



La IA entrenada con su propio conocimiento no produce nada

## Description

El riesgo de la IA entrenada con su propio conocimiento reside en la posibilidad de crear bucles de retroalimentación que refuercen sesgos y errores.

## CONTENIDOS

# La IA Entrenada con su Propio Conocimiento: La Insaciable Búsqueda de Datos

Los grandes modelos de lenguaje, como los que sustentan a ChatGPT, son sistemas de inteligencia artificial diseñados para procesar y generar texto de manera similar a los humanos. Para lograr esto, requieren ingerir cantidades masivas de datos textuales. Estos datos sirven como el material de construcción con el que los modelos aprenden a reconocer patrones, estructuras gramaticales y semánticas. Sin embargo, esta voracidad por los datos ha planteado un desafío cada vez más apremiante: el agotamiento de fuentes de información nuevas y diversas. A medida que los modelos exploran cada rincón de internet, se vuelve cada vez más difícil encontrar contenido original y de alta calidad que sirva para enriquecer sus conocimientos.



Al entrenar a la IA con su propio conocimiento, corremos el riesgo de crear sistemas que no se adapten a nuevos datos o información.

## Una Solución Tentativa: La Autoalimentación

---

Ante la escasez de datos externos, algunas voces prominentes en la industria, como el CEO de Meta, Mark Zuckerberg, han [propuesto una solución a ese dilema de datos](#): entrenar a los nuevos modelos con los textos generados por las versiones anteriores. La idea es que, al aprender de sus propias creaciones, los modelos puedan mejorar su capacidad de generar texto de forma autónoma y exponencial. Sin embargo, esta estrategia, a pesar de su atractivo inicial, oculta riesgos significativos que podrían comprometer el futuro de la inteligencia artificial.

## El Espectro del Colapso del Modelo de la IA Entrenada con su Propio Conocimiento

Investigaciones recientes han revelado un fenómeno preocupante conocido como “colapso del modelo”. Este fenómeno se produce cuando un modelo de lenguaje, al ser entrenado repetidamente con sus propias salidas, comienza a generar textos cada vez más incoherentes, repetitivos y alejados de la realidad. Es como si el modelo quedara atrapado en un bucle infinito, replicando sus propios errores y perdiendo la capacidad de aprender de nuevas experiencias.

Te Puede Interesar:

### Un Experimento Revelador: La Degradación Gradual

[Un estudio pionero, publicado en la prestigiosa revista Nature](#), ha proporcionado evidencia empírica del colapso del modelo. En este estudio, los investigadores crearon una serie de modelos de lenguaje, cada uno entrenado con las salidas del modelo anterior. Los resultados fueron sorprendentes: a medida que avanzaban las generaciones, el texto generado se volvía cada vez más abstracto, menos coherente y más alejado del texto original. En las últimas generaciones, el modelo llegó a producir un galimatías sin sentido, demostrando de manera contundente la degradación gradual que se produce cuando un modelo se alimenta de sus propias creaciones.

### La IA Entrenada con su Propio Conocimiento: Los Mecanismos Detrás del Colapso

¿Por qué ocurre el colapso del modelo? Los investigadores sugieren que este fenómeno se debe a una combinación de factores. En primer lugar, al entrenar a un modelo con sus propias salidas, se introduce un sesgo cada vez más fuerte hacia ese estilo particular de generación de texto. Con el tiempo, este sesgo domina por completo al modelo, impidiéndole explorar nuevas posibilidades y generar contenido verdaderamente original. Además, los errores que se producen en las primeras generaciones de modelos se amplifican y perpetúan en las generaciones posteriores, dando lugar a una espiral descendente de calidad.

La [posibilidad de un colapso del modelo](#) del mundo real no es una hipótesis descabellada. Actualmente, existen [incontables sitios web con artículos](#) y entradas de blog creados completamente por modelos de lenguaje de gran escala. En el afán de desarrollar nuevos modelos con la mayor rapidez posible, no es inverosímil que una gran cantidad del contenido de baja calidad producido por la IA acabe incorporándose en los conjuntos de datos de entrenamiento.



Un desafío clave en el desarrollo de la IA es evitar que la IA entrenada con su propio conocimiento se vuelva demasiado especializada y pierda su capacidad de generalización.

## Para seguir pensando

El colapso del modelo plantea un serio desafío para el desarrollo futuro de la [inteligencia artificial](#). Si no se toman medidas para evitarlo, podríamos terminar con modelos de lenguaje que generan contenido repetitivo, sesgado y poco útil. Para mitigar este riesgo, es necesario explorar nuevas estrategias de entrenamiento que combinen datos reales y generados por modelos, así como desarrollar técnicas para detectar y corregir los sesgos que se producen durante el proceso de aprendizaje. Además, es fundamental establecer estándares más rigurosos para la evaluación de la calidad de los modelos de lenguaje, garantizando que estos sean capaces de generar contenido

---

relevante, preciso y confiable.