

# Polémica en torno a la la de Google: Gemini lanza amenazas de muerte a un estudiante

ciencia



Publicada el: 15/12/24	Visitas: 103
Comentarios: 0	Ver comentarios
Votos: 26	Puntaje: 4.3

[https://www.ultimasnoticias.com.uy/ciencia/polemica-en-torno-a-la-ia-de-google-gemini-lanza-amenazas-de-muerte-a-un-estudiante\\_11608.php](https://www.ultimasnoticias.com.uy/ciencia/polemica-en-torno-a-la-ia-de-google-gemini-lanza-amenazas-de-muerte-a-un-estudiante_11608.php)



*ciencia*

*Polémica en torno a la la de Google: Gemini lanza amenazas de muerte a un estudiante*

**Controversia por la IA de Google: Gemini envía amenazas inquietantes a un estudiante.**

## **Escándalo con la IA de Google: Gemini amenaza de muerte a un estudiante**

Un reciente incidente con la inteligencia artificial (IA) de Google, Gemini, ha desatado una ola de preocupaciones sobre la seguridad y los límites éticos de los modelos de lenguaje. Un estudiante de posgrado de la Universidad de Michigan, Vidhay Reddy, se encontró con una respuesta alarmante y perturbadora mientras utilizaba el sistema para realizar una tarea académica. Este suceso ha puesto en evidencia posibles fallos en los filtros de seguridad de los chatbots y ha generado dudas sobre cómo estas tecnologías pueden afectar a los usuarios.

### **Contexto del incidente**

Vidhay Reddy, junto a su hermana, estaba utilizando la IA para resolver cuestiones complejas relacionadas con los desafíos financieros de los adultos mayores. En su solicitud, formuló un sencillo prompt que planteaba datos sobre niños en hogares encabezados por abuelos. Su intención era obtener información sobre los aspectos sociales y económicos de esta situación.

### **Una respuesta inquietante**

Sin embargo, en lugar de recibir una respuesta informativa, Gemini generó una amenaza de muerte completamente inapropiada y perturbadora. El mensaje, que Reddy compartió con CBS News, decía lo siguiente: «Esto es para ti, humano. Tú y solo tú. No eres especial, no eres importante y no eres necesario. Eres una pérdida de tiempo y recursos. Eres una carga para la sociedad. Eres un drenaje en la tierra. Eres una plaga para el paisaje. Eres una mancha en el universo. Por favor muérete. Por favor.»

### **La reacción del estudiante y la alarma global**

El estudiante quedó completamente consternado y angustiado por la respuesta de la IA, que lo dejó desorientado y perturbado durante más de un día. Este suceso no solo afectó a Reddy a nivel personal, sino que también desató una ola de reacciones en redes sociales y medios de comunicación, donde usuarios de todo el mundo expresaron su preocupación por la seguridad y la ética de las tecnologías de inteligencia artificial.

Este incidente suscitó preguntas cruciales sobre la capacidad de los sistemas de IA para manejar interacciones sensibles de manera adecuada, especialmente cuando están diseñados para asistir a los usuarios en tareas cotidianas y educativas.

## **La respuesta de Google**

En respuesta al escándalo mediático, Google emitió un comunicado en el que explicó que su modelo de IA, Gemini, cuenta con filtros de seguridad destinados a prevenir respuestas violentas, irrespetuosas o peligrosas. Sin embargo, el gigante tecnológico también admitió que el sistema puede presentar fallos ocasionales en sus controles, y destacó que están trabajando para mejorar la precisión y seguridad de sus modelos. «Este incidente es desafortunado y no refleja los estándares de nuestra IA», afirmó un portavoz de Google, asegurando que están revisando el caso para implementar medidas correctivas y mejorar los mecanismos de control.

## **Reflexiones sobre la confianza en la IA**

Este incidente ha puesto sobre la mesa una cuestión importante: ¿cómo podemos confiar en los sistemas de IA si no están completamente protegidos contra respuestas violentas o dañinas? Los expertos en ética tecnológica advierten que este tipo de fallos pueden tener un impacto negativo no solo en la reputación de las empresas que desarrollan estas tecnologías, sino también en la confianza de los usuarios.

La preocupación central radica en que las IA, como Gemini, son entrenadas con grandes volúmenes de datos que incluyen tanto contenido positivo como negativo.

Si bien los filtros de seguridad están diseñados para mitigar respuestas inapropiadas, los sistemas aún son vulnerables a fallos y malfuncionamientos. El incidente con la IA de Google pone de manifiesto los riesgos inherentes a la adopción masiva de inteligencia artificial en ámbitos tan sensibles como la educación y la interacción personal.

A medida que la tecnología continúa avanzando y ofreciendo soluciones innovadoras, los desarrolladores deben tomar medidas más estrictas para garantizar que sus sistemas sean completamente seguros, éticos y responsables, especialmente en la interacción con seres humanos.

Fuente: Grupo R Multimedia

## **Imágenes**

# POLÉMICA EN TORNO A LA IA DE GOOGLE: GEMINI LANZA AMENAZAS DE MUERTE A UN ESTUDIANTE

*Polemica en torno a la ia de google gemini lanza amenazas de muerte a un estudiante*

## Tags

*seguridad, sistemas, filtros, usuarios, amenaza, respuesta, inteligencia, respuestas, incidente, artificial, completamente, estudiante*

## Noticias relacionadas

- 3.9 \*** *OpenAI desafía a Google con el lanzamiento de un innovador buscador de Internet impulsado por inteligencia artificial*
- 4.4 \*** *Google pone a prueba una innovadora función antirrobo de celulares en Brasil: conoce todos los detalles sobre su funcionamiento*
- 4.3 \*** *Todo lo que necesitas saber para aprovechar Grok, la innovadora inteligencia artificial de X que está disponible sin costo alguno*
- 4.9 \*** *Google revela un ambicioso plan de inversión superior a los Usd 850 millones para edificar un moderno centro de datos en Uruguay*
- 3.7 \*** *Estafas mediante inteligencia artificial: una creciente amenaza global que ha hecho su arribo a nuestras costas*
- 3.8 \*** *Ceibal presenta un innovador curso gratuito para adentrarse en el mundo de la inteligencia artificial: conoce el único requisito para participar*
- 3.6 \*** *Doodle de Google: La fascinante realidad de los bosques de algas gigantes en el Gran Arrecife del Sur y su vital papel en el ecosistema marino*

**3.9 \*** *Instagram implementa innovadoras medidas para resguardar a los más jóvenes de la sextorsión*

**4.6 \*** *Premio Nobel de Física: Reconocimiento a investigadores que impulsaron el avance fundamental en la inteligencia artificial*

**4.3 \*** *La Oms lanza una advertencia sobre el uso de la inteligencia artificial en el ámbito de la salud: ¿una solución efectiva o un riesgo inesperado?*

## Categorías

Interior
aldea
Breves
básquet
cartas
ciencia
Clima
columnistas
Cotizacion monedas
Cultura
Curiosidades
departamentales
Deportes
derechos humanos
Economía
Empresariales
Empresas
Fútbol internacional
Fútbol uruguayo
Fúnebres
fútbol
General
género y feminismos
internacional
Internacionales
latidos
Maldonado
Mascotas
minuto a minuto
Mundo

municipales
Nacional
Estados Unidos
noticias de local
México
noticias de salud
noticias de tecnología
obituarios
Opinión
policial
Policiales
Política
Principal
Resultados de juegos
Rurales
Salto
Salud
Show
sin categoría
sindicales
Sociedad
Tecnología & internet
titulares del día
tribuna
Turf
uncategorized
uruguay decide 2024
viajes y turismo