

OPINIÓN / DESPUÉS DE LA PANDEMIA

Getty Images y la lógica de los datos

por Enrique Dans • X

10 enero, 2024 - 02:31

GUARDAR



No puede decirse que la relación del repositorio de imágenes digitales Getty Images y la inteligencia artificial generativa sea muy sencilla. En 2022, cuando comenzaron a presentarse los **primeros algoritmos de creación de imágenes** a partir de texto, la compañía denunció a uno de ellos, Stable Diffusion, porque cuando se le pedían determinadas imágenes, añadía también algo que se asemejaba sospechosamente a la marca de agua que Getty Images pone a sus imágenes para que los usuarios puedan verlas en su página web, pero no puedan descargarlas y utilizarlas sin pasar por caja y pagar la correspondiente factura.

Para entenderlo, es preciso explicar cómo funciona el entrenamiento de estos algoritmos: para llevarlo a cabo, no basta con una gran colección de imágenes. Además, **esas imágenes deben estar convenientemente etiquetadas** con palabras clave que describan su contenido.



PUBLICIDAD

Primero, porque un algoritmo no ve imágenes, sino mapas de píxeles. Y segundo, porque es preciso introducir la correspondencia entre una imagen o un tipo de imágenes determinadas y sus descripciones: solo tras haber introducido un cierto número de imágenes etiquetadas como "gato", puede el algoritmo empezar a aprender que un gato se corresponde con una imagen de un tipo determinado.

Si añadimos que hay gatos de muchos tipos, de diferentes colores y que pueden estar en infinidad de posturas, es fácil imaginar la cantidad de imágenes etiquetadas que hacen falta.

Ante semejante necesidad, repositorios como Getty Images

representaban, para las compañías que se afanaban en desarrollar esos algoritmos, el auténtico El Dorado: colecciones gigantescas de imágenes correctamente etiquetadas para que sus usuarios pudieran encontrar lo que buscaban. Al entrar en Getty Images y escribir "gato", obtienes automáticamente una enorme colección de gatos de todo los tipos, de todos los **colores**, de todos los **tamaños** y en todas las **posiciones**. Y además, tanto imágenes como descripciones estaban libremente accesibles en páginas abiertas... eso sí, con su correspondiente marca de agua.

Primero, porque un algoritmo no ve imágenes, sino mapas de píxeles. Y segundo, porque es preciso introducir la correspondencia entre una imagen o un tipo de imágenes determinadas y sus descripciones

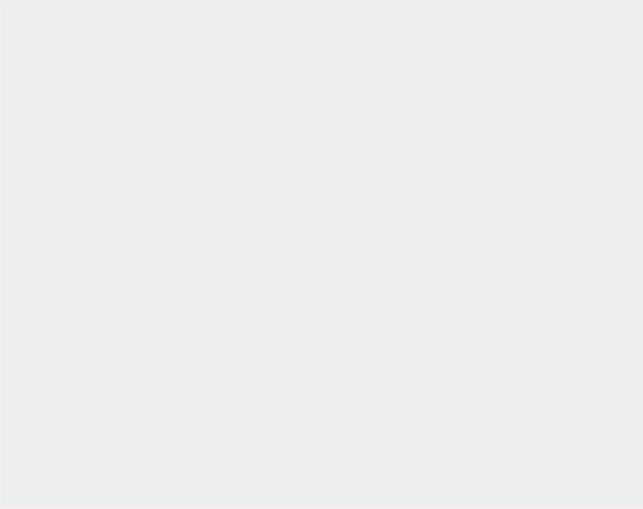
Dado que existía jurisprudencia sobre la posibilidad de utilizar información disponible en abierto en páginas web, los desarrolladores no lo pensaron dos veces: entrenaron a sus algoritmos, entre otras muchas cosas, con el repositorio de imágenes de Getty Images. El resultado es conocido: de vez en cuando, al pedirle que nos genere una **imagen determinada**, el algoritmo añade "algo" que se parece a su marca de agua... estaba en tantas imágenes, que el algoritmo pensaba que debía ser parte de ellas.

Pero hasta aquí, y hasta que se celebre el juicio correspondiente, y se determine si es Getty Images o la compañía denunciada, Stability AI, quien tiene razón, es historia.

Lo interesante viene después: Getty Images acaba de presentar Generative AI by Getty Images, un algoritmo entrenado exclusivamente con su repositorio de imágenes, y **que ha desarrollado conjuntamente con Nvidia**, la compañía que más creció en 2023 precisamente por ser la que diseña los chips más utilizados para el entrenamiento de algoritmos generativos.

Esto es, precisamente, hacer de la necesidad virtud. Si tienes unos datos que eran la base de tu negocio, que te ha llevado muchos años y mucho trabajo conseguir, y te encuentras con que llega un tercero y los utiliza para algo que tú ni siquiera sabías que se podía hacer... ¡haz lo mismo! Toma tus propios datos, y ponlos en valor haciendo lo mismo que han hecho esas compañías, con lo que habrás generado un producto complementario al tuyo principal que ahora te permite no solo que tus clientes **exploren tu repositorio para encontrar la imagen** que buscan, sino que además, si no la encuentran, puedan generarse una que satisfaga sus necesidades.

PUBLICIDAD



Con esta funcionalidad, Getty Images saca más partido a sus propios datos, pero además, puede alegar ante el juez que lo que Stability AI ha hecho es utilizar ese repositorio para hacerle la competencia, un argumento potencialmente más contundente que podría ayudar a la compañía a obtener un acuerdo extrajudicial o un veredicto más ventajoso. Porque no nos engañemos: de lo que se trata aquí, y lo que pretende Getty Images, no es que Stability AI deje de utilizar su repositorio, algo que en realidad ya ha hecho, sino de ponerle el precio correspondiente.

Pero hasta aquí, y hasta que se celebre el juicio correspondiente, y se determine si es Getty Images o la compañía denunciada, Stability AI, quien tiene razón, es historia.

Lo que Getty Images ha hecho debería **convertirse en un modelo para muchas otras compañías**. Hay muchísimas empresas, empezando por las periodísticas, que están sentadas sobre enormes repositorios de datos, en muchos casos, además, adecuadamente etiquetados.

Si es tu caso, si tu compañía posee algún repositorio de ese tipo obtenido

gracias a su actividad, o como parte de un producto o servicio determinado, tienes rápidamente que ponerte a pensar en cómo ponerlo en valor mediante la algoritmia generativa, o vendrán otros y lo harán por ti. *The New York Times*, sin ir más lejos, acaba de denunciar a OpenAI y a Microsoft solicitando daños estimados en miles de millones de dólares, porque la compañía tomó las noticias del periódico, y las utilizó para entrenar su algoritmo que ahora pretende competir con el periódico como fuente de información fiable.



Si aún no lo has considerado, es hora de empezar. Plántate qué datos tiene tu compañía, si están correctamente etiquetados o podrían estarlo con un cierto esfuerzo adicional, y qué valor podrían llegar a tener. En el momento actual, pocas cosas pueden tener más sentido.

******Enrique Dans es Profesor de Innovación en IE University.***

 SIGUE LOS TEMAS QUE TE INTERESAN

 COLUMNAS DE OPINIÓN  INTELIGENCIA ARTIFICIAL  TECNOLOGÍA