

Anthropic Selidiki Akses Ilegal ke Model AI Mythos

Category: TEKNOLOGI

written by Redaksi | April 22, 2026



Perusahaan kecerdasan buatan (AI) terkemuka, [Anthropic](#), tengah melakukan investigasi intensif terkait laporan adanya akses tidak sah terhadap model AI terbaru mereka, Mythos. Model ini sebelumnya dirancang sebagai alat khusus riset keamanan siber dan aksesnya dibatasi hanya untuk perusahaan teknologi tertentu melalui program Project Glasswing.

Berdasarkan laporan Bloomberg, sebuah kelompok pengguna mengklaim telah mendapatkan akses ke model Mythos sejak hari pertama peluncurannya. Kelompok tersebut diduga memanfaatkan celah melalui lingkungan vendor pihak ketiga yang bekerja sama dengan Anthropic.

Kemampuan Model Mythos yang Sensitif

Mythos bukan sembarang model AI. [Anthropic](#) menyebutkan bahwa alat ini memiliki kemampuan untuk mengeksploitasi kerentanan

keamanan pada sistem operasi maupun peramban web utama.

Pada saat peluncuran, Anthropic mengungkapkan bahwa model ini mampu menemukan ribuan kerentanan tingkat tinggi (high-severity) pada perangkat lunak yang digunakan sehari-hari. Sebagai contoh, Mozilla baru-baru ini melaporkan penemuan 271 kerentanan pada Firefox melalui penggunaan model Mythos. Karena potensi risikonya yang tinggi, Anthropic secara ketat membatasi akses Mythos hanya kepada 11 perusahaan teknologi terpilih agar masalah keamanan dapat diperbaiki sebelum disalahgunakan oleh pihak lain.

Tanggapan Anthropic dan Celah Keamanan

Menanggapi laporan tersebut, juru bicara Anthropic mengonfirmasi bahwa perusahaan sedang melakukan penyelidikan mendalam.

“Kami sedang menyelidiki laporan yang mengklaim adanya akses tidak sah ke Claude Mythos Preview melalui salah satu lingkungan vendor pihak ketiga kami,” ujar perwakilan Anthropic.

Hingga saat ini, pihak Anthropic menyatakan belum menemukan bukti bahwa akses tersebut meluas di luar alat milik vendor yang bersangkutan. Kelompok yang mengaku mengakses model tersebut menyatakan kepada Bloomberg bahwa mereka tidak memiliki niat untuk menyebabkan kerusakan atau penyalahgunaan. Namun, insiden ini menimbulkan kekhawatiran serius mengenai keamanan model AI tingkat lanjut yang sangat kuat.

Pentingnya Keamanan AI

Insiden ini menjadi pengingat kritis bagi ekosistem AI mengenai pentingnya kontrol akses, terutama untuk model-model yang memiliki kapabilitas offensive atau serangan siber. Meskipun kelompok yang mengklaim akses tersebut mengaku memiliki niat baik, potensi penyalahgunaan oleh pihak yang

tidak bertanggung jawab tetap menjadi ancaman nyata yang harus diantisipasi oleh para pengembang AI.

Keamanan rantai pasok (supply chain security) melalui vendor pihak ketiga kini menjadi sorotan utama bagi perusahaan teknologi yang mengembangkan teknologi AI sensitif. Kasus ini kemungkinan besar akan memicu pengetatan protokol keamanan di seluruh industri untuk memastikan bahwa model-model berisiko tinggi tetap terlindungi dari akses yang tidak semestinya.[]