

# Tabla de Contenidos

## Episodio 17: Es posible que la IA sea sensible. No hay quien vea si será una buena idea

<b>#Porvivi</b> .....	1
<b><i>Man versus Machine</i></b> .....	2
<b><i>Antonio Damasio: homeostasis</i></b> .....	2
<b><i>Judea Pearl y Yoshua Bengio: Causalidad</i></b> .....	2
<b><i>Fernando Broncano: las emociones, territorios intermedios de la mente</i></b> .....	8
<b><i>¿Cómo será la convivencia entre el humano y la Inteligencia Artificial?</i></b> .....	10
<b><i>Otras referencias</i></b> .....	10



# Episodio 17: Es posible que la IA sea sensible. No hay quien vea si será una buena idea #Porvivid



La filosofía del porvivid se proyecta hacia el futuro. Bajo esa premisa [en el primer episodio de la serie, dedicado a la inteligencia artificial](#), y a partir de los conceptos fundamentales de la filosofía de la mente, Yashir P. Casas dejó una serie de preguntas abiertas que pueden iniciar el camino para el debate:

Con el conocimiento actual, y viendo hacia donde parece que nos lleva la tecnología, **¿Cuál será el concepto de lo humano en el futuro?** Llegados a un escenario de cierta estabilidad de la inteligencia artificial y situados en este posthumanismo, **¿Cómo será la convivencia entre el hombre y la inteligencia artificial?** Al hilo de esta convivencia, surgen dos cuestiones relevantes: Inteligencia artificial frente al trabajo, cuáles serán los **modelos laborales del futuro**, y qué **implicaciones éticas** tiene esta relación entre el hombre y la máquina inteligente. En una perspectiva aún más global, establecidas las posibles relaciones de la inteligencia artificial con el mercado y el consumo, **¿Cómo será la interacción entre la inteligencia artificial y el ecosistema y la sostenibilidad?**

Hoy comenzaremos indagando cómo podrían ser los seres artificiales inteligentes del futuro y cuál será nuestra convivencia con ellos. Lo haremos de la mano de los trabajos del neurocientífico Antonio Damasio y del informático y filósofo Judea Pearl.

“La filosofía no sirve para nada” es un podcast sin pretensiones en el que reflexionaremos sobre el presente.

Participan: Joaquín Herrero @joakinen, José Carlos García @quobit, Juan Carlos Barajas @sociologiadiver y Sergio Muñoz @smunozroncero

<b>Fecha</b>	18 de noviembre de 2019
<b>Participan</b>	Joaquín Herrero <a href="#">@joakinen</a> José Carlos García <a href="#">@quobit</a> Juan Carlos Barajas <a href="#">@sociologiadiver</a> Sergio Muñoz <a href="#">@smunozroncero</a>
<b>Descarga</b>	Puedes descargar todos los episodios en <a href="#">iVoox</a> , en <a href="#">Spotify</a> , en <a href="#">iTunes</a> , <a href="#">Google Podcasts</a> y en nuestro <a href="#">canal de Telegram</a> . Si tienes un lector de podcasts que admite enlaces RSS, <a href="#">este es el enlace RSS a nuestro podcast</a> .
<b>Sintonía</b>	<a href="#">Mass Invasion</a> , Dilo, álbum <b>Robots</b> (2004)
<b>Fotos</b>	<a href="#">PÍO,PÍO... &gt;&gt;&gt;</a> , Julyandme, Flickr
<b>Intro</b>	Conversación con Alexa

<b>Twitter</b>	En <a href="#">@FilosofiaNada</a> publicamos noticias que nos interesan y conversamos.
<b>Canal Telegram</b>	Puedes seguir la preparación de nuevos episodios suscribiéndote al canal <a href="#">@FilosofiaNada</a> en Telegram
<b>Grupo de opinión</b>	Únete a nuestro grupo de opinión <a href="#">Opina FilosofiaNada</a> para opinar sobre el episodio en preparación y enviarnos audios con preguntas o críticas con humor para nuestra intro



00:00 TIMELINE

## Man versus Machine

Quizá la idea de la relación que los seres humanos mantendremos con la inteligencia artificial esté bastante influenciada por una semejanza ideal. Parece que flota en la cultura popular que tendemos a una equivalencia entre la forma en que se desenvuelve la mente humana y la mente artificial, cuando el conocimiento que tenemos de la primera no nos permite poder reconstruirla en la segunda. Los robots inteligentes descritos en la ciencia ficción han contribuido sin duda a esta visión de una futura relación entre iguales, e incluso de sumisión del hombre a la máquina una vez que ésta haya alcanzado la -hipotética- singularidad.

## Antonio Damasio: homeostasis

No debemos desdeñar sin más estas líneas de la ficción especulativa, porque a nuestro juicio suponen una fuente muy fiable e interesante de anticipación a los problemas que podríamos llegar a afrontar. En este sentido, en el [capítulo 244 del podcast "Coffe Break: Señal y Ruido"](#) han comentado el reciente estudio publicado por el neurocientífico Antonio Damasio en Nature, en el que reivindica una exploración de las capacidades de supervivencia de la inteligencia artificial como medio para su evolución, a través de sus conocida tesis de la [homeostasis](#) aplicada a la IA.

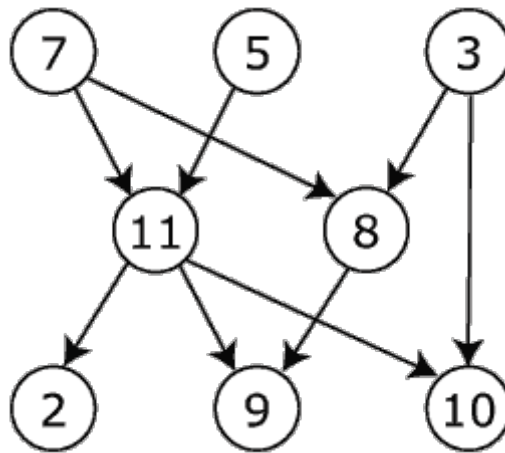
- en torno al minuto 50 empiezan a hablar del algoritmo de google
- en el 59.30 empiezan con la homeostasis en IA, después de la pausa del programa de radio

## Judea Pearl y Yoshua Bengio: Causalidad

Hicimos referencia al artículo de Damasio en [este hilo de Twitter](#) en el que decíamos: *"¡Interesante! Qué añadir a un robot para que surja la inteligencia creativa: deben gestionar su propia supervivencia, sentir dolor/placer/empatía, comprender la causalidad y ser afectados por contrafacticos. Un robot así nos permitiría estudiar filosóficamente la consciencia. El trabajo se centra en qué materiales permitirían a un robot ser vulnerable y sentir dolor pues se propone la necesidad de supervivencia como una clave de la creatividad. Pero nos han llamado la atención otras cosas filosóficamente relevantes. Por ejemplo, el hecho de que haya que implementar en el robot la*

comprensión de qué es causalidad y qué es un contrafáctico. El aprendizaje no consistiría solo en buscar correlaciones en muchos datos pues hay casos individuales tan relevantes que deben cambiar el comportamiento. La “inferencia casual”, propuesta por Judea Pearl y bien resumida en su libro “The Book of Why”, planteaba la necesidad de incluir razonamiento contrafáctico en el campo de la Inteligencia Artificial. Parece que su mensaje está calando, como él mismo dice.”

Citábamos un tuit del propio Judea Pearl en el que dice: “Dana Mackenzie y yo estamos contentos de que una de las personas que ayudó a desarrollar el aprendizaje profundo entienda esta limitación. Esperamos escribir una segunda edición de #Bookofwhy donde el capítulo diez será reemplazado por historias de éxito sobre cómo la IA superará su ceguera causal”. Pearl se está refiriendo a un tuit de un Paul Hünermund que dice “Yoshua Bengio está trabajando ya en inferencia causal y DAGs. El tema está ganando más y más impulso en la comunidad de ML e IA”. Al decir “DAGs” se refiere a los [Gráficos Acíclicos Dirigidos](#) que Judea Pearl usa al modelar la causalidad.



Hünermund, a su vez, está citando un artículo de la revista Wired sobre [Yoshua Bengio](#) titulado [An AI Pioneer Wants His Algorithms to Understand the 'Why'](#) y que anuncian con un tuit que dice “El aprendizaje profundo es bueno para encontrar patrones en una gran cantidad de datos, pero no puede explicar por qué conectados. El ganador del Premio Turing, Yoshua Bengio, quiere cambiar eso.”

Bengio recibió en 2019 el ACM A.M. Turing Award, “el premio Nobel de la computación”, junto con Geoffrey Hinton y Yann LeCun por los avances conceptuales y de ingeniería que han hecho de las redes neuronales profundas (Deep Learning) un componente crítico de la computación actual.

Entendemos la alegría de Judea Pearl porque hasta el presente las técnicas desarrolladas son capaces de encontrar correlaciones imposibles de encontrar para nosotros los humanos pero no pueden explicar ninguna de esas correlaciones. Lo explicábamos en el episodio 15 con el ejemplo de la botella y el tapón de rosca de Gary Marcus.

En el episodio del pasado 11 de diciembre de su porcast “Artificial Intelligence” Lex Fridman entrevista a Judea Pearl.



Judea Pearl: Causal Reasoning, Counterfactuals, Bayesian Networks, and the Path to AGI | AI Podcast



Judea Pearl: Causal Reasoning, Counterfactuals, Bayesian Networks, and the Path to AGI | AI Podcast



Judea Pearl: Causal Reasoning, Counterfactuals, Bayesian Networks, and the Path to AGI | AI Podcast



Judea Pearl: Causal Reasoning, Counterfactuals, Bayesian Networks, and the Path to AGI | AI Podcast



Judea Pearl: Causal Reasoning, Counterfactuals, Bayesian Networks, and the Path to AGI | AI Podcast

- “The do-operator connects the operation of doing with something that you can see. So as opposed to the purely observing, you're making the choice to change a variable.”
- En lugar de encontrar la causa de que en la naturaleza haya pasado en  $x$  causado por  $y$  y lo que hacemos es intervenir, hacer  $y$ . Entonces comprobamos qué sucede a  $x$ . También podemos operar en  $x$  y ver qué sucede con todos los  $y$  que supuestamente son causales de  $x$ .
- Para operaciones en las que no se puede intervenir en la realidad de  $x$  entonces intervenimos en el modelo que hemos creado de  $x$
- La idea de “modelizar” es la idea contrafáctica: qué sucede si lo cambio.
- Puede que el modelo no tenga suficiente información para averiguar relaciones causales. Entonces tenemos que hacer más intervenciones en el modelo (o en la realidad que se ha modelado) para agregar datos. Es decir, partimos de un modelo que un humano ha hecho y mediante experimentación podemos agregar más datos a porciones del modelo que sean cualitativas (¿incluso automáticamente?)
- El modelo comienza con una serie de suposiciones sobre qué causa qué, qué variables están causalmente relacionadas. Y entonces mediante intervenciones enriquecemos el modelo con datos.
- La presión inicial es sobre el humano que tiene que codificar su conocimiento sobre algo en un grafo. Pero hay que tener la precaución de que las flechas con las que se fabrica el grafo sean sobre conocimiento cierto o sobre conocimiento fruto de la experimentación. Si hay muchas flechas que marcan una hipótesis (como hacen los economistas) entonces nunca puedes averiguar causalidades porque el grafo se vuelve muy enrevesado y todo afecta a todo.

Para Judea Pearl la IA basada en redes neuronales es **aprendizaje por asociación** mientras que el

do-calculus en [redes bayesianas](#) codificadas como grafos DAG, es un **aprendizaje por intervención**. Para información más detallada ver [Gated Graphs and Causal Inference. Networks, processes and causality](#), de Microsoft Research.

¿Qué diferencia un contrafáctico de una intervención? Si me tomo una aspirina y se me pasa el dolor de cabeza el pensamiento contrafáctico es “¿seguiría el dolor de cabeza si no te hubieras tomado la aspirina permaneciendo el resto de factores igual?” Es un contrafáctico porque ambas cosas (tomar la aspirina y no tomar la aspirina) no pueden suceder a la vez y por eso este tipo de razonamiento es la fuente de nuestras explicaciones. Los físicos razonan así todo el tiempo. La forma matemática de este razonamiento es la **ecuación**, igualando dos factores de tal forma que si altero uno tengo que alterar el otro para mantener la igualdad que describe la ecuación. Una ecuación nos permite pensar contrafácticamente. Un robot no puede hacer esto. Un robot no puede hacer una declaración causal a menos que tenga un modelo causal del mundo y haga deducciones a partir de él.

¿Cómo construimos un modelo causal sin un equipo de humanos expertos que lo hagan? Parece que los niños construyen un modelo contrafáctico del mundo y lo hacen interviniendo sobre el mundo mediante sus juegos. Además de esto está la guía de los padres, las conversaciones con otros niños y los rumores que se propagan. Reciben información de muchas fuentes y la integran para completar su inicialmente precario modelo del mundo.

ESTO TIENE MUCHA RELACION CON LA FILOSOFIA DEL CUERPO Y CON LA PROPUESTA DE DAMASIO PORQUE NUESTRA **INTERVENCIÓN EN EL MUNDO USANDO NUESTRO CUERPO** Y RECIBIENDO FEEDBACK POR NUESTROS SENTIDOS COMPLETA NUESTRA RED BAYESIANA INTERNA CON LA QUE NOS RELACIONAMOS CAUSALMENTE CON EL MUNDO. Por tanto la inteligencia no es solo cuestión de tener un cerebro, es cuestión también de intervenir en el mundo con nuestro cuerpo.

Los humanos también somos capaces de COMBINAR DIFERENTES MODELOS CAUSALES SIMPLES procedentes de nuestras experiencias en diversas disciplinas y somos también capaces de usar metáforas sobre las experiencias. Pearl considera que LAS METAFORAS SON LA BASE DE LA INTELIGENCIA HUMANA ya que son un SISTEMA EXPERTO, son una forma de enfrentarse a un problema desconocido haciéndolo corresponder con un problema conocido, que te resulta familiar. ¿Qué es un problema conocido? Es uno en el que las respuestas a ciertas preguntas están explícitas, no hace falta derivarlas. Un buen jugador de ajedrez tiene ciertas reglas interiorizadas en un conocimiento explícito mientras que los que no sabemos jugar bien al ajedrez tenemos que pensar qué hacer en cada jugada, tenemos que derivar lo que para el experto es explícito. Esas respuestas explícitas han sido conocidas a través de repetir experimentos, intervenciones en el mundo (como cuando un niño juega con pelotas y aprende las leyes del movimiento y de los choques elásticos). Pearl indica que toda nuestra inteligencia está construida alrededor de metáforas que hacen corresponder lo desconocido con lo conocido. Esto es “razonamiento por metáforas”. Sin embargo aún no hemos sido capaces de construir algoritmos que metaforicen, esto es, que relacionen lo desconocido con un modelo del mundo que tenemos interiorizado.

*Nota: En el [episodio 3](#) estuvimos hablando de la metáfora y dijimos que el uso más fuerte de la metáfora es el que hace la filosofía orientada a los objetos, una filosofía natural contemporánea que pretende captar una realidad más completa que la que proporcionan los modelos científicos y asegura que solo podemos relacionarnos con la “realidad ahí fuera” mediante aproximaciones metafóricas. El libro de Ian Bogost *Alien phenomenology. What is like to be a thing* hace un repaso a esta filosofía que parte de la idea de Thomas Nagel (*What is like to be a bat?*) y llega a conclusiones diferentes. El capítulo 3 del libro está dedicado entero al metaforismo de esta propuesta filosófica y que se titula así precisamente: “Metaphorism”.*

Indica Pearl que **estamos haciendo grandes progresos en el dominio del razonamiento**



**causal.** Estos progresos son para él una **revolución causal**. En 30 años hemos hecho cosas que empujamos lo conseguido en el resto de historia humana. Estamos arrancando en el ejercicio de desarrollar modelos cualitativos del mundo, de cómo funciona el universo, que las máquinas usarán como punto de partida para derivar, aprender, todo lo que las matemáticas nos permitan aprender y llegar a respuestas cuantitativas a nuestras preguntas.

Respecto al **Big Data** muchas veces lo importante no es tanto tener muchos datos como tener datos de muchas fuentes. Esto último es mucho más importante porque nos permite extrapolar resultados médicos, por ejemplo, de intervenciones hechas en una población de deducir cómo afectarán a otra población diferente.

¿Qué es la **causalidad**? No solo la relación temporal entre eventos. Es también una relación estructural. La precedencia temporal no puede reemplazar la causalidad a pesar de que la causalidad nunca va “hacia atrás”. Pearl usa grafos DAG para representar la causalidad. El do-calculus es una relación (de intervención) entre dos eventos que es estructural, es independiente del tiempo aunque la relación es una flecha dirigida que tiene un sentido y que no puede ir hacia atrás (por eso los Grafos de Pearl son DAG, la D significa “dirigido”. Es acíclico, la “A”, porque no contiene caminos que empiecen y acaben en el mismo punto). Los grafos de Pearl son una **generalización de la lógica booleana**, según él mismo indica. En este grafo se representan ciertas relaciones causales y podemos preguntarle sobre otras relaciones causales que no están representadas en el grafo, relaciones que rompen las reglas y se preguntan qué sucedería si las cosas se hubieran desarrollado de otra forma (pensamiento contrafáctico). Lo que Pearl desea es una máquina que sea capaz de hacer eso, de aplicar un pensamiento contrafáctico a un modelo causal del mundo y preguntarse sobre cosas que no están contempladas en el modelo.

¿Qué sería el **aprendizaje** según Pearl? Una máquina podría ir almacenando todos los hechos sucedidos relativos a algo en todos los lugares del mundo pero no tendría la capacidad para establecer la dirección causal de los hechos que sea capaz de relacionar. No es capaz de responder a una pregunta contrafáctica “qué sucedería si el evento X no hubiera sucedido”. El machine learning tal como está concebido hoy solo ve datos, hechos y calcula correlaciones a ciegas, sin referirlas a causalidad. ¿Cómo se podría modificar el machine learning actual? Según Pearl solo mediante intervención en el mundo y comparar el dato antes de la intervención y el dato tras la intervención: *playful manipulation* (jugar a manipular el mundo). Podemos bajo ciertas circunstancias evitar la manipulación porque a veces de forma aleatoria sucede que los sucesos en el mundo cambian su curso de acción por accidente (los humanos enferman, tienen accidentes) y a veces ese “ruido” si es puramente aleatorio (y por tanto no sujeto a causalidad) nos puede servir para observar un suceso contrafáctico del tipo “¿y si X no hubiera sucedido?”.

El objetivo final, dice Pearl, es conseguir que una máquina sepa responder a cuestiones sofisticadas como los **contrafácticos**, la **compasión**, **lamentar algo**, la **responsabilidad** y el **libre albedrío**. No tenemos un equivalente al test de Turing para el libre albedrío pero sería posible hacerlo: si los robots fueran capaces de comunicarse como lo hacemos nosotros, es decir, con recompensas y castigos y dándole uno al otro un toque en la mano diciendo “no deberías de haber hecho eso” entonces el robot se cambiaría a sí mismo su software y haría las cosas de otra forma. Es parecido a como actuamos los humanos: no podemos cambiar al otro pero sí podemos motivarle y darle razones para que él mismo se cambie. Nosotros cambiamos nuestro patrón mental de comportamiento y los robots cambian su software, pero el efecto es el mismo: aprendizaje por contrafácticos. De esa forma un robot aprendería a hacer mejor algo que ya hace bien. Sin un aluvión de datos, solo con un buen argumento contrafáctico. Como hacemos nosotros.

**¿Cómo hacer que una máquina sea ética?** Necesita saber reconocer la causalidad y, a partir de ella, desarrollar empatía y deducir lo que es bueno para ti. Lo que llamamos compasión, es decir,

hacer corresponder mi modelo causal interior (el único que conozco) con el modelo causal de otro (que no conozco). "Ah, eres como yo, entonces no haré esto porque supongo que te dolerá, como me dolería a mí".

Nota: Aquí es importante recordar el **\*\*argumento de analogía de la voluntad\*\*** de Schopenhauer que, de forma similar, le permite hipotetizar sobre el funcionamiento de la naturaleza, la "voluntad" de la naturaleza, a partir de autoexaminar cómo funciona su propia voluntad, la única que conoce.

La **consciencia** sería ese modelo causal que cada uno tiene de sí mismo en contraposición con el entorno. Es un modelo causal, no es una lista detallada de reglas. La consciencia es un plano de tu propio software (*a blueprint of your own software*). Sobre ese modelo podemos pensar (reflexión) y decir "si cambio esto en mí sucederá esto otro" y entonces nos empezamos a desarrollar como personas. Eso es **libre albedrío**, la deducción de qué sucede si interferimos en el modelo causal y la capacidad de hacerlo.

**¿Qué hay de preocupante en el tema de la Inteligencia Artificial?** Pearl indica que estamos construyendo una nueva especie que nos va a superar en capacidades y no sabemos qué capacidad tendremos de controlarla. Ni siquiera sabemos qué es lo que tendremos que controlar. La única vez que ha sucedido algo similar ha sido con la evolución humana. Ha sido bueno para nosotros pero probablemente ha sido malo para otras especies. El problema es que tenemos un solo caso, un  $n=1$ , la evolución humana. Pero eso no es no saber nada porque también tenemos teorías y conjeturas sobre lo que podría suceder. Pero es difícil deducir nada a partir de ese conocimiento, solo preocupación.

**¿Deberíamos de equipar a un robot con creencias religiosas, con una metáfora de un dios?** Bueno, el robot tendrá un modelo de su entrenador, ya sea que dicho entrenador sea un humano u otro software. Todo lo que suceda en el mundo será referido a ese modelo que incluye a un entrenador último, un "dios", la relación más primigenia, más fundamental. No hay Dios, hay creencia en Dios. Es interesante que Pearl dice que la religión funciona siempre que tengas en cuenta que es solo una metáfora. Pero entonces (esto no lo resuelve Pearl) el robot tendría que reconocer en su modelo **qué es la realidad**, si es su modelo o aquello a lo que el modelo se refiere, esto es, el mundo. Y esta es una pregunta filosófica por excelencia.

## Fernando Broncano: las emociones, territorios intermedios de la mente

Una manera interesante de recorrer este camino puede ser recurrir a Fernando Broncano y sus notas acerca de las emociones como territorios intermedios de lo mental:

[Las emociones: territorios intermedios en la mente - Fernando Broncano](#)

En el artículo [El parque Westworld no ha contratado a ningún filósofo. Y debería.](#) Joaquín cita el trabajo de Broncano describiéndolo de esta forma:



El filósofo [Fernando Broncano](#) en su trabajo *Las emociones, territorios intermedios de la mente* plantea una tesis muy sugerente para abrir boca en el tema de qué son las emociones. Su tesis es que *las emociones conforman un sistema de señales o mensajes que porta y procesa contenido, aunque quizás este contenido es no-conceptual.*<sup>22)</sup>

Fijémonos en los dos aspectos de la tesis de Broncano: en la mente hay un transporte de mensajes/señales y hay contenido mental que no es conceptual.

Comenta Broncano, además, este hecho relevante acerca del sistema emotivo: su funcionamiento es parcialmente autónomo, esto es, no manejamos de forma consciente nuestras emociones. De hecho experimentamos la emoción de forma fulminante, mucho más rápidamente de lo que reflexionamos sobre la situación que nos está creando la emoción.

Pues bien, a partir de estos hechos ha elaborado una *teoría funcional del sistema emotivo* que define las emociones como un mecanismo de señales sobre planes de conducta. Es decir, según esta teoría el ser humano organiza su conducta de forma consciente en planes y el mecanismo de señales se dispara de forma inconsciente en función de lo que percibimos del entorno. Estas emociones añadidas a lo percibido permitirían que dichos planes persistan y avancen. Esta es una teoría de *información electro-química* si se quiere, pues se basa en señales que indican a la mente transiciones de planes: o bien lo detectado por nuestros sentidos facilita nuestros planes o los obstaculiza. Es decir, las emociones categorizan el mundo de forma simple pero rápida y eficaz. Nos permiten sentir afecto hacia lo que nos facilita la vida y rechazo hacia lo que la estorba. La base fisiológica de esa teoría se basa en el funcionamiento del sistema límbico, muy rápido y especializado en emociones pero conectado con el cortex, que procesa el pensamiento consciente.

Esta teoría del sistema emotivo es, como hemos dicho, una teoría funcional, es decir, describe cómo podría ser el modelo de funcionamiento de nuestras emociones pero no indaga en por qué funciona así. Para el moderno *hakim* es insuficiente decir que una emoción es cuando el sistema límbico interacciona muy velozmente con el sistema endocrino y lo hace sin que medien estructuras superiores del cerebro porque **lo que quiere es una explicación última** y una explicación descriptiva es tan solo una etapa intermedia en el camino hacia la comprensión total de algo. Por eso necesitamos algo más. De hecho **necesitamos un nuevo paradigma**, una nueva forma de pensar. Para conseguir esta nueva forma de pensar podemos inspirarnos en la forma de pensar de los filósofos que inauguraron la filosofía natural. Vamos a hacerlo.

En este sentido, también Zamora Bonilla en su libro “Sacando Consecuencias” hace una distinción bastante acertada del estado de la cuestión: las normas de inferencia por las que los humanos *sacamos consecuencias* “no son solo algoritmos computacionales, algo así como líneas de un programa informático (...) el ordenador saca consecuencias que está programado para sacar, no las saca porque le parezcan correctas. (...) El que nuestra deliberación sea un proceso al que cabe aplicar normas sobre lo que debemos o no debemos concluir, y no un ciego algoritmo que calcula las conclusiones a partir de las premisas, no implica de manera inevitable que nuestros procesos mentales no puedan ser reproducidos por un ordenador. (...) tal ordenador tendría que poseer al menos dos tipos de programas, uno que le permitiese realizar inferencias, y otro que le generase los estados que consisten en darse cuenta de si la inferencia realizada es correcta o incorrecta. (...) Para complicar las cosas, también tendría que tener un programa que le llevase a desear hacer ciertas cosas y a evitar otras”. (Sacando consecuencias, p. 37)

El artículo [Los científicos se cansan de la moda de la Inteligencia Artificial y piden prohibir algunos de sus usos](#) contiene este gráfico que muestra muy bien los logros y carencias de la forma como actualmente estamos usando las técnicas de Inteligencia Artificial.

### Grado de fiabilidad de la Inteligencia Artificial



FUENTE: PRINCETON UNIVERSITY

# ¿Cómo será la convivencia entre el humano y la Inteligencia Artificial?

Hablamos sobre qué es mejor, si una inteligencia artificial autónoma que aprenda de su propia interacción con el mundo y se autocorrija o los sistemas actuales que buscan correlaciones en medio de un Big Data de información. ¿Queremos crear alguien semejante al personaje [DATA de la serie Star Trek](#) aún a riesgo de que pueda convertirse en algo parecido a lo que era su hermano en dicha serie, el [personaje LORE](#)?

## Otras referencias

[Netflix: Documental sobre AlphaGo](#)  
[Entrevista a Jose Antonio Marina](#)

From:  
<http://filosofias.es/wiki/> - **filosofias.es**

Permanent link:  
<http://filosofias.es/wiki/doku.php/podcast/episodios/17>

Last update: **2022/06/23 09:40**

