

Tabla de Contenidos

Episodio 15: Inteligencia Artificial porvivir, que son dos días	1
<i>El Porvivir y su filosofía</i>	2
<i>Filosofía de la mente</i>	3
<i>Darmouth: diversidad de técnicas desde los inicios de la Inteligencia Artificial</i>	3
<i>Panorama contemporáneo de la industria de la Inteligencia Artificial</i>	6
<i>Antecedentes: 1981, premio Nobel de medicina, Hubel y Wiesel</i>	6
<i>¿Dualismo o monismo anómalo?</i>	6
<i>Antecedentes: 1943, modelo neuronal de McCulloch-Pitts.</i>	7
<i>Veranos e Inviernos de la Inteligencia Artificial</i>	7
<i>Peter Norvig y la confianza en la tecnología</i>	7
<i>Cómo ven los jóvenes la inteligencia en las máquinas</i>	8
<i>La ciudadanía digital ante una tecnología con sesgos e inexplicable</i>	9
<i>¿Correlaciones o causalidad? Judea Pearl y los contrafacticos</i>	9
<i>Críticas a las redes neuronales: Geoff Hinton, Gary Marcus y Marvin Minsky</i>	9
<i>La curiosa ocultación de la eficacia del teorema de Bayes</i>	11
<i>Curiosidades sobre Marvin Minsky</i>	12
<i>Un caso práctico de chatbot con Inteligencia Artificial híbrida</i>	12
<i>¿Estamos delegando nuestra responsabilidad en máquinas?</i>	12
<i>1999, Declaración de Santo Domingo sobre ciudadanía digital</i>	12
<i>El papel de la filosofía</i>	13

Episodio 15: Inteligencia Artificial porvivir, que son dos días



Inauguramos nuestra serie de la “Filosofía del Porvivir” con este episodio sobre Inteligencia Artificial. ¿Qué mundo estamos construyendo?

La breve pero intensa historia de este campo de investigación, 60 años, nos enseña que no está todo dicho respecto a cómo hacer que las máquinas tengan comportamiento inteligente. Acudiremos a Geoff Hinton, Gary Marcus y Marvin Minsky para obtener pistas de cómo podría ser la Inteligencia Artificial del futuro.

¿Es la Inteligencia Artificial un conocimiento acabado y del que tan solo cabe explotar su utilidad? ¿Hay modelos de Inteligencia Artificial en pugna? ¿Se está fabricando una tecnocracia en el tema de la Inteligencia Artificial que excluye a los ciudadanos de las decisiones sobre qué modelo usar? ¿Qué podemos hacer los ciudadanos para desarrollar una visión crítica sobre este campo tan especializado?

“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, Yashir P. Casas, Rodrigo Barajas y Sergio Muñoz @smunozroncero

Fecha	18 de noviembre de 2019
Participan	Joaquín Herrero @joakinen José Carlos García @quobit Juan Carlos Barajas @sociologiadiver Yashir P. Casas Rodrigo Barajas Sergio Muñoz @smunozroncero
Descarga	Puedes descargar todos los episodios en iVoox , en Spotify , en iTunes , Google Podcasts y en nuestro canal de Telegram . Si tienes un lector de podcasts que admite enlaces RSS, este es el enlace RSS a nuestro podcast .
Sintonía	Mass Invasion , Dilo, álbum Robots (2004)
Fotos	Future NEXT EXIT >>> , Daniel Voyager, Flickr
Intro	Conversación con el asistente de Google
Twitter	En @FilosofiaNada publicamos noticias que nos interesan y conversamos.
Canal Telegram	Puedes seguir la preparación de nuevos episodios suscribiéndote al canal @FilosofiaNada en Telegram
Grupo de opinión	Únete a nuestro grupo de opinión Opina FilosofiaNada para opinar sobre el episodio en preparación y enviarnos audios con preguntas o críticas con humor para nuestra intro



00:00 Introducción y presentaciones
07:20 Repaso a la comunidad "Opina"
10:20 Ada Lovelace vs. Margaret Hamilton
19:30 El Porvivir y su filosofía
26:00 Filosofía de la mente
34:38 Darmouth: dualidad de enfoques desde los inicios de la Inteligencia Artificial
44:20 Panorama contemporáneo de la industria de la Inteligencia Artificial
1:08:40 Antecedentes: 1981, premio Nobel de medicina, Hubel y Wiesel
1:11:09 ¿Dualismo o monismo anómalo?
1:13:00 Antecedentes: 1943, neurona de McCulloch-Pitts.
1:15:00 Veranos e Inviernos de la Inteligencia Artificial
1:16:60 Peter Norvig y la confianza en la tecnología
1:22:00 Cómo ven los jóvenes la inteligencia en las máquinas
1:26:30 La ciudadanía digital ante una tecnología con sesgos e inexplicable
1:35:00 ¿Correlaciones o causalidad? Judea Pearl y los contrafacticos
1:39:00 Críticas a las redes neuronales: Geoff Hinton, Gary Marcus y Marvin Minsky
1:44:15 La curiosa ocultación de la eficacia del teorema de Bayes
1:46:00 Curiosidades sobre Marvin Minsky
1:42:30 Un caso práctico de chatbot con Inteligencia Artificial híbrida
1:54:00 Gary Marcus: la falta de comprensión conceptual en las máquinas
1:59:00 ¿Estamos delegando nuestra responsabilidad en máquinas?
2:00:00 1999, Declaración de Santo Domingo sobre ciudadanía digital
2:02:30 El papel de la filosofía

El Porvivir y su filosofía

Da título a esta serie de episodios del podcast el verso del poeta Ángel González que dice **"Te llaman porvenir porque no vienes nunca"**.

La palabra "porvenir" vislumbra un próximo encuentro con el futuro pero nos deja a los humanos en el lado pasivo de dicho encuentro, esperando una venida. Una alternativa sería "porllegar", que es un neologismo interesante si no fuera porque sufre el mismo defecto: el futuro es algo que llega y yo solo tengo que esperar, no tengo control sobre lo que llega.

He decidido usar un neologismo diferente: **"porvivir"**, porque transmite la idea de que es un futuro que vamos a tener que vivir irremediamente pero del que cabe la duda de si es diseñable, si podemos hacer algo en el presente para que el futuro sea diferente. Esta **idea optimista de un futuro que se puede construir** es la que transmite el libro "Inventar el futuro" del filósofo Nick Srnicek aunque con la duda de si se puede intervenir en un **presente caótico que parece ingobernable**.

Así que "porvivir" deja abierta esta duda sobre nuestra posible intervención para diseñar el futuro y lo

único que asegura es que, sea cual sea, vamos a tener que vivir en el futuro. Vivimos en un mundo muy cambiante tecnológica, política y económicamente. No hay nada dicho de forma definitiva sobre cómo será el futuro. Por eso la “filosofía del porvenir” trata de esos aspectos del presente que no están cerrados ni terminados y sobre los que parece que se puede influir. También tratará de ser una filosofía de la sospecha que revele formas de imposición o manipulación que están siendo usadas, conscientemente o no, para condicionar el futuro.

La idea de la filosofía del porvenir es proyectar hacia el futuro el contenido de la especialidad filosófica CTS (Ciencia, Tecnología y Sociedad).

El [poema completo de Ángel González](#) parece contener todos los ingredientes de esta filosofía: la incertidumbre sobre el futuro.

Te llaman porvenir
porque no vienes nunca.
Te llaman: porvenir,
y esperan que tú llegues
como un animal manso
a comer en su mano.
Pero tú permaneces
más allá de las horas,
agazapado no se sabe dónde.
... Mañana!
Y mañana será otro día tranquilo
un día como hoy, jueves o martes,
cualquier cosa y no eso
que esperamos aún, todavía, siempre.

Filosofía de la mente

[Introduction to Philosophy. Philosophy of Mind](#), Rebus Open Textbooks

EN PREPARACIÓN

Darmouth: diversidad de técnicas desde los inicios de la Inteligencia Artificial

El campo de la Inteligencia Artificial nace oficialmente en 1956 en la llamada [Conferencia de Darmouth](#) si bien más que conferencia fue un taller (workshop) de dos meses de duración. Ya hacía años que se investigaba la posibilidad de inteligencia no humana y había ciertos avances en el conocimiento del funcionamiento de las neuronas y del comportamiento inteligente, dos investigaciones estas, la fisiológica y la psicológica, diferentes aunque complementarias y que darán lugar a técnicas bien diferentes dentro del paradigma de la Inteligencia Artificial: las investigaciones simbólicas y sub-simbólicas.

La gran narradora de cómo se inicia el campo de la Inteligencia Artificial es [Pamela McCorduck](#), que ha sido entrevistada por Lex Fridman en el episodio [Pamela McCorduck: Machines Who Think and the](#)

Early Days of AI. Ella comenta cómo el campo de la IA comienza apostando por la **Inteligencia Artificial simbólica** y durante 20 años este sería el paradigma dominante en la investigación y el que atrajo más fondos para fundar centros de investigación.

En estas imágenes tomadas de la documentación del curso **Symbolic vs. Subsymbolic AI** del MIT Media Lab se resumen las técnicas que agrupamos dentro de estas dos grandes categorías: IA simbólica e IA no-simbólica.

Symbolic vs. Subsymbolic	
Explicit symbolic programming	Bayesian learning
Inference, search algorithms	Deep learning
AI programming languages	Connectionism
Rules, Ontologies, Plans, Goals...	Neural Nets / Backprop
	LDA, SVM, HMM, PMF, alphabet soup...
Henry Lieberman • MIT	
Symbolic vs. Subsymbolic	
Introspection more useful for coding	More robust against noise
Easier to debug	Better performance
Easier to explain	Less knowledge upfront
Easier to control	Easier to scale up
Not so Big Data	Big Data
More useful for explaining people's thought	More useful for connecting to neuroscience
Better for abstract problems	Better for perceptual problems
Henry Lieberman • MIT	

El primer éxito real de las técnicas de Inteligencia Artificial Simbólica es la denominada como **Sistema experto**, creada por **Ed Feigenbaum**. Sobre esto la Wikipedia indica en su artículo "Sistema Experto":

Los sistemas expertos fueron introducidos formalmente alrededor de 1965 por el Proyecto de Programación Heurística de Stanford dirigido por Edward Feigenbaum, a quien a veces se le llama el "padre de los sistemas expertos". Otros investigadores clave fueron Bruce Buchanan y Randall Davis. Los investigadores de Stanford intentaron identificar dominios donde la experiencia era altamente valorada y compleja, como diagnosticar enfermedades infecciosas (Mycin) e identificar moléculas orgánicas desconocidas (Dendral). La idea de que "los sistemas inteligentes derivan su poder del conocimiento que poseen en lugar de los formalismos específicos y los esquemas de inferencia que

usan”, como dijo Feigenbaum, fue en ese momento un importante paso adelante, ya que la investigación anterior se había enfocado sobre métodos computacionales heurísticos, que culminan en intentos de desarrollar solucionadores de problemas de propósito muy general (principalmente el trabajo conjunto de Allen Newell y Herbert Simon). Los sistemas expertos se convirtieron en algunas de las primeras formas verdaderamente exitosas de software de inteligencia artificial (IA).

En los años 1970, de la mano de John Henry Holland, surgió una de las líneas más prometedoras de la inteligencia artificial, la de los [algoritmos genéticos](#), son llamados así porque se inspiran en la evolución biológica y su base genético-molecular. Estos algoritmos hacen evolucionar una población de individuos sometiéndola a acciones aleatorias semejantes a las que actúan en la evolución biológica (mutaciones y recombinaciones genéticas), así como también a una selección de acuerdo con algún criterio, en función del cual se decide cuáles son los individuos más adaptados, que sobreviven, y cuáles los menos aptos, que son descartados. Ver [Inteligencia artificial: algoritmos genéticos](#), en el que Daniel Riera, Doctor Ingeniero en Informática hace un recuento de los usos de los algoritmos genéticos en Inteligencia Artificial explicando también en qué consisten:

La implementación de la idea de John H. Holland (en los años 70) no es excesivamente compleja: Partimos de una población de individuos, en que cada uno representa una solución factible del problema que se quiere resolver. Cada solución tiene una calidad asociada (fitness) y evidentemente, debe ser factible. En este punto, aplicamos un conjunto de mecanismos evolutivos que producen una nueva generación de individuos. Si todo va bien, tras un cierto número de generaciones, iremos obteniendo individuos (soluciones) que mejoran la solución inicial. Pero ¿cuáles son esos mecanismos? Pues básicamente tres: cruce (simulamos reproducción sexual, por lo que escogeremos parejas de soluciones, las partimos y cruzamos para generar su descendencia), mutación (incluimos mutaciones al azar en los individuos para evitar la convergencia de la población hacia un máximo/mínimo local) y «selección» (eliminamos individuos que representen soluciones no factibles - inviables- y además, en cada generación, nos quedamos con un subconjunto de padres e hijos, normalmente siguiendo el criterio de mejor fitness). Si como en la realidad este proceso se itera millones de veces (por suerte, los ordenadores nos permiten que los algoritmos genéticos simulen a una velocidad más cercana a la evolución de las bacterias que a la de los primates) nos iremos acercando a una población de soluciones de calidad normalmente muy aceptable.

No debemos olvidar que los Algoritmos Genéticos son metaheurísticas y por lo tanto no hacen una búsqueda completa (exhaustiva) sobre el espacio de soluciones, por lo que no aseguran encontrar la solución óptima. Pero dado que el tipo de problemas a los que se aplican suelen ser de complejidad NP, ya es lo esperado. Estos algoritmos suelen tener pocos parámetros (tamaño de las poblaciones, número de generaciones máximo, probabilidad de mutación, ...) pero son extremadamente dependientes del modelado de los individuos (soluciones), ya que es clave facilitar que la producción automática (mediante cruce) de nuevos individuos genere mayoritariamente individuos viables. También tiene bastante impacto la elección de la población inicial (debido a lo que en BM se llamaría deriva genética, principalmente).

Desde los años 70, los Algoritmos Genéticos han sido una de las «ideas» de la IA más investigadas y utilizadas, dando solución a multitud de problemas combinatorios y siendo utilizadas en aplicaciones reales desde el diseño de los horarios de los JJOO hasta la planificación de rutas logísticas. Y por supuesto, no se han escapado de «sufrir» en sus propios huesos la misma evolución que implementan dando lugar a cientos de aproximaciones y mejoras.

Panorama contemporáneo de la industria de la Inteligencia Artificial

EN PREPARACIÓN

Antecedentes: 1981, premio Nobel de medicina, Hubel y Wiesel

El punto de partida del despegue de las técnicas basadas en redes neuronales es la comprensión de cómo el cerebro procesa la información visual, debida a [David Hubel](#) y [Torsten Wiesel](#), premiados con el [Premio Nobel en Medicina en 1981](#) “por sus descubrimientos sobre el procesamiento de información en el sistema visual”, que fueron ver en el cerebro los siguientes procesos:

- El sistema de procesamiento de la visión consta de varias capas de neuronas
- Las capas más cercanas al ojo detectan límites entre contrastes: “edge detectors”
- Cada neurona se dedica a reconocer patrones de límites en una zona limitada del campo de visión
- Las siguientes capas reconocen patrones más complejos
- Hay flujo de datos entre capas en ambas direcciones: forward-propagation y back-propagation

¿Dualismo o monismo anómalo?

“Monismo, porque mantiene que los acontecimientos psicológicos son acontecimientos físicos; anómalo, porque insiste en que los acontecimientos no caen bajo leyes estrictas cuando se describen en términos psicológicos” -Donald Davidson en *Psychology as Philosophy, Essays on Actions and Events*, citado en *Filosofía de la mente y de la Ciencia cognitiva*, de José Hierro-Pescador.

Donald Davidson llega al monismo anómalo a través de tres supuestos aparentemente incompatibles:

- Los sucesos mentales interactúan causalmente con sucesos físicos, unos pueden ser causa de los otros.
- Los sucesos que causan otros sucesos caen bajo una estricta –esto es, sin excepciones– ley natural.
- No hay ninguna ley natural estricta acerca de los sucesos mentales.

Estos tres supuestos parecen contradecirse, ya que el 2º y el 3º parecen implicar la negación del 1º, el 1º y el 2º la negación del 3º, y el 1º y el 3º la negación del 2º. Davidson, no obstante, tiene a los tres por verdaderos, por lo que intenta elaborar una posición que satisfaga los tres supuestos. Esta es, a su juicio, el monismo anómalo.

El monismo anómalo interpreta los supuestos 1 y 3 de la siguiente manera:

1* Los sucesos mentales individuales (casos) interactúan causalmente con sucesos físicos, unos pueden ser causa de los otros y viceversa. 3* No hay leyes naturales estrictas acerca de los tipos de sucesos mentales.

El supuesto 1* es válido, ya que todo suceso mental individual es idéntico a un suceso físico, y en cuanto suceso físico cae también, en el sentido del supuesto 2, bajo una estricta ley natural. Esta suposición conforma la parte monista de la filosofía de Davidson. No obstante, aun cuando todo suceso mental, en cuanto suceso físico, cae bajo una estricta ley natural, no existen leyes psicológicas o psicofísicas generales. En virtud de esta tesis, Davidson da a su filosofía la denominación de «anómala» (de «nomos», ley).

El argumento central de Davidson en favor del monismo anómalo tiene por tanto la forma de una conclusión obtenida a partir de la mejor explicación posible. Los supuestos 1 y 3 son ciertos. Sólo si el monismo anómalo es cierto puede explicarse que los supuestos 1 y 3 sean ciertos. Luego el monismo anómalo es cierto.

Fuente: [Monismo anómalo](#), Wikipedia

Otras mentes:

Podemos aprender mucho del funcionamiento de la mente humana investigando el funcionamiento de mentes radicalmente diferentes a la nuestra. Es la idea fundamental del filósofo de la ciencia e historiador [Peter Godfrey-Smith](#), quien en su libro [Otras mentes. El pulpo, el mar y los orígenes profundos de la consciencia](#) hace un recorrido por la evolución de la mente distribuida de los cefalópodos y especula acerca del origen común de mentes tan distintas de la de los homínidos.

[Reseña del libro en la Revista de la Universidad de México](#)

Antecedentes: 1943, modelo neuronal de McCulloch-Pitts.

[Neurona de McCulloch-Pitts](#), Wikipedia

Veranos e Inviernos de la Inteligencia Artificial

Ver el artículo de la Wikipedia en inglés [History of artificial intelligence](#)

Peter Norvig y la confianza en la tecnología

La Unión Europea está regulando el [Derecho a la Explicación](#) respecto a decisiones automatizadas.



A esta problemática se refiere el científico de la computación Peter Norvig en [esta entrevista que le hace Lex Fridman](#) en su podcast Artificial Intelligence (minutos 18 a 25), aunque Norvig desplaza el problema desde el tema de la "explicabilidad" al tema de la confianza en la exactitud de los resultados proporcionados por sistemas de 'caja negra'.

secondly we could somehow learn yeah there's this rule that you can remove one grain of sand and you can do that a bunch of times but you can't do it the new your infinite amount of times but on the other hand when you're doing induction on the integers sure then it's fine to do it an infinite number of times. And if we could somehow we have to learn when these strategies are applicable rather than having the strategies be completely neutral and of it available everywhere anytime using you know networks anytime you learn from data form representation from day in an automated way it's not very explainable as to or it's not introspective to us humans in terms of how this neural network sees the world where why does it succeed so brilliantly on so many in so many cases and fail so miserably in surprising ways and small. So what do you think is this is the future there can simply more data better data more organized data solve that problem or is there elements of symbolic systems they need to be brought in which are a little bit more explainable. So I prefer to talk about trust and validation and verification rather than just about explain ability and then I think explanations are one tool that you use towards those goals and I think it is an important issue that we don't want to use these systems. Unless we trust them and we want to understand where they work and where they don't work and in an explanation can be part of that right. So I apply for loan and I get denied I want some explanation of why. You have in Europe we have the GDPR that says you're required to be able to get that but on the other hand an explanation alone is not enough. So you know we were used to dealing with people and with organizations and corporations and so on and they can give you an explanation and you have no guarantee that that explanation relates to reality right right so the bank can tell me well you didn't get the loan because you didn't have enough collateral and that may be true or it may be true that they just didn't like my religion or or something else. I can't tell from the explanation and that's that's true whether the decision was made by computer or by a person. So I want more I do want to have the explanations and I want to be able to have a conversation to go back and forth and said well you gave this explanation but what about this and what would have happened if this had happened and what would I need to change that. So I think a conversation is a better way to think about it than just an explanation as a single output and I think we need testing of various kinds right so in order to know was the decision really based on my collateral or was it based on my religion or skin color or whatever I can't tell if I'm only looking at my case. But if I look across all the cases then I can detect the pattern that's right so you want to have that kind of capability you want to have these adversarial testing all right.

Cómo ven los jóvenes la inteligencia en las máquinas

EN PREPARACIÓN

La ciudadanía digital ante una tecnología con sesgos e inexplicable

Parece necesario que los ciudadanos comprendamos que determinadas decisiones que nos afectan en nuestra vida diaria estarán, si no lo están ya, ejecutadas por algoritmos. Desafortunadamente el contenido de la “caja negra” en la que se toman estas decisiones no es accesible para cualquiera, pero la legislación europea ha ido dando pasos en la protección del ciudadano frente a estas decisiones automatizadas basadas en grandes conjuntos de datos que incluyen los suyos y su identidad, de forma que cualquiera puede, al menos potencialmente, decidir no ser objeto de estas decisiones.

El artículo 21.1 del **Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo** de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (de aplicación obligatoria en todos los estados de la UE desde el 25 de mayo de 2018), dice textualmente:

“Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar.”

[Texto completo en pdf del BOE](#)

¿Correlaciones o causalidad? Judea Pearl y los contrafacticos

- [Judea Pearl](#), Wikipedia
- [Keynote: Judea Pearl - The New Science of Cause and Effect](#), PyData 2018, Los Angeles (YouTube) 😊

Joaquín Herrero (2018) explica el modelo de Pearl en su TFG, páginas 21-35: [Causalidad en redes neuronales y Big Data](#), en *La causalidad en entornos de complejidad*.

Críticas a las redes neuronales: Geoff Hinton, Gary Marcus y Marvin Minsky



En esta entrevista de 2004 disponible en YouTube titulada [Prof. Marvin Minsky Forecasts The Future of Artificial Intelligence](#) Marvin Minsky, uno de los “cuatro padres” de la Inteligencia Artificial junto con John McCarthy, Allen Newell y Herbert Simon y que desde el comienzo apostó por un enfoque simbólico de la Inteligencia Artificial ha ido evolucionando su punto de vista y últimamente defendía la idea de que la virtud de cualquier sistema de inteligencia artificial sería poder usar para cada problema la técnica más apropiada en vez de intentar usar una sola técnica para todos los problemas: “el progreso fue muy rápido hasta alrededor de 1980 y luego disminuyó la velocidad”.

A este respecto comenta:

Es difícil saber las razones pero, por lo que puedo ver ahora, lo que sucedió es que un máquina muy inteligente debería de tener muchas diferentes formas de hacer las cosas. Pero la mayoría los investigadores eligen un camino e intentan demostrar que por ese camino se podrían resolver la mayoría de los problemas. Así que ahora estamos en una nueva fase donde estamos tratando de ver cómo puedes hacer una máquina que sepa hacer las cosas de muchas maneras diferentes y que también tiene el conocimiento para saber cuál forma utilizar para cada tipo de problema”.

Por tanto se puede decir que Minsky defiende un enfoque híbrido de la Inteligencia Artificial.



Esta es la misma postura que defiende Gary Marcus, como se puede ver en la entrevista que le hizo Lex Fridman en octubre de 2019: [Gary Marcus: Toward a Hybrid of Deep Learning and Symbolic AI](#). También dijo lo mismo en el artículo [Deep Learning: A Critical Appraisal](#) en cuyo resumen dice: “Aunque el aprendizaje profundo tiene raíces históricas que se remontan a décadas, ni el término “aprendizaje profundo” ni el enfoque eran populares hace poco más de cinco años, cuando el campo fue reactivado por documentos como Krizhevsky, Sutskever y el

modelo de red profunda ahora clásico de Hinton (2012) de Imagenet. ¿Qué ha descubierto el campo en los cinco años siguientes? En un contexto de progreso considerable en áreas como el reconocimiento de voz, el reconocimiento de imágenes y el juego, y un entusiasmo considerable en la prensa popular, presento diez preocupaciones por el aprendizaje profundo, y sugiero que el aprendizaje profundo debe complementarse con otras técnicas si queremos llegar a la inteligencia artificial general.”.

Criticando las soluciones estadísticas a problemas de la vida real pone Marcus el ejemplo de un tapón de rosca y una botella: una red neuronal captará fácilmente que existe una correlación entre tapones y botellas pero no comprenden el ingenioso mecanismo de rosca que hemos ideado para que la botella quede sellada con el tapón y el líquido no salga. A comprender estos mecanismos lo llamamos “sentido común” y la comprensión de la correlación entre tapones y botellas es, para nosotros, una consecuencia de comprender el mecanismo de rosca. La percepción de las máquinas es muy buena

pero se basa en la correlación entre pixels de imágenes, no en la aplicación de conceptos a las imágenes ya que la máquina no tiene una representación de los conceptos sino una representación de las correlaciones. El sentido común es MUY complicado de desarrollar en máquinas.

Indica Marcus que parte de nuestro sentido común parece responder a reglas más o menos fijas y parte de nuestro sentido común parece responder a intuiciones de las que no sabríamos dar explicación. A partir de una descripción conceptual del mundo como la que hacen las técnicas GOFAL (sistemas expertos, algoritmos, programas) podemos hacer inferencias sobre el mundo. A partir del aprendizaje de correlaciones propio de las redes neuronales también. ¿Por qué no mezclar ambas técnicas?

Ver <https://skymind.ai/wiki/symbolic-reasoning> para casos en los que la IA simbólica se ha mezclado con redes neuronales para solucionar la rigidez de los enfoques simbólicos, ya que las reglas están codificadas en programas.

El último artículo hace referencia a este otro

<https://www.axios.com/artificial-intelligence-pioneer-says-we-need-to-start-over-1513305524-f619efbd-9db0-4947-a9b2-7a4c310a28fe.html> en el que Geoff Hinton (informático británico galardonado con el Premio Turing en 2018 junto con Yoshua Bengio y Yann LeCun por su trabajo en deep learning) da por muerta la técnica de backpropagation y apuesta por empezar otra vez a pensar las técnicas de inteligencia artificial:

En 1986, Geoffrey Hinton fue coautor de un artículo que, tres décadas después, es fundamental para la explosión de la inteligencia artificial. Pero Hinton dice que se debe prescindir de su método innovador y encontrar un nuevo camino hacia la IA.

Hablando con Axios al margen de una conferencia de inteligencia artificial en Toronto el miércoles, Hinton, profesor emérito de la Universidad de Toronto e investigador de Google, dijo que ahora es "profundamente sospechoso" de la propagación hacia atrás, el método de trabajo que subyace a la mayoría de Los avances que estamos viendo en el campo de la inteligencia artificial hoy en día, incluida la capacidad de ordenar fotos y hablar con Siri. "Mi punto de vista es tirar todo y comenzar de nuevo", dijo.

El resultado final: Otros científicos en la conferencia dijeron que la propagación hacia atrás todavía tiene un papel central en el futuro de la IA. Pero Hinton dijo que, para avanzar materialmente, probablemente habrá que inventar métodos completamente nuevos. "Max Planck dijo: 'La ciencia progresa un funeral a la vez'. El futuro depende de algún estudiante graduado que sospeche profundamente de todo lo que he dicho ".

La curiosa ocultación de la eficacia del teorema de Bayes

La aproximación probabilista: Teorema de Bayes. ¿Qué hace la regla o teorema de Bayes? Sirve para evaluar el grado de fiabilidad de una hipótesis en función de la nueva evidencia que vamos descubriendo.

- [Teorema de Bayes](#), Wikipedia
- [Understanding Bayes' Theorem](#), YouTube

Un ejemplo típico de uso del teorema de bayes es para interpretar el resultado de pruebas clínicas, tal como indica la entrada [Doctors don't know Bayes' Theorem](#) del blog de la Cornell University.

Sharon Bertsch McGrayne en su charla [Sharon Bertsch McGrayne: "The Theory That Would Not Die"](#), disponible en YouTube, nos cuenta la historia de esta forma de comprobación de hipótesis muy usada en la Inteligencia Artificial y la curiosa historia de cómo se ha ocultado su eficacia.

Curiosidades sobre Marvin Minsky

El 3 de abril de 1968 se estrenó '2001: Una odisea del espacio', la película visionaria de Stanley Kubrick que imaginó una misión espacial en la que una máquina inteligente llamada HAL se rebelaba contra sus tripulantes. Con motivo de este aniversario, [recuperamos hoy el vídeo de un fascinante diálogo](#) que organizó la Fundación BBVA entre Marvin Minsky, uno de los padres de la inteligencia artificial que asesoró el guión del filme, y el cineasta español [Kike Maíllo](#). El encuentro, disponible en el vídeo que encabeza esta noticia, tuvo lugar cuando Minsky visitó Madrid en junio de 2014 para recibir el Premio Fronteras del Conocimiento en Tecnologías de la Información y la Comunicación.

Un caso práctico de chatbot con Inteligencia Artificial híbrida

El proyecto [Dej@lo](#), un ensayo clínico para evaluar la eficacia de chatbots como apoyo al abandono del tabaquismo, incluye diversos enfoques de Inteligencia Artificial:

- como sistema experto es un un grafo de estados que implementa un guión de práctica médica de probada eficacia
- procesa el lenguaje natural mediante mecanismos Bayes
- para averiguar el siguiente estado del grafo y dentro de ese estado ejecutar un guión de un sistema experto se puede consultar una red neuronal en lugar de seguir el rígido guión experto

En este vídeo se puede ver el funcionamiento del chabot en lo que respecta a la implementación del sistema experto: [Demostración del uso del bot Dej@lo](#)

¿Estamos delegando nuestra responsabilidad en máquinas?

Ver la entrevista de 5 minutos a Cathy O'Neil [Tech & Society: Entrevista a Cathy O'Neil | #TechSociety](#)

1999, Declaración de Santo Domingo sobre ciudadanía digital

"Declaración de Santo Domingo" hecha en 1999 por la UNESCO. El informe se titula [La ciencia para el siglo XXI: una nueva visión y un marco de acción](#) y dice en el punto 2.2:

Por su parte, las comunidades de investigadores deben:

- (i) *contribuir, especialmente en el caso de problemas en los que están involucradas, a la*

presentación de alternativas sobre las cuales la ciudadanía pueda informarse y pronunciarse

- *(ii) tener en cuenta las opiniones de la sociedad y dialogar efectivamente con ella*
- *(iii) luchar contra el entronizamiento de tecnocracias amparadas en conocimientos científicos y tecnológicos, reales o supuestos.*

Es decir, la UNESCO apuesta por una divulgación crítica, es decir, de corte filosófico. No divulgar para que la gente entienda lo que se ha hecho, sino divulgar para que la gente entienda lo que se podría hacer y que tenga voz respecto a decisiones políticas y tecnológicas que podrían condicionar el futuro.

El papel de la filosofía

Este uso de la filosofía está en consonancia con lo que comenta Scott Soames, profesor de filosofía en la University of Southern California, cuando, en la introducción de su último libro, [The World Philosophy Made: From Plato to the Digital Age](#) dice:

"Los filósofos están profundamente involucrados en todas las áreas importantes de interés intelectual, incluidas las artes, las ciencias y las humanidades. Bien entendida, la filosofía no es una disciplina aislada, sino el socio de prácticamente todas las disciplinas."

"Aunque la historia es una parte importante del tema, los filósofos de hoy generan nuevas preguntas filosóficas, al tiempo que ofrecen mejores respuestas a las preguntas tradicionales que las formuladas por pensadores anteriores. El conocimiento filosófico está aumentando. Los filósofos ayudan proporcionando nuevos conceptos, reinterpretando viejas verdades y reconceptualizando preguntas para expandir sus espacios de solución."

"A medida que la ciencia avanza, hay más, no menos, que hacer por parte de la filosofía. Nuestro conocimiento del universo y de nosotros mismos crece como una esfera de luz en expansión desde un punto de iluminación. Según viaja la luz en todas las direcciones, el volumen de la esfera, nuestro conocimiento seguro, crece exponencialmente. Pero también lo hace el área de la superficie de la esfera, esto es, el límite donde el conocimiento se desdibuja y vuelve la incertidumbre metodológica. La filosofía controla la frontera, lista para ayudar a trazar nuestro próximo movimiento."

From:

<https://filosofias.es/wiki/> - **filosofias.es**

Permanent link:

<https://filosofias.es/wiki/doku.php/podcast/episodios/15>

Last update: **2022/06/15 10:40**

