

Discurso de aceptación

16 de junio de 2022

Judea Pearl, galardonado en la categoría de Tecnologías de la Información y la Comunicación (XIV edición)

Miembros de la Presidencia, autoridades, colegas premiados y distinguidos invitados:

Agradezco a la Fundación BBVA y al jurado del premio que hayan considerado mi trabajo digno de este prestigioso galardón.

Me considero excepcionalmente afortunado por poder describir mi trabajo en términos sencillos, sin jerga matemática, porque los problemas que he intentado resolver en las últimas cuatro décadas los conocen todos y cada uno de ustedes. Basta con examinar actividades cotidianas, como conducir un coche, cruzar una calle o acudir a un médico: todo ello lo hacemos en medio de un océano de ruido e incertidumbre y, curiosamente, ustedes y yo nos las arreglamos para navegar por ese océano bastante bien, con sorprendente comodidad y gran destreza.

¿Cómo lo hacemos? Los científicos no suelen plantearse esta pregunta, porque se espera que los humanos poseamos esas aptitudes. La pregunta surge al intentar emularlas en una máquina digital que trabaja con ceros y unos, verdadero y falso. Es decir, ¿cómo podría una máquina inteligente reconocer todos los matices de gris llenos de ruido que nos rodean en la vida cotidiana e integrarlos en una comprensión coherente de en qué punto está y qué debe hacer a continuación?

Al preguntarnos *cómo lo haría una máquina*, nos hacemos una idea de cómo lo hacemos NOSOTROS, porque las máquinas son como laboratorios flexibles para probar teorías diversas del pensamiento humano y ver cuál de ellas se desenvuelve tan bien como nosotros.

Yo lo llamo *un patio de recreo del entendimiento humano*.

De hecho, la investigación en inteligencia artificial ha desvelado algunos secretos básicos del razonamiento, tanto en el hombre como en la máquina.

He descubierto, por ejemplo, que lo que llamamos *conocimiento* se compone de tres capas distintas, cada una de las cuales responde a un tipo diferente de preguntas y puede hacer cosas que la capa inmediatamente inferior no puede realizar.

La capa más sencilla es, sin duda, la predicción: al oír el canto del gallo, predecimos el amanecer. Al oír el tañido de una campana, el perro de Pavlov predice la llegada de la comida. Tanto la estadística como el aprendizaje automático destacan en esta tarea.

Pero como sabe cualquier estudiante del primer curso de estadística, correlación no es causalidad. Hacer cantar al gallo no acelera la salida del sol, y tocar la campana no da de comer al perro. Así pues, he aquí la segunda capa, la intervención, que responde a preguntas sobre las acciones y los experimentos: ¿Se me quitaría el dolor de cabeza si me tomo una aspirina? ¿Vacunarse prevendría las infecciones de covid-19?

Sin embargo, las predicciones pasivas y los experimentos activos no llegan a cubrir todo el espectro del conocimiento humano. Por encima de estos dos elementos, encontramos la capa de la imaginación, también llamada *contrafactualidad*, responsable de las explicaciones, la retrospectión, el arrepentimiento, el mérito y la culpa: ¿Y si hubiera actuado de otra forma y hubiera tomado el camino que no tomé? ¿Y si el tren hubiera salido a tiempo? ¿Fue la aspirina la que me curó el dolor de cabeza, o fue la buena noticia que me trajo mi mujer? ¿He herido sus sentimientos? Estos enunciados *contrafácticos* son los bloques de construcción del pensamiento científico, la confianza social y el comportamiento moral.

Pero ¿cómo puede una máquina lógica preguntarse por universos alternativos y calcular lo que habría sucedido si los hechos reales no hubieran ocurrido? Lo que mis colegas y yo hemos hecho en las tres últimas décadas ha sido plantear y analizar esta difícil cuestión filosófico-computacional en toda su complejidad. Y me enorgullezco de poder decirles que, hoy en día, tenemos la teoría y los algoritmos que permiten a un robot realizar esos cálculos.

Esta capacidad de imaginar mundos hipotéticos y utilizarlos para razonar tiene enormes repercusiones prácticas y científicas. Permite a las máquinas predecir el efecto de acciones y políticas que no se han probado nunca; por ejemplo, qué pasa si prohibimos el consumo de cigarrillos, o qué pasa si utilizamos un fármaco nuevo en un nuevo país. También nos permite predecir el efecto de los tratamientos en un paciente individual que no había sido tratado antes, lo que inaugura la era de la medicina personalizada y la toma de decisiones personalizada.

Desde el punto de vista científico, esta capacidad nos permite conocernos mejor a nosotros mismos: cómo aprenden los niños una disciplina nueva, y cómo generalizamos lo que hemos aprendido de una materia para aplicarlo a otra disciplina. Para decirlo de forma más poética, por primera vez en la

historia podemos *entender* lo que significa entender, ahora que tenemos un modelo computacional de la *comprensión profunda*.

Me gustaría decir unas palabras sobre lo que implica el premio de la Fundación BBVA para la teoría que les he descrito. No es ningún secreto que mi investigación se ha topado con su cuota de escepticismo: por un lado el escepticismo de quienes ven en ella una amenaza a la sabiduría convencional, y por otro el de quienes no han tenido ocasión de examinar su potencial.

Espero que el prestigio y la visibilidad de este premio anime a los científicos en activo de todas las disciplinas a evaluar las poderosas herramientas que les proporciona la inferencia causal.

Aludía concretamente a los problemas de transparencia, generalización y explicabilidad, los tres obstáculos que impiden que se acepten sistemas de inteligencia artificial para aplicaciones empresariales, sanitarias o jurídicas. Es comprensible que resulte difícil confiar en una *caja negra* si no podemos saber qué supuestos han llevado a su decisión o recomendación. Afortunadamente, estos tres escollos ya pueden eliminarse con la ayuda de las capas segunda y tercera de causalidad.

Quiero terminar con un enorme mensaje de gracias a todos mis alumnos, colegas y familia por su constante apoyo a lo largo de mi investigación. Y a mis grandes maestros de Israel, que me educaron en el amor a la ciencia y a la humanidad.

Gracias.