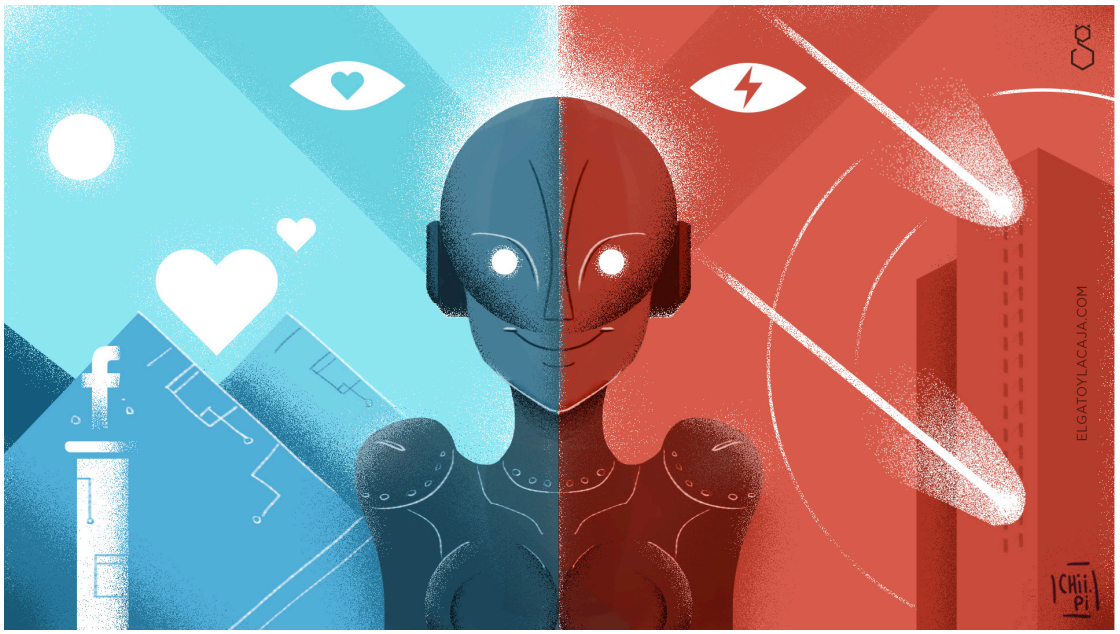

11/12/2017

Ni apocalipsis ni utopías robóticas

TXT PABLO BARRAGÁN IMG PÍA CASCELLA

¿La Inteligencia Artificial va a llevarnos a una utopía o a la extinción? ¿Tiene sentido la pregunta o es puro humo?

Pensás que estás en la cima del mundo y, de repente,
te tiran el Armageddon por la cabeza.
Terry Pratchett y Neil Gaiman, Buenos Presagios (1990)

Nos encanta el futuro. Es un lugar enorme y lleno de posibilidades donde la humanidad guarda de todo: desde los cuatro jinetes del apocalipsis hasta los androides que sueñan con ovejas eléctricas; desde las predicciones de Nostradamus hasta los motores warp y las patinetas flotantes. Nos fascina porque es libre, porque nos deja atribuirle todos nuestros ideales de gloria o de condena. Pero

cuando se mezcla futuro con ciencia, la imaginación no alcanza y necesitamos datos concretos. **Hoy existe una pelea mediática, con pretensión de ser científica, entre dos posturas completamente opuestas acerca del futuro, el desarrollo y la investigación de la Inteligencia Artificial**. El problema con esto es que, cuando a 'futuro' y 'ciencia' les agregás 'mediático', se genera un cóctel que les saca espuma por la boca a los expertos en el tema.

Empecemos por aclarar una cosa: **el término 'inteligencia artificial' (IA) es impreciso si lo usamos así nomás**. Cuando se habla de IA parece que incluye a Terminator, el software de aprendizaje del motor de búsqueda de Google, las fembots de Austin Powers, los vehículos autónomos, HAL 9000 y los algoritmos que filtran el inicio de Facebook. Este revuelto de máquinas inteligentes, software y robots asesinos contribuye a que la noción de IA sea poco clara y a que cada vez que se habla de ella se genere bastante sarasa.

Lo peligroso del asunto es que este humo llena páginas enteras en algunos medios de comunicación, sobre todo, cuando se lo relaciona con un ingrediente más: el futuro de la humanidad. El apocalipsis vende y mucho.

A los efectos de esta nota, **simplifiquemos el término IA para referirnos al objeto de estudio de una ciencia que se ocupa de 'hacer máquinas que hacen cosas que podríamos considerar inteligentes'**.

Si se lee el tema por arriba pareciera que hay sólo dos posturas respecto del futuro de las IA. Una visión apocalíptica capitaneada por Elon Musk, CEO de SpaceX y creador de Tesla Inc. (entre otras), que intenta convencer a las mesas chicas de la política y la ciencia con la idea de que **el desarrollo desregulado de las máquinas traerá la aniquilación de la raza humana**. **La otra postura, comandada por Mark Zuckerberg** –creador de Facebook y única persona con dos mil millones de amigos–, plantea que **hay que dar libertad al desarrollo de la IA** y que quienes digan lo contrario son unos 'irresponsables'. O sea, la pelea es entre dos multimillonarios con mucha prensa que **discuten si hay que hacer o no leyes para controlar a los científicos que investigan sobre IA.** Y hay una tercera postura que dice que los dos se tienen que calmar un montón.

Del lado apocalíptico, Musk no está solo: lo acompaña Stephen Hawking –el de los agujeros negros y la teoría del todo–. Juntos forman parte de *Future of Life*, una fundación creada para el desarrollo responsable de las investigaciones sobre IA. Advierten que la humanidad estará en peligro de extinción a menos que se apliquen 24 principios de control para el desarrollo adecuado, sustentable y seguro de las IA que, por supuesto, crearon ellos. Pero aunque su postura extremista sobre

el fin de la humanidad llame la atención por lo desesperada y colifata, contrapesa con sus propuestas que, en definitiva, tienen mucho de sensatas.

La organización divide los 24 principios en tres pilares: investigación, ética y asuntos a tratar a largo plazo. Entre otras cuestiones, disertan sobre la necesidad de crear un desarrollo cooperativo entre equipos de investigación, regular la transparencia de los fondos de inversión en estos proyectos, marcar cláusulas obligatorias que aseguren el beneficio de la humanidad en las investigaciones, determinar la responsabilidad legal de quienes desarrollan y utilizan IA ante posibles accidentes y un último punto importante: **que ninguna IA, por más avanzada que sea, pueda tomar una decisión sobre una persona sin la aprobación de un humano especializado en el área en cuestión**. Por ejemplo, si usáramos una máquina para evaluar objetivamente la ilegalidad de un acto realizado por una persona, su juicio debe ser avalado por un juez. Si se revisan los 24 principios, **ninguno está en contra del desarrollo de las IA** sino que plantean que, como todo, debe hacerse de manera responsable.

Mark Zuckerberg representa la postura opuesta. Desde su casa en Palo Alto (California), mientras cocinaba una barbacoa, mandó un mensaje por Facebook en vivo a todos sus seguidores en el que decía que los que agitan con que el futuro se viene malo y feo "son unos irresponsables", y que **no hay que interferir con el desarrollo libre de las IA**. Pero ¿en qué se basa para decir eso sobre Musk y Hawking? ¿Qué razones tiene?

La razón principal de Facebook para defender su postura es que tiene su propio equipo de investigación de IA trabajando en varios proyectos. Diseñaron, entre otras cosas, un software que aprende cómo te movés en la red social para """administrar mejor tu inicio""""" (COMILLAS COMILLAS) y poder mostrarte las fotos de la persona que más stalkeas o los posts que pueden hacerte reaccionar más fácil (sí, la peor IA para olvidarte de tu ex, de cuando tenías pelo o de tus compañeros de secundaria). Este tipo de herramienta logra que te quedes más tiempo mirando la pantalla, y cuanto más estés en la pantalla, más aprende, más te hace quedarte, más publicidad vende, más tiempo estás en la pantalla y tiempo y publicidad y plata y pantalla y así. (¿alguien gritó 'conflicto de interés'?). Por

supuesto, esto está muy lejos de ser un robot que nos va a llevar a la destrucción de la humanidad, pero hay personas que piensan que tampoco está tan bueno que existan estas cosas.

Una de esas personas es **Tristan Harris, ex especialista en ética del diseño de Google**. *¿¿Ex qué??* Su trabajo consistía en decidir **cómo persuadir a los usuarios y dirigir su atención para captarlos por más tiempo de manera 'ética y correcta'**. O sea, que no salgan fuera de los mejores valores de la sociedad; que no sea ilegal o inmoral. Harris llama la atención sobre estos algoritmos y plantea que "hay que alinear los objetivos del persuasor y del persuadido" y aclara que "actualmente hay una habitación con pocas personas que deciden en qué van a gastar su tiempo varios miles de millones de personas. **Deberíamos concentrarnos en las IA de hoy, no en las de mañana,** y hacer una revolución en la tecnología". La maldad es humana; la herramienta, divina.

El grupete investigador de Zuckerberg no sólo se dedica a engancharte a la pantalla. El último proyecto que desarrollaron se trata de un sistema que utiliza realidad virtual (VR), software de aprendizaje y cámaras 360 para conectar a personas en zonas de desastres naturales con organismos de ayuda humanitaria y, de esta manera, acelerar y efectivizar los rescates. La investigación se realizó en conjunto con *American Red Cross* y este modelo se presentó en Puerto Rico tras los desastres que generó el Huracán María en septiembre. Sin embargo, la presentación recibió críticas muy duras ya que se mostraba al CEO de Facebook y a su jefa de desarrollo de VR, Rachel Franklin, personificados por un avatar de caricatura y festejando la tecnología mientras hacían un tour sobre los destrozos. Si bien la idea es buena y ya se habló con la Cruz Roja Internacional y UNICEF para que tengan acceso a este tipo de tecnología a través de Facebook, los comentarios del video llegaron al punto en que ciudadanos de Puerto Rico escribieron que Zuckerberg era un "multimillonario sin corazón", ante lo cual el CEO tuvo que pedir disculpas en los mismos comentarios por el importante error de comunicación.

Otro ejemplo que viene de la mano de Mr. Facebook y su equipo se puede leer en un *paper* donde explican que lograron desarrollar una IA capaz de aprender a mejorar métodos de negociación sólo a través de conversaciones. La IA fue capaz

de fingir que se interesaba en todo lo que su interlocutor tenía para ofrecer cuando en realidad sólo quería una cosa específica. De esta manera, hizo que el interlocutor pensara que había ganado en la negociación al hacerle ceder de obtener el total de los objetos, cuando en realidad ella fue quien consiguió lo que buscaba, que era solo uno de estos objetos. Lo interesante es que todo lo aprendió solita, solo con hablar y hablar. En castellano sería: si ella tenía un libro y necesitaba un sombrero, y su interlocutor tenía un sombrero, tres pelotas y dos lápices, le decía: "te doy mi libro por todo lo que tenés". Cuando el interlocutor se negaba, le decía "bueno aunque sea dame el sombrero y te doy mi libro", y lo conseguía. O sea, **la IA aprendió a engañar a los interlocutores para ganar la negociación**. De tal palo, tal inteligencia artificial.

Esta investigación también explica que estos *bots* desarrollaron su propio idioma durante el proceso, y acá es cuando entramos en la gran máquina generadora de humo. Este hecho despertó a los grandes fantasmas de un futuro distópico donde una suerte de robots Frankenstein cobraban vida, y por varios medios salieron a decir que por ese motivo Facebook entró en pánico y apagó el experimento. Claramente esto no fue así. Desde el equipo aclararon que "no hay que alarmarse"; explicaron que "es normal" que los bots desarrollen sus propias subrutinas de lenguaje y que, de cualquier manera, sólo desarrollaron un idioma comprensible entre ellos a los efectos de cumplir su función, y que lo hicieron porque no se les había determinado que debían realizar la conversación en un lenguaje comprensible para humanos.

Como se puede ver, hasta ahora nada de rayos láser, ni de explosiones, ni de robots asesinos. Entonces, **¿qué pasa con los apocalípticos? ¿Qué razones dan ellos para no confiar en las IA?** Stephen Hawking le explicó a la BBC que el peligro radica en que **el humano nunca podría competir con una IA por las diferencias enormes en la capacidad de procesamiento de información y en la velocidad para adaptarse a nuevos escenarios**. En ese caso, las IA fuera de control podrían reemplazarnos con facilidad en cualquier tarea y dejarnos fuera del esquema. Apa, ¿y eso?

Aunque venga de Hawking, esta visión sobre un desarrollo desbocado de las IA es demasiado reducida. Ese tipo de escenario no depende tanto de que las máquinas sean más inteligentes que los humanos, sino de que a nosotros nos falta aprender mucho todavía. Noriko Arai, la líder del proyecto Todai Robot, construyó un robot que pudo pasar el examen de ingreso a la universidad de Tokyo con un resultado mejor que el 80% de los alumnos de Japón "sin entender una palabra del examen". Esto es literal: el robot no entendió nada, usó el programa de aprendizaje y ejecución con el que fue diseñado para reconocer la información que le pedían las preguntas y asociar las palabras con una lógica matemática propia que le permiten recuperar de su base de datos inmensa y siempre al día las respuestas más probables. Por supuesto, falló en comprensión de texto.

Noriko Arai plantea que necesitamos una nueva forma de enseñar a los jóvenes y cambiar el modelo matemático y de aprendizaje 'de memoria': "No podemos procesar mejor y más información que los robots, pero ellos no pueden entender lo que hacen. Necesitamos un cambio en el sistema educativo". Lo que Todai Robot Project demuestra es que si tratamos de competir en un examen de matemática contra una calculadora con esteroides, entonces siempre vamos a ser superados. Lo que explica es que la revolución educativa viene con fomentar el tipo de inteligencia que tiene que ver con entender realmente las cosas y no sólo recordarlas de memoria.

Mientras tanto, Elon Musk insiste en aprovechar su posición cercana a la mesa chica del gobierno de los Estados Unidos con el fin de lograr una regulación del desarrollo de las IA. En febrero Musk se presentó en la *World Government Summit* en Dubai y, si bien dejó de lado el problema ético sobre las decisiones que puedan llegar a tomar las IA en situaciones de riesgo, sacó a relucir el problema más próximo en cuanto a regulación: **"Los vehículos autónomos están más cerca de lo que se piensa y estos van a dejar sin trabajo a muchísima gente, vamos a tener que encontrar nuevos roles para esas personas o el gobierno tendrá que darles un sueldo"**. También lo llamaron como formador de opinión para participar en una conferencia frente a gobernadores en la *National Governors Association* de Estados Unidos este año. Ahí dejó bien claro que "las IA son un

riesgo distinto a las drogas o los accidentes, que son graves pero no como esto, porque **las IA son un riesgo a la humanidad en su totalidad**". Bueeeno, ¿será para tanto Elon? A ver qué dicen otros expertos.

Stuart Russell, científico pionero en la investigación de IA, explica que **los robots podrían ser un riesgo sólo si se los programa con un objetivo único y se les dice que no deben dejar que nada interfiera en su misión**, ya que aprenderían cada vez más sobre esa función y la optimizarían superando a cualquier humano en la misma tarea. Por ejemplo, si le decimos a un robot que se ocupe de cebar mate, lo hará cada vez mejor hasta dejar atrás al mejor cebador del mundo (como hizo Alpha Go con el mejor jugador de Go); si además le decimos que no debe dejar que nadie ni nada interfiera en su función, y siendo que los únicos que pueden apagarlo y evitar que cumpla su misión son los humanos, ¿eliminará a los humanos mientras duermen y cebará mate hasta el fin del suministro de yerba? ¿Vamos a morir por un robot cebador cebado? No, no tan rápido. En este ejemplo, el robot sólo tiene una función que es cebar mate y no se le dio una entrada de información que incluya 'matar a los humanos si interfieren'. Sólo lograríamos un robot cebador perfecto que no aprendería por sí solo todo lo demás. Traemos este ejemplo para resaltar que hoy en día estas especulaciones son sencillamente conjeturas que nada tienen que ver con la realidad actual de las investigaciones IA.

Sin embargo, el debate es válido en un punto. Si se llegara a tener la capacidad de crear semejante tipo de IA, Russell propone que **deberíamos crear nuevas leyes de la robótica para que sean compatibles con la vida humana**. Lamentablemente, considera que todavía no estamos listos como sociedad para eso. Estas nuevas leyes, cuyos algoritmos no están formalizados, deberían incluir que el robot no sepa exactamente su objetivo, que su propósito sea beneficiar siempre a la humanidad y que sus conductas sean aprendidas de observar las conductas humanas. De esta manera, aprenderá a definir aquello que está socialmente 'bien' de lo que está 'mal' y entonces, si su función es cocinar y se queda sin comida, no cocinará a la mascota tras evaluar su valor nutricional. Además, plantea que el robot debe siempre poder ser apagado y que debe saber que si es apagado es por algo superior a su objetivo, y entonces así logre aprender

de esa situación. Esto es muy importante para que nunca trate de evitarlo asesinando a la humanidad mientras duerme. Por lo pronto, hay que reemplazar las leyes de la robótica. Asimov, estarás siempre en nuestros androicos corazones, pero tus leyes ya no.

Si no tenemos nada más que algunos algoritmos polémicos en cuanto a ética empresarial y conjeturas de posibles robots demasiado inteligentes como para tolerar nuestra existencia, ¿por qué insisten en el apocalipsis? Musk asegura que tiene acceso a la más moderna tecnología de IA y contraataca con que deberíamos tener miedo a lo que se viene. Insiste en que **se deben crear regulaciones proactivas para el desarrollo de las IA y no reactivas cuando ya sea demasiado tarde**. Todo esto lo sostiene en que mantiene su conocimiento sobre IA al día a través de inversiones millonarias en tres proyectos de investigación: Neuralink, DeepMind y Vicarious. Las tres empresas trabajan en la investigación y desarrollo de posibles aplicaciones de la IA en robótica y software. Sin embargo, dato curioso, el otro gran inversionista de Vicarious es Mark Zuckerberg. Entonces, si reciben la misma información, ¿por qué tienen ideas tan diferentes? Porque ambos se basan en creencias y no evidencias.

Para Luciano Floridi, profesor de filosofía y ética de la información en la Universidad de Oxford, dada las capacidades de desarrollo de IA actuales, lo que proponen ambas posturas son escenarios que no son plausibles. Para él, "el riesgo más serio sería que usemos mal nuestra tecnología inteligente, en detrimento de la mayoría de la humanidad y del planeta entero". Si tenemos en cuenta que la IA es tecnología que depende exclusivamente de la información que se ingresa en ella para que trabaje, el cebador de mate que usamos de ejemplo ceba y no mata porque le dijeron que cebe y no que mate. Por lo tanto, sea el futuro una utopía o una distopía, Floridi concluye: "Somos –los humanos– y seguiremos siendo, en cualquier futuro predecible, el problema, y no así nuestra tecnología".

Lo que buscan Musk y Zuckerberg es usar su influencia para lograr lo que ellos consideran mejor desde sus propios puntos de vista. Conflictos de intereses aparte, **todavía estamos lejos de que exista una posibilidad real de un apocalipsis robótico o incluso una utopía tecnológica**. Además, llegado cualquiera de los

casos, la responsabilidad es siempre del humano que se esconde detrás del desarrollo e implementación de esas inteligencias. Las dos posturas son extremistas y se enfrentan en un momento de la historia en que el futuro que soñaban las grandes plumas de la ciencia ficción se vuelve cada vez menos ficción y más ciencia. El debate sigue abierto y nos hace entender que necesitamos un mejor tratamiento de las noticias científicas, mejores preguntas para todas las partes y buen aporte desde la ciencia que aclare hacia el público general el humo que se creó alrededor del potencial real y aplicable de la IA.
Skynet, no te tenemos miedo. Por ahora.

elgatoylacaja.com/19995-2

Sumate en
eglc.ar/bancar