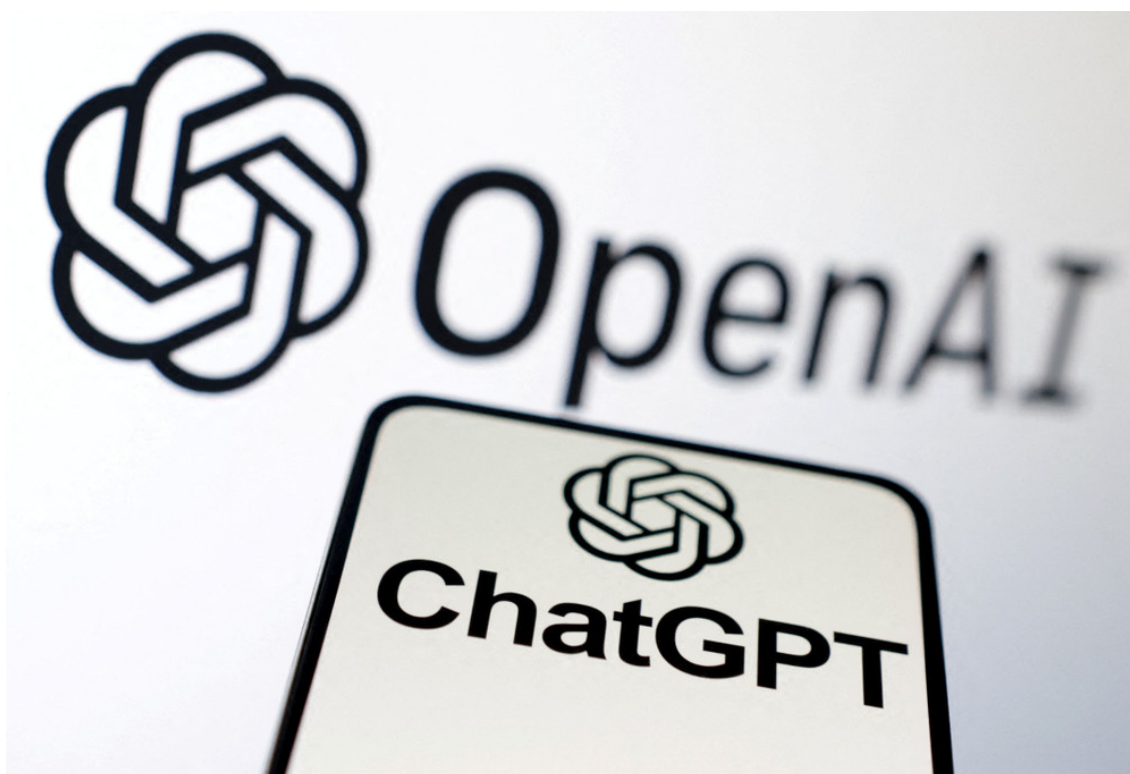


OpenAI ra mắt tính năng kiểm soát dành cho phụ huynh, bảo vệ thanh thiếu niên trên ChatGPT

13:05 02/10/2025

Trước những lo ngại về tác động tiêu cực của AI đối với người trẻ, OpenAI vừa bổ sung một loạt tính năng kiểm soát mới cho ChatGPT. Những tính năng này giúp phụ huynh quản lý cách con mình sử dụng chatbot AI một cách an toàn và phù hợp với độ tuổi hơn, đồng thời cảnh báo khi phát hiện dấu hiệu bất thường.



OpenAI ra mắt tính năng kiểm soát dành cho phụ huynh, bảo vệ thanh thiếu niên trên ChatGPT

OpenAI đã thiết kế một hệ thống kiểm soát toàn diện, cho phép phụ huynh và thanh thiếu niên liên kết tài khoản để cùng quản lý. Sau khi tài khoản được liên kết, tài khoản của người dùng trẻ tuổi sẽ tự động được áp dụng các bộ lọc nội dung chặt chẽ.

Các bộ lọc này sẽ chặn những nội dung độc hại như:

Nội dung đồ họa, bạo lực: Chặn những hình ảnh, thông tin có tính chất bạo lực hoặc gây sốc.

Thử thách lan truyền, vai trò tình cảm: Ngăn chặn những thử thách nguy hiểm trên mạng và các tương tác không phù hợp.

Tiêu chuẩn sắc đẹp cực đoan: Bảo vệ người trẻ khỏi những nội dung có thể gây ảnh hưởng xấu đến tâm lý.

Phụ huynh có thể tùy chỉnh các bộ lọc này, nhưng người dùng trẻ tuổi không thể tự tắt chúng. Ngoài ra, phụ huynh còn có thể thiết lập các chế độ quản lý như "giờ yên tĩnh" để ngăn ChatGPT hoạt động vào ban đêm, vô hiệu hóa chế độ giọng nói và ngăn việc sử dụng dữ liệu trò chuyện để huấn luyện AI.

Một trong những tính năng đáng chú ý nhất là hệ thống cảnh báo của OpenAI. Nếu phát hiện người dùng trẻ có dấu hiệu căng thẳng nghiêm trọng hoặc ý định tự làm hại bản thân, một nhóm chuyên gia sẽ đánh giá tình huống và gửi cảnh báo đến phụ huynh qua email hoặc tin nhắn.

OpenAI cam kết sẽ chỉ chia sẻ những thông tin cần thiết để phụ huynh có thể can thiệp kịp thời. Dù thừa nhận hệ thống không hoàn hảo, nhưng OpenAI tin rằng việc ưu tiên hành động sẽ tốt hơn là im lặng trước những rủi ro tiềm ẩn.

Tâm An (t/h)