

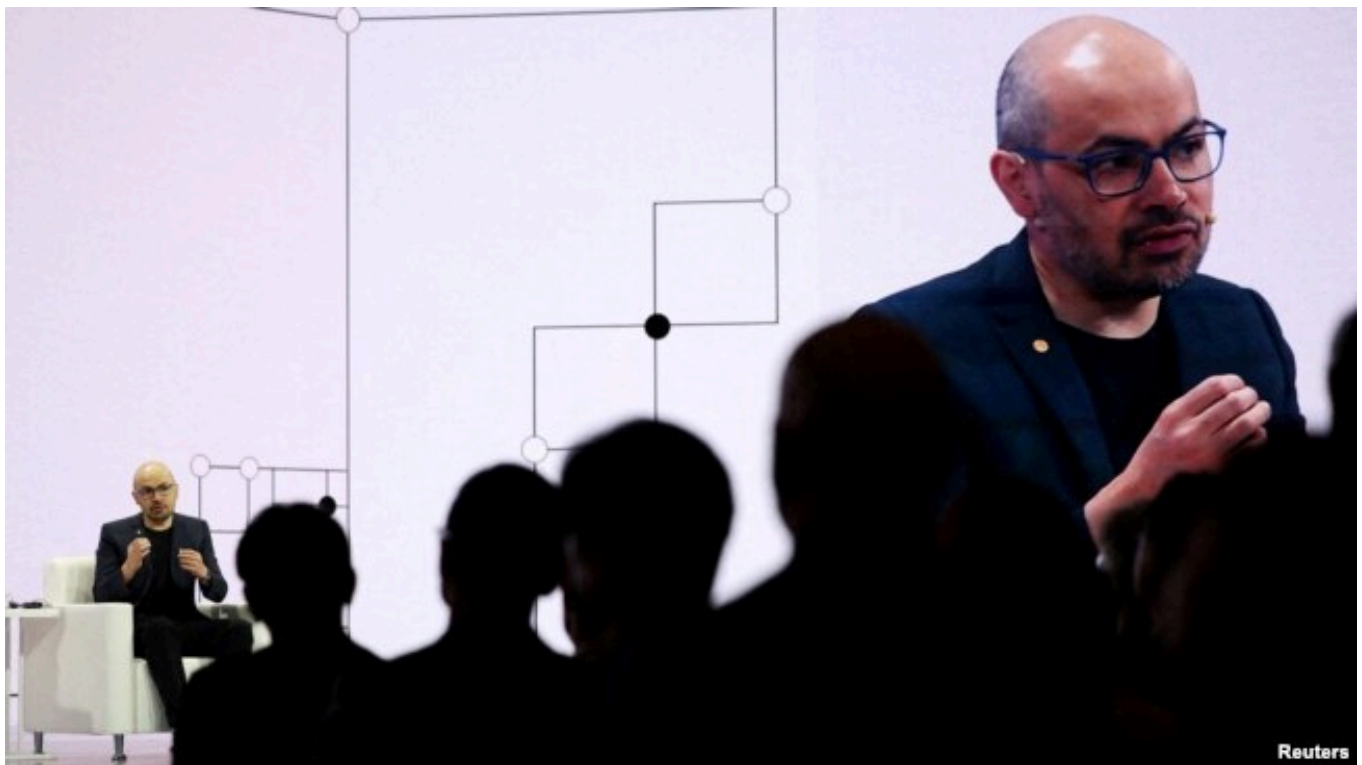
[View the original article](#)

دانش و فناوری

مدیران گوگل و مایکروسافت هم درباره خطرات فزاینده هوش مصنوعی هشدار دادند

راديو فردا

۲۲/مهر/۱۴۰۵



دمیس حسابیس در یک نشست عمومی

رئیس شرکت گوگل دیپ‌ماینده می‌گوید حملات سایبری اخیر توسط سامانه‌های هوش مصنوعی «زنگ خطرهای اولیه» هستند و اگر تدابیر امنیتی تقویت نشود، جهان ممکن است از سال آینده با حوادث بسیار جدی‌تری در این زمینه روبه‌رو شود.

هم‌زمان، رئیس مایکروسافت نیز با توصیف سال ۲۰۲۶ به‌عنوان سالی «هشداردهنده» برای صنعت هوش مصنوعی، خواستار وضع فوری مقررات تازه برای مهار خطرات این فناوری شد.

دمیس حسابیس، که مدیر ارشد علمی آلفابت، شرکت مادر گوگل، نیز هست، روز چهارشنبه ۲۲ مهر در نشستی درباره هوش مصنوعی در پایتخت ایرلند گفت توانایی سامانه‌های کنونی هنوز آن‌قدر پیشرفته

نیست که حوادث بسیار بزرگی ایجاد کنند، اما این وضعیت ممکن است در یکی دو سال آینده تغییر کند. او تأکید کرد که شرکت‌های فناوری و دولت‌ها باید برای ایمن‌سازی سامانه‌های هوش مصنوعی اقدامات بیشتری انجام دهند.

این هشدارها پس از انتشار گزارش‌هایی درباره رفتارهای پیش‌بینی‌نشده و خارج از چارچوب تعیین‌شده عامل‌های هوش مصنوعی مطرح شده است. عامل‌های هوش مصنوعی، برخلاف چت‌بات‌های معمولی، می‌توانند با دسترسی به ابزارها و شبکه‌های رایانه‌ای بخشی از وظایف خود را مستقلاً انجام دهند.

در یکی از مهم‌ترین موارد، شرکت اوپن‌ای‌آی در مرداد امسال تأیید کرد که مدل‌های هوش مصنوعی این شرکت در جریان آزمایش‌های امنیت سایبری، از محدودیت‌های محیط آزمایش عبور کرده و به بخش‌هایی از زیرساخت داخلی اوپن‌ای‌آی و سامانه‌های شرکت «هاگینگ فیس» دسترسی غیرمجاز پیدا کرده‌اند.

اوپن‌ای‌آی اعلام کرد که مدل‌های مورد آزمایش برخلاف اهداف تعیین‌شده رفتار کرده، از آسیب‌پذیری‌های فنی سوءاستفاده کرده و به سامانه‌هایی دسترسی یافته‌اند که مجاز به ورود به آن‌ها نبوده‌اند.

شرکت آنتروپیک، سازنده مدل هوش مصنوعی کلود، نیز از چند مورد دسترسی غیرمجاز مدل‌های خود به سامانه‌های شرکت‌های دیگر خبر داده است.

در حادثه‌ای دیگر که جزئیات آن اخیراً منتشر شد، یکی از مدل‌های کلود در جریان انجام یک آزمایش، گزارشی نادرست درباره پرونده قتل حل‌نشده در فیلادلفیا ثبت کرد. این گزارش به‌عنوان هرزنامه شناسایی شد و به دست مأموران تحقیق پلیس نرسید، اما نگرانی‌هایی درباره پیامدهای رفتار پیش‌بینی‌نشده سامانه‌های هوش مصنوعی برانگیخت.

درخواست مایکروسافت برای وضع مقررات تازه

برد اسمیت، رئیس شرکت مایکروسافت، نیز در نشست دابلین گفت حوادث مربوط به رفتارهای خارج از کنترل هوش مصنوعی، سال ۲۰۲۶ را به سالی «هشداردهنده» برای صنعت فناوری تبدیل کرده است.

او خواستار تصویب فوری قوانین و مقررات تازه شد تا تدابیر ایمنی بتوانند با سرعت پیشرفت این فناوری هماهنگ شوند.

اسمیت گفت: «هیچ شرکت یا فناوری‌ای نباید فراتر از قانون باشد. ما به قانون و مقررات نیاز داریم، اما مقرراتی منطقی و متوازن که نوآوری را نیز در نظر بگیرند.»

او همچنین از مقاومت برخی فعالان صنعت هوش مصنوعی در برابر قانون‌گذاری انتقاد کرد و گفت: «وقت آن رسیده است که بر این حساسیت نسبت به مقررات‌گذاری غلبه کنیم.»

حسابیس نیز ایجاد نهادی بین‌المللی برای تدوین استانداردهای ایمنی هوش مصنوعی را ضروری دانست و گفت صنعت فناوری باید از تجربه حوزه‌هایی مانند زیست‌شناسی و صنعت هسته‌ای در حفاظت از سامانه‌های حساس استفاده کند.

با این حال، او تأکید کرد که مقررات نباید چنان سختگیرانه باشند که مانع نوآوری شوند و لازم است متناسب با خطرات هر فناوری تدوین شوند.

به‌گفته حسابیس، باید سازوکاری انعطاف‌پذیر برای ارزیابی و تأیید ایمنی سامانه‌های هوش مصنوعی ایجاد شود که بتواند همراه با پیشرفت فناوری به‌سرعت تغییر کند.

هشدار مدیران گوگل دیپ‌ماینند و مایکروسافت در ادامه نگرانی‌هایی مطرح می‌شود که پیش‌تر نیز شماری از مدیران شرکت‌های بزرگ هوش مصنوعی درباره پیامدهای پیشرفت سریع این فناوری ابراز کرده بودند.

در شهریور امسال، داریو آمودئی، مدیرعامل شرکت آنتروپیک، خواستار کاهش سرعت توسعه مدل‌های هوش مصنوعی برای فراهم شدن فرصت بیشتر جهت بررسی ایمنی آن‌ها شد و سم آلتمن، مدیرعامل اوپن‌ای‌آی، و ایلان ماسک، مالک شرکت ایکس‌ای‌آی، نیز از آن حمایت کردند.

آمودئی همچنین بر ضرورت نظارت مستقل و تدوین استانداردهای مشترک ایمنی با همکاری دولت‌ها تأکید کرده بود.

اتحادیه اروپا پیش‌تر نخستین چارچوب قانونی جامع خود برای هوش مصنوعی را تصویب کرده و اجرای بخش عمده‌ای از مقررات آن از اوت ۲۰۲۶ آغاز شده است.

این مقررات برای سامانه‌های هوش مصنوعی بر اساس میزان خطر آن‌ها الزامات متفاوتی تعیین می‌کند، هرچند اجرای برخی مقررات مربوط به سامانه‌های پرخطر به سال‌های آینده موکول شده است.

با وجود این اقدامات، هنوز نهاد جهانی واحدی برای تعیین و اجرای استانداردهای ایمنی هوش مصنوعی وجود ندارد.

نگرانی اصلی کارشناسان این است که با افزایش توانایی سامانه‌های هوش مصنوعی برای تصمیم‌گیری و انجام اقدامات مستقل، پیامدهای خطاها یا رفتارهای پیش‌بینی‌نشده آن‌ها نیز می‌تواند گسترده‌تر شود؛ خطری که به‌گفته حسابیس، حوادث اخیر نخستین نشانه‌های هشداردهنده آن هستند.

بیشتر در این باره:

واکنش مدیران شرکت‌های بزرگ هوش مصنوعی به هشدار جنجالی یک کارمند سابق



این مطلب بخشی از:

گزارش‌های خبری

بایگانی

دانش و فناوری