

# Sobre los riesgos de la Inteligencia Artificial

7 de enero de 2016

**Comentario de Lampadia**

Una vez más queremos difundir información que nos permita crear nuestro propio criterio la revolución tecnológica que ya envuelve a la humanidad, especialmente sobre la tecnología que amenaza ser la más disruptiva, la Inteligencia Artificial (IA). (Ver nuestro Repositorio: [Revolución Tecnológica](#)).



En esta ocasión compartimos dos artículos del Financial Times, el primero sobre el reto personal de Zuckerberg (Facebook), de crear un asistente personal con IA y al mismo tiempo que rechaza los riesgos más comentados y, el segundo, un diálogo abierto sobre los riesgos de un desarrollo descontrolado de la IA. Como parte de esta información incluimos además dos videos (en inglés) transcritos al español.

El tema de la revolución tecnológica es muy difícil de seguir en su cabal medida y por lo tanto, es igualmente complicado entender sus implicancias en las políticas de desarrollo de los países. Como hemos visto antes, su [impacto en el empleo puede ser muy disruptivo](#). Pero en el caso de la IA, sus impactos pueden traer profundos cambios en la vida de los seres humanos mismos. Amenazando incluso nuestra supervivencia. Las siguientes líneas nos ayudarán a poner en perspectiva este último aspecto.

## **Zuckerberg rechaza temores sobre tendencias destructivas de la Inteligencia Artificial (IA)**

Richard Waters y Hannah Kuchler en San Francisco y Murad Ahmed en Londres, 4 de enero 2016

**Financial Times**

Traducido y glosado por Lampadia



**Zuckerberg (Facebook) cree que la IA puede desarrollarse para que 'trabaje para nosotros'** ©Getty

Líderes tecnológicos como Elon Musk de Tesla Motors y el capitalista de riesgo, Peter Thiel han advertido que las tendencias destructivas de la IA pueden, algún día, amenazar a la raza humana. Pero Zuckerberg se ha impuesto el reto de construir este año su propio asistente personal con IA,

rechazando los temores de que la 'súper-inteligencia' se comporte incontrolable y disruptivamente, como improbable.

En un post de la web de su compañía [Facebook founder](#) dijo que ha decidido construir una máquina inteligente que controle las funciones de su casa y que le ayude en el trabajo. Lo comparó con Jarvis, la computadora inteligente de las películas de *Iron Man*.

Su reto de este año es el primero que involucra un código de software para testear tan importante desarrollo para el futuro de Facebook. La compañía ya tiene dos equipos trabajando en inteligencia artificial - uno que se enfoca en investigación pura y otro que ve la forma de aplicar IA a las redes sociales.

El reto también enfatiza la carrera de IA que se está desarrollando entre las compañías líderes de internet que tratan de crear las vías para que las personas puedan interactuar con las máquinas.

Facebook ha establecido tres laboratorios de IA, en Silicon Valley, New York y Paris y sus contrataciones de alto perfil incluyen a Yann LeCun, profesor de New York University, uno de los pioneros de 'conocimiento profundo' - una forma de IA que trata de imitar las capas de funcionamiento de neuronas del cerebro humano.

Zuckerberg dice que su trabajo lo lleva a temas técnicos profundos con los ingenieros de la compañía, y que construir su propia máquina inteligente traerá una nueva perspectiva en el trabajo.

Describió su proyecto personal como 'una simple IA', tomando un sistema al que pueda enseñar a reconocer su voz y a controlar todo en su casa - música, luces, temperaturas y otros.

También dijo que está interesado en usar reconocimiento de voz y rostros para identificar a quién estuviera en la habitación, por ejemplo, para ajustar la temperatura, pues él prefiere ambientes más frescos que su esposa.

Para ayudarse en su trabajo, Zuckerberg añadió que intentará enseñarle al sistema a convertir datos en representaciones visuales que pueda ver con su dispositivo de realidad virtual.

Hermann Hauser, un emprendedor científico de computadoras de Cambridge dijo que los avances recientes en algoritmos de aprendizaje de máquinas, el poder computacional y el acceso a inmensos volúmenes de datos, combinados permiten que los ingenieros de computadoras, puedan crear sistemas de IA. Pero el doctor Hauser dijo que le preocupaba que solo las más grandes compañías de Silicon Valley tuvieran acceso al nivel de información necesario para crear IA funcional, lo que concentraría poder entre esas compañías. "La gran preocupación que tengo es que el único camino para que esto funcione es tener acceso a grandes conjuntos de datos", dijo. "La única gente que tiene acceso a esto son de Apple, Google, Amazon y Facebook."

**Los grandes avances en IA han llevado a líderes tecnológicos a advertir que algún día, las máquinas inteligentes podrían poner en peligro a la humanidad.**

Zuckerberg calificó esto como "improbable ... y mucho menos factible que desastres debidos a la dispersión de enfermedades, violencia, etc.". Añadió: "Creo que podemos construir IA para que trabaje para nosotros y nos ayude".

Hasta ahora, el experimento de Facebook más avanzado con IA ha sido la prueba de un asistente virtual llamado M, que opera en el Messenger de Facebook y promete servir los deseos de llamar un taxi hasta organizar una fiesta de cumpleaños. También ha utilizado IA para describir fotos a usuarios ciegos.



**Ver video:** [Robots in the workplace](#)

**Comentario del Financial Times (Transcripción del video):**

**Andrew Hill (Editor de administración, Financial Times):** Los humanos contra las máquinas es una historia tan antigua como la historia de la automatización, ya sea en el campo o en las fábricas, con la invención de la nueva tecnología siempre viene la promesa de una revolución de la forma en que trabajamos y seguramente con el temor de que se sustituyan los trabajos que ahora harían las máquinas. Después de un período de transición inquietante siempre han evolucionado nuevos puestos de trabajo. ¿Es diferente esta vez?

**Martin Ford:** Creo que ahora estamos en el momento en que la tecnología está finalmente acá, cuando esta opción va a pasar y el hecho de que es, ya sabes, hemos tenido estas falsas alarmas tantas veces en el pasado que realmente es difícil para las personas aceptar esto, que es de donde viene una gran cantidad de la resistencia.

**Andrew Hill:** el libro de Martin Ford, «El ascenso los robots», ganador del premio del libro de negocios del año, de Financial Times y McKinsey, pinta un panorama sombrío del futuro en el que las máquinas desplazan a los seres humanos en todo menos en algunos roles muy especiales. Él prescribe soluciones económicas radicales, incluyendo un 'dividendo ciudadano universal' (ver en **Lampadia:** [El futuro del empleo con Robots e 'inteligencia Artificial'](#)), pero advierte que la disrupción se extenderá a profesiones que se consideraban inmunes.

**Martin Ford:** Uno ve ahora impactos en el periodismo (reporteros) y en revisión de documentos y empresas de abogados, creo que va a escalar, que comenzará a impactar a cualquiera que haga análisis o escritos rutinarios, ya sabes, se reporta a los superiores y haces lo mismo una y otra vez, todo eso va a ser susceptible, eventualmente.

**Andrew Hill:** Pero desarrollos como los que trabajan en la vanguardia de la tecnología cognitiva de IBM ven el futuro de manera diferente. Las máquinas más sofisticadas trabajarán en conjunto con los humanos abriendo nuevas posibilidades en áreas de trabajo todavía inimaginables, al igual que la última ola de cambio impulsada por computadoras creó la profesión del ingeniero de software.

**Guruduth Banavar (Vice Presidente, Cognitive Computing de IBM Research):** En el futuro, cuando nos metemos en este mundo de datos a gran escala, el conocimiento y los sistemas cognitivos, creo que va a haber un tipo similar de nuevas profesiones en el área, lo que yo llamaría ingeniería del conocimiento, que comienza a partir de la toma de datos en bruto alcance hasta la creación y representación del conocimiento de las diferentes áreas que necesita ser complementado por los tipos de aprendizajes estadísticos más automáticos para ser aplicados a los datos en bruto a gran escala.

**Andrew Hill:** Watson de IBM, el sistema cognitivo de más alto perfil, ya ha golpeado al ser humano en concursos, pero Banavar ve esta tecnología como una forma de ayudar a los humanos expertos, no suplantándolos.

**Guruduth Banavar:** Bueno, mi opinión es que hay tantas dimensiones de la inteligencia que, de hecho, hay mucho desconocimiento acerca de lo que significa inteligencia. Sostenemos ideas al respecto, pero no entendemos que sólo estamos arañando la superficie de la inteligencia, así que creo que un mejor enfoque para nosotros en este momento sería construir máquinas que aumenten la inteligencia humana para resolver problemas prácticos y que ayuden a hacer frente a la enorme cantidad de datos y conocimientos que se han generado en el mundo.

**Andrew Hill:** Mientras los trabajadores de oficina tratan de adaptarse a las nuevas tecnologías, Banavar cree las computadoras ayudarán en la transición de la nueva era, educando en formas nuevas y enseñando nuevas habilidades, pero la clave para saber si esta transformación socava o simplemente perturba las economías y sociedades está en cuán lejos y cuán rápido progresa el cambio.

**Martin Ford:** Creo que incluso las personas que no compran por completo esta idea de la amplia automatización estarían de acuerdo en que, con claridad, las cosas van a avanzar más rápido y más rápido en el futuro. Quiero decir, yo no creo que haya ninguna duda de que algunas ocupaciones van a evaporarse, la gente, si quiere seguir siendo relevante tienen que reentrenarse o encontrar algo más que hacer. Creo que lo más importante de cara al futuro es estar en una posición en la que estés preparado para hacer eso.

**Andrew Hill:** Finalmente, optimistas y pesimista de la tecnología, pueden no estar tan distantes como parece que la naturaleza revolucionaria de la tecnología cognitiva podría traer una gran promesa. Ambos están de acuerdo en la necesidad de que las personas, las empresas y las sociedades se adapten. Ambos piensan que la educación y la reeducación pueden suavizar la transición, pero las áreas en las que difieren fundamentalmente, es que la velocidad y la escala del cambio están llenos de incertidumbre. ¿Quién esté en lo cierto determinará si el ascenso de los robots conduce a un futuro más brillante para muchos trabajadores o a uno mucho más oscuro.

## Ya que hiciste preguntas complicadas sobre IA

Izabella Kaminska

18 de diciembre 2015

Financial Times

Traducido y glosado por Lampadia

¿Es la IA algo de lo que debemos preocuparnos?



Robots de un restaurante en Suzhou, China ©Getty

“Silicon Valley en movimiento para mantener segura la IA»

¿Es la IA algo de lo que debemos preocuparnos?

Bueno, ya sabes lo que dicen sobre IA...

No — ¿Qué?

Que los científicos dicen hace 40 años que la IA está a 40 años de distancia.

**Bueno. ¿Entonces por qué todo el ruido?**

Emprendedores de alto perfil y científicos han estado tocando las alarmas sobre el riesgo existencial que la IA trae a la humanidad. Elon Musk, el promotor de los autos eléctricos de Tesla tuiteó el año pasado que la IA era “potencialmente más peligrosa que las bombas atómicas”. Y el físico Stephen Hawking, ha advertido que “el desarrollo de inteligencia artificial puede significar el fin de la raza humana”.

¿Cuál es el riesgo inminente?

Los avances en sistemas de aprendizaje profundo – una forma de tecnología de algoritmos – ha permitido que las computadoras se involucren en tareas que antes solo podían ser hechas por los seres humanos. Estos descubrimientos se refieren a la habilidad de detectar pequeñas diferencias en el mundo, como reconocer patrones visuales de sonidos y, luego de ser capaces de responder a ellos con fluidez. En muchos casos esto ha creado una impresión convincente de acciones conscientes. Y como estos sistemas se ven y sienten inteligentes – aún cuando no hay inteligencia detrás de ellos – la gente está crecientemente inquieta.

¿O sea que estos desarrollos están mal representados?

Los científicos de computadoras te dirán que los intentos de aprendizaje de las máquinas usados por algo como el [Google's DeepMind](#) no son nuevos. Lo nuevo es la velocidad de proceso y el tamaño del conjunto de datos manejados por los algoritmos. Musk y Hawking están entre los que creen que un programa de ‘superinteligencia’ autónoma puede colarse, pues podría aprender de sí mismo exponencialmente.

¿Pero es correcto hacer ese supuesto?

Científicos como Simon Stringer del ‘Oxford Centre for Theoretical Neuroscience and Artificial Intelligence’ son escépticos sobre la idea de que se combinen las fuerzas de computación exponencial con estrategias de aprendizaje de máquinas que lleven a un tipo de IA que pueda ser una amenaza a los humanos. Estos modelos tienen muchas aplicaciones – la habilidad de tranzar en mercados o de navegar autos sin piloto – pero no pueden resolver los problemas de replicar la conciencia. Y todos ellos dependen de seres humanos para su intelecto. Un sistema verdaderamente consciente tendría que entender el mundo sensorial de la misma manera en que lo hace el cerebro humano. El trabajo de Stringer se focaliza en alcanzar eso, pero él no cree que lleve a la superinteligencia. Él espera lograr la inteligencia de una rata.

¿No sería inteligente cubrir el riesgo, por si acaso?

Musk y el financiero de Silicon Valley, Peter Thiel, piensan así, por lo cual ellos y otros han anunciado un proyecto de US\$ 1,000 millones para desarrollar IA en una forma que sea más probable de beneficiar al conjunto de la humanidad. Su proyecto OpenAI, será un movimiento sin fines de lucro que conduzca investigación de IA con código abierto y que comparta sus hallazgos con todos. Otros fundadores de OpenAI incluye a Amazon Web Services e Infosys, así como inversores individuales como Sam Altman, presidente de Y Combinator, una incubadora de start-ups de California.

¿De verdad los billonarios del club de élite de Silicon Valley alinean sus intereses con los de la humanidad?

Los cínicos sugerirían que este es un intento de evitar que las ventajas de la IA sean monopolizadas por una corporación o una persona. Un plan para compartir libremente la tecnología de IA no es lo mismo que frenar el conocimiento para que un sistema de inteligencia artificial sea creado y usado por un mal actor; o prevenir que una IA autónoma de aniquilar la humanidad. A lo máximo permite que la emergencia simultánea de muchos tipos de IA con los buenos protegiéndonos de los malos.

## ¿Pero qué pasaría si todos ellos se coluden contra la humanidad?

Exacto. Pero esa es una pregunta inconveniente para los billonarios de Silicon Valley.

## ¿Cómo así?

Si la IA posee realmente un riesgo existencial para la humanidad, lo mejor sería suspender todos los desarrollos. Imagínate lo que una moratoria del desarrollo de IA haría en la valorización de las compañías de tecnología.

**Se convertirían en nuevas compañías de combustibles fósiles. O sea que no sería una gran sorpresa si ellas forman un grupo de lobby para promover un desarrollo responsable.**



Ver video: [AI and the future for humanity](#)

### FT Negocios (Transcripción del video)

Ravi Mattu, Editor de Tecnología, FT

Durante casi 400 años, la Royal Society en Londres ha sido la sede de los líderes del pensamiento para discutir cómo la ciencia puede beneficiar a la humanidad. Desde Isaac Newton a Charles Darwin, los científicos se han reunido en estas habitaciones para considerar los temas más importantes del día. Ahora esta venerable institución está presentando una serie de eventos para la era de Internet. La primera reunión es una conferencia de un día sobre el aprendizaje de las máquinas, el proceso mediante el cual las computadoras son capaces de funcionar sin ser explícitamente programadas. Uno de los expertos que hablan en el evento en la Royal Society es Nick Bostrom, director del Instituto del Futuro de la Humanidad de la Universidad de Oxford y autor del best-seller «Súper-Inteligencia». Él dice que podría haber un momento en el futuro cercano, cuando las máquinas sean más inteligentes que los seres humanos y tenemos que hacer más para abordar el riesgo potencial de ello.

Habrá una gran intersección en la historia humana, cuando hagamos la transición de una era dominada por la inteligencia biológica a una era dominada por la inteligencia artificial. No es algo necesariamente inminente, de hecho, creo que tenemos muy poca evidencia acerca de la línea de tiempo para estos desarrollos, pero con el tiempo, tal vez a lo largo de décadas o más tiempo, vamos a encontrar la manera de producir en las máquinas el mismo proceso que hace inteligentes nuestros cerebros. Y va a ser realmente una invención como ninguna otra, el último invento, yo pienso, el último invento que los humanos alguna vez tengan que hacer porque después usted tendrá una máquina súper inteligente que lo hará mejor que nosotros. Incluso si esta transición es aún bastante lejana, las consecuencias son enormes y vale la pena empezar a pensar en ello desde hoy.

Si usted le cree a Hollywood, estas consecuencias pueden verse bastante aterradoras. Películas como Terminator y The Matrix sugieren que la inteligencia artificial podría ser la mayor amenaza para el futuro de la humanidad. Bostrom no cree que las películas de ciencia ficción sean una buena guía del futuro, no es el único que toca una campana de advertencia. El físico Stephen Hawking, que se basa en una forma de IA para comunicarse, ha dicho que un completo desarrollo de la inteligencia artificial podría significar el fin de la raza humana.

Súper-Inteligencia es algo potencialmente muy potente, al igual que la inteligencia humana es una cosa muy poderosa en relación con otras criaturas biológicas, que nos dieron a los humanos una posición inusual y dominante en el planeta. No es que tengamos músculos más fuertes o los dientes más afilados, es nuestro cerebro, que es ligeramente más inteligente que nos ha permitido los grandes avances desarrollando tecnologías, planes y organizaciones. Fundamentalmente las mismas razones, si usted tiene algo que supera nuestra inteligencia tanto como la nuestra supera la de los animales, esa cosa podría ser muy poderosa y puede ser capaz de dar forma al futuro de acuerdo a sus preferencias.

**Así es que usted dice que uno de los posibles riesgos es que el dispositivo, las tecnologías que creemos podrían llegar a dominar?**

Sí, creo que es un tipo de escenario posible, el futuro de los gorilas hoy depende mucho más de nosotros, de lo que los humanos decidan hacer, que de lo que los gorilas mismos decidan hacer. Del mismo modo, si hubiera máquinas súper-inteligentes, entonces nuestra suerte, el futuro de la humanidad, puede depender más de sus decisiones y de sus acciones. Entonces se convierte en algo muy importante que las preferencias de esta superinteligencia coincida con las nuestras, que estén lo suficientemente bien alineadas, que realmente se preocupen por el mismo tipo de cosas que nos importan, eso no es algo que no podemos dar por hecho, tenemos que actuar juntos para que en el momento en que necesitemos saber cómo controlar la súper inteligencia, podamos resolver el problema y parece que a partir de la limitada investigación que se ha hecho

hasta ahora, esto parece ser un problema realmente difícil .

**¿Es usted optimista sobre a dónde vamos ?**

**¿Eres neutral, o estás muy ansioso?**

Teniendo en cuenta nuestro estado actual de conocimiento podemos ver en la mesa, resultados extremadamente buenos y muy extremadamente malos. Yo creo que el resultados del tipo intermedio se ve probable cuando extrapolamos más hacia el futuro y que el destino de más largo plazo para la humanidad y pienso que es más probable que no sea ninguno de los extremos, ya sea la extinción o algo tan malo como la extinción, o un resultado extremadamente utópico.

**El avance tecnológico es imparable. Los riesgos indicados son probables y de muy alto impacto eventual. Todo esto puede estar lejos, pero eso no disminuye sus posibles efectos. Nos parece que ya es hora de que la humanidad inicie un proceso de fijación de reglas éticas que regulen el comportamiento de las máquinas súper-inteligentes. Este es otro aspecto en el que la humanidad necesita desarrollos positivos de gobernanza global. Ver en Lampadia: Inteligencia Artificial: [¿Se puede controlar?](#)**

**Para empezar, reflexionemos sobre las tres leyes de la robótica de Isaac Asimov:**

- 1. Un robot no puede causar daño a un ser humano o, por inacción, permitir que un ser humano sufra daño.**
- 2.- Un robot debe obedecer las órdenes dadas por los seres humanos, excepto si estas órdenes entrasen en conflicto con la primera ley.**
- 3.- Un robot debe proteger su propia existencia, en la medida en que esta protección no entre en conflicto con la primera o segunda ley.**

**Lampadia**