

OpenAI anuncia la creación de un comité de seguridad y establece un nuevo modelo de IA para reemplazar al GPT-4

La startup con sede en San Francisco, OpenAI, anunció en un comunicado de blog el martes que ha comenzado a formar un comité de seguridad y seguridad que asesorará al consejo de administración en "decisiones críticas de seguridad y seguridad" para sus proyectos y operaciones.

El comité de seguridad llega en medio de un debate en curso sobre la seguridad de la IA en la empresa, que fue lanzada al centro de atención después de que un investigador, Jan Leike, renunciara y criticara a OpenAI por permitir que la seguridad "ocupara un asiento trasero en productos relucientes". El cofundador de OpenAI y científico jefe Ilya Sutskever también renunció, y la empresa disolvió el equipo "superalineación" centrado en los riesgos de la IA que lideraron conjuntamente.

OpenAI dijo que "ha comenzado recientemente a entrenar su próximo modelo de frontera" y que sus modelos de IA lideran la industria en capacidad y seguridad, aunque no hizo mención de la controversia. "Damos la bienvenida a un debate robusto en este momento importante", dijo la empresa.

Los modelos de IA son sistemas de predicción que se entrenan en vastos conjuntos de datos para generar texto, imágenes, video y conversaciones humanas a pedido. Los modelos de frontera son los sistemas de IA más poderosos y de vanguardia.

El comité de seguridad está compuesto por insiders de la empresa, incluidos el CEO de OpenAI, Sam Altman, y su presidente, Bret Taylor, y cuatro expertos técnicos y políticos de OpenAI. También incluye a los miembros de la junta directiva Adam D'Angelo, que es el CEO de Quora, y Nicole Seligman, ex asesora general de Sony.

La primera tarea del comité será evaluar y desarrollar aún más los procesos y medidas de seguridad de OpenAI y presentar sus recomendaciones al consejo de administración en 90 días. La empresa dijo que luego publicará las recomendaciones que está adoptando "de una manera consistente con la seguridad y la seguridad".

Informações do documento:

Autor: symphonyinn.com

Assunto: slotsstar

Palavras-chave: **slotsstar - symphonyinn.com**

Data de lançamento de: 2024-09-09