LITORALPRESS

Fecha: 27-05-2026
Medio: El Mercurio
Supl. : El Mercurio - Cuerpo A
Tipo: Noticia general
Título: El cofundador de Anthropic que acompañó al Papa en presentación de su encíclica

Pág. : 4
Cm2: 214,8
VPE: $ 2.821.730

Tiraje: 126.654
Lectoría: 320.543
Favorabilidad: No Definida

{ PERFIL | Christopher Olah }

El cofundador de Anthropic que acompañó al Papa en presentación de su encíclica

EL COMERCIO/PERÚ/GDA Y AGENCIAS

Durante la presentación de su encíclica *Magnifica humanitas*, el lunes en el Vaticano, el Papa León XIV estuvo acompañado por expertos en inteligencia artificial (IA, el tema principal de su texto), entre ellos el cofundador de la compañía Anthropic, Christopher Olah.

"Necesitamos que más del mundo —comunidades religiosas, sociedad civil, académicos, gobiernos— haga lo que Su Santidad ha hecho aquí: tomarse esto en serio, mirar de cerca y empujar los acontecimientos en una mejor dirección", dijo Olah, durante la presentación, y agregó: "Necesitamos críticos informados que les digan a los laboratorios cuándo estamos fallando. Necesitamos voces morales que los incentivos no puedan doblegar".

Olah tiene 33 años, es canadiense y puede ser considerado un producto directo del entorno tecnológico estadounidense. A pesar de haber abandonado la universidad a los 18 años, un año más tarde recibió una beca de la fundación de Peter Thiel —cofundador de PayPal— que le permitió formarse en investigación y proyectos independientes.

El joven trabajó en Google Brain y luego fue parte de OpenAI, donde lideró un equipo de análisis de redes neuronales, siendo su campo de especialidad la interpretabilidad mecanicista. Su actual labor al interior de Anthropic sigue enfocada en este campo, el cual busca comprender cómo funcionan los mecanismos internos de la inteligencia artificial.

THE ASSOCIATED PRESS

CHRISTOPHER OLAH, esta semana en el Vaticano, donde fue parte de la presentación de *Magnifica humanitas*.

En el tiempo reciente, la empresa ha vivido un tenso pulso con el gobierno del Presidente Donald Trump, por su negativa a retirar sus limitaciones al uso militar de su sistema de IA.

No puede ser utilizado para dañar a los humanos

El punto clave de esta discrepancia se sitúa alrededor del modelo Claude de Anthropic, el cual se rige por una constitución propia dentro de su sistema que —entre otras cosas— le impide ser usado para dañar físicamente a seres humanos y contribuir a la creación de armas. Trump busca retirar esta restricción para que este modelo, ampliamente usado en el Departamento de Defensa, pueda emplearse de forma más agresiva en las operaciones bélicas emprendidas en territorios como Irán.

Olah frecuentemente usa la palabra "cultivar" para describir el desarrollo de los modelos de inteligencia artificial, dejando de lado el determinismo que había en los sistemas de programación anteriores.

"No programamos ni hacemos estos modelos de IA, los cultivamos. Nosotros creamos el soporte en el que se desarrollan y la luz hacia la que crecen: lo que creamos es esta entidad casi biológica u organismo que estudiamos", contaba hace un año en el pódcast de Lex Fridman.

El jefe de interpretabilidad de Anthropic sostiene que esta capacidad de generar procesos propios es la que hace que se deba revisar bajo qué patrones funcionan modelos como Claude.