente, sin muñequito parlanchín; aunque el coche sí que habla:

—Buenos días, bienvenido al servicio de coches Eureka. Según mis datos, vamos a la fábrica de lechugas «La Verdura Fresca».

—Sí, así es.

Usted se monta y entonces el coche empieza a charlar.

—¿Sabe? Estoy frustrado.

—¿Por qué?

—Pues porque yo podía haber sido astrofísico o cirujano, y ya ve, he terminado de taxista. Mi *hardware* y mis programas de aprendizaje profundo (*deep learning*) son idénticos a los de mi compañero de promoción, DaVinci5, que es un cirujano de gran prestigio; o a Hawking3, que se dedica a analizar las fotos multiespectrales de varios telescopios espaciales. O podría invertir en Wall Street como mi colega Bolonia17. Eso me frustra mucho.

Supongo que la primera vez que nos encontramos con un taxi así nos haría gracia; pero si se repitiera varias veces, pediríamos que se callase y que se limitase a conducir, y a lo sumo, a ofrecernos música o películas.

Y ese es el tema. No nos interesa construir máquinas que sean como los humanos. Nos interesa hacer máquinas que lo hagan mejor que nosotros en campos especializados. Por ejemplo, cuando se construyeron las grúas, no se intentó hacer una máquina con la fuerza de un humano. Se hizo una máquina que era capaz de mover pesos mucho mayores que lo que nosotros podíamos. Cuando se hicieron los aviones,

no se intentó que fueran un pájaro; se hicieron máquinas que volaban mucho más rápido que los pájaros y que nos llevaban en su panza. Cuando quisimos resolver ecuaciones diferenciales, creamos programas al estilo de *Mathematica*, que lo hace mucho más rápido que nosotros. Cuando quisimos construir una biblioteca universal, no hicimos un gran edificio con millones de libros; creamos enormes bases de datos, al estilo de Google, a las que podemos acceder desde cualquier lugar del mundo.

Eso es lo que necesitamos y eso es lo que construiremos: máquinas que serán mucho mejores que nosotros en campos específicos. Mejores cirujanos, mejores médicos diagnosticadores, mejores constructores, mejores aspiradores, mejores descubridores de fármacos, mejores inventores de algoritmos, mejores programadores, mejores profesores... Es incluso posible que en un futuro veamos anuncios similares a este: «Compre un Domo23, un robot para los trabajos del hogar, SIN AUTOCONSCIENCIA».

Porque en ningún momento parece ser que el que tengan autoconsciencia sea una ventaja para nosotros. Y si no tienen autoconsciencia, todas las ideas de la superación de los humanos por las máquinas no dejan de ser una película de ciencia ficción. Desde siempre, desde la primera máquina que inventamos —¿una piedra para sacar el contenido interior de un fruto seco?— han sido, en su tarea específica, mejores que nosotros. Y nunca hemos temido que las piedras nos sustituyeran, o que eso fuera grave. La máquina *arco* permitía mandar flechas más lejos que nosotros; la máquina *escritura* permitía enviar los mensajes más lejos, tanto en el espacio como en el tiempo... y nunca nos hemos sentido amenazados. ¿Por qué ahora?

