

Ricardo de Querol

Gemma Galdón, auditora de algoritmos:

“La inteligencia artificial es de muy mala calidad”

El País, 11 de diciembre de 2023.

La fundadora de Éticas asesora a organismos internacionales para evitar la discriminación. Desconfía de las expectativas del sector: “Plantear que un sistema de datos va a dar un salto a la consciencia es una alucinación”.

La inteligencia artificial no es solo cosa de ingenieros. Se puede ser de letras y, a la vez, una referencia en el debate mundial sobre las repercusiones sociales y éticas de lo que hacen estos sistemas. [Gemma Galdón](#) (Mataró, Barcelona, 47 años) se graduó en Historia Contemporánea y se doctoró en políticas públicas tecnológicas; es la fundadora y primera ejecutiva de Éticas Research and Consulting, compañía que examina los algoritmos para asegurar su uso responsable. “Ser consciente de cómo la sociedad ha resuelto problemas antiguos me da una perspectiva útil para trabajar con problemas nuevos”, cuenta en una cafetería de Madrid. “Hace 12 años, cuando me doctoré, éramos muy pocas las personas de las ciencias sociales que trabajáramos con tecnología”. Su empresa asesora a organismos europeos y estadounidenses. Tiene la maleta hecha: enseguida regresa a Nueva York, donde reside y donde el jueves recibió uno de los [Hispanic Star Awards](#), premios a los agentes del cambio de la comunidad hispanohablante, en un acto en Naciones Unidas. Tuvo que mudarse, cuenta, porque en EE UU “el mercado está más receptivo a la IA responsable”.

Pregunta. ¿Cómo es eso de auditar algoritmos?

Respuesta. Pues es inspeccionar sistemas de inteligencia artificial para ver cómo funcionan, pero en primer lugar para asegurar que sus impactos en la sociedad son justos, que no hay [ninguna discriminación](#). Y, además, que los sistemas hacen lo que dicen que hacen.

P. ¿Y qué problemas se encuentra?

R. Al principio estos sistemas son tan discriminatorios como la sociedad, pero al cabo de muy poco tiempo ya son mucho más discriminatorios que la sociedad. Porque lo que hace la IA es coger muchos datos de entrenamiento y buscar un patrón. Y el patrón siempre es el hombre blanco con trabajo estable; en el caso de los bancos, será el cliente ideal. Cualquier perfil que sea minoritario o anecdótico lo elimina de la muestra. Con lo cual una mujer tiene muchas menos posibilidades de que le diagnostiquen endometriosis mediante IA, porque históricamente no hemos diagnosticado endometriosis.

P. Hay quien dice que no se puede examinar a fondo a la IA porque ni sus creadores entienden del todo cómo funciona, sino que aprende sola.

R. Falso. Esa idea de la caja negra es un mito, puro *marketing*. Creo que hay cierta voluntad por parte del sector de la IA de que se la vea como algo mágico, que nos creamos que es algo que no podemos entender y quitarnos la capacidad de intervenir. Lo que hemos visto es que podemos auditar cuando nos contrata un cliente y nos enseña prácticamente todo, pero también desde fuera podemos hacer ingeniería inversa y ver cómo funciona un sistema en función de sus impactos.

P. Ha asesorado a instituciones políticas para [regular la IA](#). ¿Qué es lo que pretenden?

R. Lo que ha pasado en los últimos años es que el legislador, con muy buenas intenciones, ha generado una regulación muy abstracta, muy basada en principios, y la industria se ha quejado de no tener prácticas concretas. Tenemos una industria nacida al albur de Silicon Valley, acostumbrada a esa idea de “muévete rápido y rompe cosas”, sin ser consciente de que lo que pueda romper sean derechos fundamentales o leyes. A veces hay una cierta obsesión con pedir el código o los modelos fundacionales. A mí nunca me han sido útiles. Estamos pidiendo un nivel de transparencia que no es útil para auditar, para inspeccionar los impactos. Si sabes que hay un momento de inspección en el que te valoramos ciertas métricas, tienes que empezar a hacer cambios. Con lo cual cambiamos los incentivos de la industria tecnológica para que tengan en cuenta el impacto y el sesgo, cualquier tipo de disfunción.

P. ¿Le decepciona o le satisface [la ley de IA](#) que ha acordado la Unión Europea?

R. Me parece un paso de gigante en la regulación: es la primera ley sobre estos temas en Occidente. Lo que me decepciona es el papel de Europa a la hora de ir más allá, de crear un mercado vinculado a la IA responsable. Tanto Estados Unidos como Asia y China se están poniendo las pilas.

P. ¿Es inteligencia artificial todo lo que se presenta como tal?

R. Estamos rodeados de inteligencia artificial de muy mala calidad. Ya no es un tema de sesgos, es que no hace lo que dice que hace, y toma decisiones que los humanos nunca tomaríamos. Un ejemplo es el sistema que se implantó para valorar el desempeño de los profesores en el sistema educativo de varios estados de EE UU. Algunos trabajadores que vieron cómo su desempeño cambiaba en la valoración manual y en la algorítmica lo llevaron a los tribunales. El tribunal ordenó una auditoría y se observa que los únicos *inputs* que se tienen en cuenta para decidir si eres buen profesor son los resultados de tus alumnos en matemáticas y en lengua. Es un Excel glorificado. Si a los directores de esas escuelas les hubieran ofrecido esto como una hoja de cálculo que apunta los resultados en matemáticas y en lengua, nunca se lo habrían vendido.

P. ¿Se impondrá una IA responsable?

R. Soy optimista. Cuando auditamos, nos encontramos con sistemas sesgados que también funcionan mal. La inteligencia artificial es de muy mala calidad y en algún momento la industria va a tener que hacerla mejor. Estos sistemas nacieron de [herramientas de entretenimiento como Netflix](#), que pueden tener un margen de error elevado. Si la película que te recomienda Netflix no es la que quieres ver después de otra, no pasa nada. Pero si la IA quiere trabajar en el ámbito médico recomendando un tratamiento; o en la selección de personal decidiendo a quién contratamos o a quién despedimos; o en la asignación de recursos públicos... tiene que funcionar bien. Ahora mismo, la IA que aceptamos no solo es sesgada, sino que, además, no funciona bien. Lo bueno es que los dos problemas se resuelven a la vez. Cuando se aborda el problema del sesgo, se abordan también las otras ineficiencias.

P. La salida y reincorporación de Sam Altman como consejero delegado de OpenAI se han relacionado con [un supuesto avance sensacional](#) hacia la Inteligencia Artificial General (IAG), o superinteligencia, algo que amenaza a la humanidad. ¿Se lo cree?

R. La Inteligencia Artificial General está tan cerca como cuando Platón hablaba de las posibilidades de otros tipos de mundos y de vidas. La humanidad siempre ha soñado con reproducir de manera automática la conciencia humana. Hemos sido capaces de soñar

futuros de ciencia ficción. Hay un debate sobre la IAG que no tiene nada que ver ahora mismo con las capacidades tecnológicas.

P. ¿No van a superarlos las máquinas a los humanos?

R. La forma en que pensamos los humanos, la creatividad, lo nuevo, no tiene nada que ver con la IA. Un ejercicio muy sencillo: si le damos a un sistema toda la obra de Picasso antes de 1937 y le preguntamos: ¿cuál va a ser la siguiente pintura de Picasso?, nos sacará cualquier cosa. Y en 1937 él pinta el *Guernica*. Las personas evolucionamos en nuestra forma de expresarnos, de amar, de trabajar, de escribir, de crear. Plantear que en algún momento un sistema de datos estadístico y matemático va a dar un salto a la consciencia es una alucinación.

P. También se llama alucinación a [lo que hace ChatGPT](#) cuando se inventa respuestas a las preguntas. Es poco fiable, ¿verdad?

R. Es el caso de un abogado que trabaja defendiendo a víctimas de pederastia y ChatGPT le hace una biografía de pederasta. ¿Por qué? Porque su nombre aparece con estas palabras la mayoría de veces o más veces con estas palabras que con otra, con lo cual te asocia esta palabra y ya está.

P. Usted estudia el impacto social de la IA. ¿Y qué hay del impacto ecológico? Porque los centros de datos se han convertido en [grandes derrochadores de agua y energía](#).

R. No tiene ningún sentido que ahora mismo, cuando te hacen una auditoría ambiental en tu empresa, vengan a ver qué tipo de bombillas tienes y no te miren dónde están los servidores y cuánto tiene que viajar la información. No ha habido ninguna voluntad de cuantificar el impacto ambiental de los procesos de datos y de incentivar que la industria tenga servidores cercanos al lugar de provisión de la información. Es un debate que aún no hemos tenido. En la era del cambio climático no tiene sentido que casi todo el mundo hable de la tecnología como la solución y no como uno de los problemas.

P. De [criptomonedas](#), entonces, ni hablamos. Con lo que gastan.

R. Igual que estamos sacando coches contaminantes de la calle, en algún momento nos tenemos que plantear sacar tecnologías contaminantes de la circulación. Igual tenemos que empezar a prohibir arquitecturas *blockchain* cuando el valor social no se percibe. Lo que aportan las criptomonedas es una herramienta de especulación, un mecanismo de inversión más parecido a una estafa piramidal... Si estuviéramos salvando vidas diría: mira, pues igual se justifica.