

<https://elpais.com/tecnologia/2022-02-16/judea-pearl-los-ordenadores-pronto-seran-capaces-de-explicar-el-mundo-y-a-ellos-mismos.html>

Judea Pearl: “Los ordenadores pronto serán capaces de explicar el mundo y a ellos mismos”

Manuel G. Pascual

El estadounidense, premio Fundación BBVA Fronteras del Conocimiento, cree que en menos de 25 años la inteligencia artificial “experimentará una revolución” al entender el entorno



El informático estadounidense

Judea Pearl, de 85 años, fue galardonado la semana pasada con el premio Fundación BBVA Fronteras del Conocimiento. **FUNDACIÓN BBVA**

Nadie sabe qué razonamiento [siguió en 2016 AlphaGo](#), un programa [desarrollado por DeepMind](#) (Google), para derrotar sin paliativos al campeón del mundo de Go, un juego de origen chino considerado mucho más complejo que el ajedrez. Ese sistema usó redes neuronales, que se apoyan en el *deep learning* o aprendizaje automático. Se le enseñaron las reglas del juego, en el que el factor intuitivo es clave, y luego echó varias partidas contra sí mismo. Dicen los entendidos que AlphaGo desarrolló unas tácticas nunca antes vistas. Su victoria se interpretó como la prueba definitiva de que las máquinas ya no tienen rival. Aunque no seamos capaces de entender cómo llegó el programa a ese manejo.

Pero este modelo de desarrollo de la inteligencia artificial (IA), en el que la opacidad de los algoritmos es casi total, no es el único en la disciplina. Frente a la propuesta del aprendizaje automático hay otra también muy potente, basada en sistemas de causalidad, que pretende arrojar luz sobre el proceso de toma de decisiones de la máquina. El padre de esta corriente es el estadounidense Judea Pearl, que la semana pasada [fue distinguido con el premio Fundación BBVA Fronteras del Conocimiento](#) en la categoría de Tecnologías de la Información y la Comunicación. Un galardón más para sus pobladas vitrinas, en las que luce desde 2011 el premio Turing (2011), considerado el Nobel de la informática.

Pearl nació en Tel Aviv en 1936, antes de la formación del Estado de Israel (entonces Mandato británico de Palestina). En 1969 se incorporó a la Universidad de California en Los Ángeles (UCLA), donde todavía ejerce como profesor a sus 85 años. Su gran contribución a la disciplina es la aplicación de las redes

bayesianas para reducir la incertidumbre (lograr que la máquina no se vea abrumada ante la multiplicidad de variables que le impactan). De la mano de la estadística, consiguió desarrollar modelos de inteligencia artificial que traducían las variables en una especie de árboles de decisión en los que solo aparecían las opciones viables. Este sistema se iba perfeccionando, de manera que el programa no tenía que empezar a razonar de cero cada vez.

Los métodos de Pearl se enseñan hoy en todas las facultades de Informática y sus libros “han inspirado avances trascendentales en la comprensión del razonamiento y el pensamiento”, destacó el jurado del premio español. Su “amplio y profundo impacto” se percibe en multitud de áreas y aplicaciones, como “en el desarrollo de ensayos clínicos médicos no sesgados y eficaces, en la psicología, la robótica y la biología”.

2022 será un año agridulce para el científico estadounidense. Por un lado, ha conseguido uno de los galardones mejor remunerados del mundo; por otro, se cumplen 20 años del asesinato de su hijo, [el periodista Daniel Pearl](#), que fue secuestrado, torturado y asesinado en Pakistán por yihadistas. El trágico suceso no consiguió interrumpir la carrera investigadora de Pearl. El científico atiende por teléfono a EL PAÍS desde su casa de Los Ángeles.

Pregunta. ¿En qué momento se encuentra el desarrollo de la inteligencia artificial?

Respuesta. Creo que en poco tiempo, dentro de entre cinco y 25 años, veremos una revolución. Los ordenadores pronto se comportarán de forma más inteligente que hasta ahora, en parte gracias a los avances en la ciencia de la causa y el efecto. Los ordenadores serán capaces de explicar por qué tomaron sus decisiones, por qué es bueno que tú hagas una u otra cosa y qué va a pasar si no lo haces. Vamos hacia ordenadores capaces de explicar el mundo y a ellos mismos, y de volver hacia atrás y modificar su propio *software*. Y eso va a suceder en los próximos años. Supongo que ahora me querrá preguntar si esto es peligroso, si acabará por desarrollarse una especie de organismo que controle el mundo. Me preocupa esa posibilidad, pero también creo que deberíamos ser capaces de controlar la situación.

P. ¿Cree que la gente pone demasiadas esperanzas en el aprendizaje automático?

R. El *deep learning*, no lo olvidemos, es una forma muy simple de IA. Está en el primer escalón de complejidad, es fundamentalmente estadística: solo es capaz de predecir cosas similares a las que ya ha visto anteriormente. Es cierto que se ha confiado en exceso en el *deep learning* debido a los grandes avances que ha impulsado en visión computacional, reconocimiento de voz y vehículos autónomos, por ejemplo. Eso nos llevó a pensar que esta progresión era imparable, que la inteligencia general estaba a la vuelta de la esquina. Ahora sabemos sus limitaciones. No podemos predecir los resultados de una intervención solo mirando pasivamente los datos. Tampoco se puede generar una explicación a partir de su contemplación.



Judea Pearl trabaja en el despacho de su casa de Los Ángeles. **FUNDACIÓN BBVA**

P. ¿Cuál es entonces la principal limitación de la IA?

R. Cruzar la frontera entre la predicción y la explicación de la acción. Se trata de una diferencia que no se reconocía antes. Imagino que habrá otras grandes limitaciones; yo le puedo hablar de esta porque es con la que estoy familiarizado. Ahora sabemos cuál es el problema y creo que lo superaremos.

P. ¿Debería regularse el uso de algoritmos de IA que afecten a las vidas de las personas?

R. Creo que es prematuro legislar antes de entender qué significa protegernos de ello, qué se puede considerar un fracaso. Los abogados no deberían traducir sus miedos infundados en una legislación. Por supuesto que no queremos que la gente sea discriminada por su color de piel, origen nacional o estatus económico. Habrá que entender qué significa esta correlación. Ahora tenemos los medios para clasificar los algoritmos en discriminatorios o justos. Debería haber alguna regulación, pero no antes de entender qué es la justicia algorítmica. Y justicia es una noción causal.

P. ¿Qué quiere decir con eso?

R. La definición de justicia viene de un modelo causal. Porque ya fuera la raza o el sexo lo que aportó la apariencia de discriminación o inequidad, depende del modelo causal que tengas. Podría ser una coincidencia estadística, pero si hay intención de discriminar deben ser prohibidos. Algunos tipos de discriminación se relacionan causalmente con variables protegidas, y por tanto deberían ser regulados.

P. ¿Qué le parece que las redes bayesianas que usted desarrolló sean ahora aprovechadas por empresas privadas para amasar grandes ingresos sin mostrar su código?

R. No lo he visto todavía, pero sé que hay una tendencia a usarlas. Soy muy cauteloso con estas cosas. Por otra parte, si veo gente que desarrolla y usa adecuadamente el algoritmo que yo desarrollé, genial. No necesito que mi nombre esté en los créditos del producto. Me han dicho que hay una red bayesiana en cada iPhone, no lo sé. Google tampoco me dice cómo lo usan.

P. Si empezase ahora su carrera como investigador, ¿por qué campo se interesaría?

R. Por la medicina personalizada. Y si me pusiera más filosófico, por la inteligencia artificial social.

P. ¿Qué es la inteligencia artificial social?

R. Consiste en conseguir que los ordenadores se comuniquen entre ellos igual que hacemos las personas. Con la idea de que cada uno de nosotros es un agente con voluntad propia, tiene creencias y deseos y que como especie social trabajamos juntos y formamos una sociedad en la que los unos confíen en los otros. Hay que establecer relaciones extremadamente básicas entre los ordenadores, que tengan los planos de otros ordenadores. Hay que buscar formar empatía, confianza, responsabilidad, arrepentimiento, mérito y acusación. Si conseguimos eso, los ordenadores trabajarán mucho mejor en sociedad que cada uno por su cuenta.

P. ¿Eso podría ayudarnos a entendernos mejor a nosotros mismos?

R. Exacto. Lo más interesante de ese desarrollo sería llegar a comprender qué nos enfada, qué nos inspira confianza, por qué sentimos compasión por otros. Los ordenadores serán un laboratorio para las ideas en ciencias sociales. Podremos implementar un planteamiento en algunas máquinas, apartarlas, cambiar algún componente y ver qué pasa. Crearíamos un avance tremendo en la comprensión de nuestra vida social.

P. ¿Qué consejo le daría a quien esté ahora mismo empezando su carrera de investigación en IA?

R. No aceptes el *no* como respuesta. No te metas en el aprendizaje profundo [por el *deep learning*], métete en el entendimiento profundo. Evita el dogmatismo, recela de quienes insistan en una perspectiva y no atiendan la influencia de otros campos en el suyo. Algunas disciplinas son más abiertas, otros más cerradas... Elige una que muestre ser abierta, eso se puede ver entre otras cosas por el tipo de vocabulario técnico que emplee.