

TESTIMONIOS

**Eneko Agirre**  
Director de Hitz



«No saben que les escribes en castellano. Te responden así por probabilidad»

**Mikel Díaz**  
Directivo de IBM



«Algunos robots de limpieza también usan Inteligencia Artificial»

**Alex Rayón**  
Experto en IA



«Ajustan sus respuestas como los 'súper' ajustan los precios de los productos»

o penalizadas. «Así es como los robots de limpieza al principio se chocan contra los muebles y al poco tiempo ya no lo hacen», continúa Díaz.

Limitaciones

En el caso de los modelos de lenguaje, ya no se trata de ir palabra a palabra. Se les enseñan tareas como escribir un correo electrónico al jefe, un currículum para un trabajo o explicar cómo aterrizaba una nave en la luna a un niño de seis años. «Se le dan ejemplos de lo que tiene que contestar y se le indica que no tiene que dar determinadas respuestas. No le estás diciendo qué tiene que decir cada vez, sino 'esto que has dicho está bien y esto está mal'. Así consigues las respuestas que quieres», afirma el experto de Hitz, que pone otro ejemplo: «Es como el niño que lee de todo y empieza a decir palabrotas -aprendizaje por supervisión- y otro niño con un educador que le dice lo que tiene que decir y otro que le dice no, eso está mal».

«Son muy listos y muy tontos a la vez. Listos, porque gracias a todo lo que han 'leído' saben mucho sobre el mundo. Cuanto más se les enseña, mejor hacen lo que se les pide. Y tontos porque realmente no entienden lo que dicen. Si les hablas en castellano, te responden en este idioma porque su cálculo de probabilidades le dice que es lo más acertado, pero no saben que hablan en castellano. Tampoco saben reaccionar ante un problema complejo, caen en contradicciones y no pueden generalizar, solo recopilan millones de datos y extraen patrones...», subraya Agirre.

Un experimento reciente ha demostrado que la IA sí puede ser capaz de generalizar, de relacionar ideas como hacemos los humanos. «Tal y como es ahora, ChatGPT no podrá hacerlo por mucho tiempo que dedique a recabar información».

# Cómo la IA convirtió al Papa en rapero

Uno de los grandes peligros de esta tecnología es la manipulación de imágenes. Un experto explica cómo lo hacen

J. GARAY

BILBAO. La manipulación de imágenes es una de las capacidades más llamativas de las nuevas herramientas de Inteligencia Artificial. ¿Cómo logran estas herramientas desnudar digitalmente a una treintena de chicas de Almedralejo, detener a Trump antes de que fuera arrestado o vestir al papa Francisco con una parca de moda entre los raperos?

Estas dos últimas fueron creadas con un programa llamado Midjourney -la primera, la de las menores de Badajoz se hizo con una aplicación para móvil-. Si ChatGPT trabaja con textos, Midjourney -y otros como Dall-e o Stable Diffusion- genera una imagen a partir de unas instrucciones que se llaman 'prompts'. «Es como desmontar una foto para volver a montarla», explica Gorka Azkune, profesor en Hitz, el Centro Vasco de Tecnología de la Lengua de la UPV/EHU. La tecnología clave que está detrás de este proceso son los modelos de difusión. «La idea es coger la imagen real -el papa Francisco, por ejemplo- e ir añadiéndole ruido paso a paso. Esto consiste en coger píxeles -los puntos que forman las imágenes- y cambiarles el color. Al principio siguen siendo reconocibles -el papa sigue siendo el papa-, pero progresivamente se le añade más ruido hasta que al final queda completamente difuminada. Algo así como el Canal + cuando estaba codificado», explica el experto.

Con esto se consigue una secuencia que parte de una imagen real y llega hasta otra en la que no se distingue nada, con todos los pasos intermedios.

5.000 millones de imágenes

En este punto entran en juego una red neuronal, una herramienta estadística de gran complejidad a la que se enseña a recuperar la imagen original. «Es como ir marcha atrás. Coge la imagen con ruido y se lo va quitando paso a paso hasta conseguir de nuevo la imagen real». Tras desmenuzar la imagen del pontífice, volveríamos a verla como al principio. Para entrenar a estas redes neuronales, se utilizan bancos de datos públicos en Internet que contienen 5.000 millones de imágenes.

El proceso empieza entonces con el 'prompt', en este caso, algo similar a «imagina al papa sentado en una silla con una parca blanca y unas zapatillas deportivas». Una red neuronal diferente a la antes citada convierte el texto en números -vectores-. Hecho esto, la segunda red neuronal -la del modelo de difusión- tiene ahora tanto la orden del usuario como las imágenes de Francisco y el abrigo, estas gracias a la gigantesca base de datos que maneja. Las descompone y recompone para juntarlas en un montaje que se puede ir ajustando con más 'prompts' hasta conseguir el resultado buscado.



El papa Francisco, en el montaje con la parca blanca. MIDJOURNEY

**CIRCO DE LOS HORRORES**  
PRESENTA

**REQUIEM**  
SINFONIA FINAL

**BILBAO, PLAZA DE TOROS**  
DEL 22 NOV. AL 10 DIC.

**WWW.CIRCODELOSHORRORES.COM**

COLABORA: 