چهار شنبه 25 بهمن 1402 - 07:15

خیال خام کنترل ایمن هوش مصنوعی

دکتر رومن وی یامپولسکی از کارشناسان برجسته هوش مصنوعی می‌گوید ما با یک رویداد تقریباً تضمین شده با پتانسیل ایجاد یک فاجعه وجودی روبرو هستیم.

دکتر رومن وی یامپولسکی از کارشناسان برجسته هوش مصنوعی می گوید ما با یک رویداد تقریباً تضمین شده با پتانسیل ایجاد یک فاجعه وجودی روبرو هستیم.

به گزارش ایسنا، یک متخصص در زمینه هوش مصنوعی، بررسی کاملی روی ادبیات علمی در مورد هوش مصنوعی انجام داده و به این نتیجه رسیده است که هیچ مدرکی برای حمایت از این تصور وجود ندارد که هوش مصنوعی را می توان به طور ایمن کنترل کرد.

علاوه بر این، حتی اگر مکانیسم ها یا روش هایی برای کنترل هوش مصنوعی پیشنهاد شده باشد، دکتر رومن وی یامپولسکی(Roman V. Yampolskiy) معتقد است که این اقدامات برای تضمین ایمنی کافی نخواهند بود.

به نقل از آی ای، دکتر یامپولسکی دانشیار دانشگاه لوئیزویل(Louisville) است که نتایج تحقیقات خود را در کتابی با عنوان «هوش مصنوعی غیر قابل توضیح، غیر قابل پیش بینی، غیر قابل کنترل» منتشر کرده است.

کار او در زمینه «ایمنی هوش مصنوعی» تا حدی توسط ایلان ماسک و موسسه آینده زندگی(FLI) که توسط مکس تِگمارک استاد دانشگاه MIT و ژان تالین یکی از بنیانگذاران اسکایپ تاسیس شده است، پشتیبانی شده است.

موسسه آینده زندگی همان موسسه ای است که برای توسعه ابزارهای هوش مصنوعی قدرتمندتر از GPT-4 درخواست کرد این پروژه با یک توقف شش ماهه روبرو شود.

نامه سرگشاده این موسسه توسط بیش از 33 هزار نفر از جمله ایلان ماسک، استیو وزنیاک هم بنیانگذار اپل، یوشوا بنجیو(Yoshua Bengio) پدرخوانده هوش مصنوعی و سایر کارشناسان امضا شده است.

نگرانی در مورد مدیریت و تنظیم مقررات در مقابل خطرات هوش مصنوعی

با نفوذ ابزارهای جدید هوش مصنوعی در بازار طی یک سال گذشته، دکتر یامپولسکی پیشنهاد می کند که سیستم های هوش مصنوعی پیشرفته به دلیل غیر قابل پیش بینی بودن و استقلال ذاتی، با وجود مزایای بالقوه شان، همیشه خطراتی را به همراه خواهند داشت.

دکتر یامپولسکی در یک بیانیه مطبوعاتی گفت: چرا بسیاری از پژوهشگران تصور می کنند که مشکل کنترل هوش مصنوعی قابل حل است؟ تا آنجا که ما می دانیم، هیچ گواهی برای آن وجود ندارد، هیچ مدرکی وجود ندارد. قبل از شروع تلاش برای ساخت یک هوش مصنوعی کنترل شده، مهم است که نشان دهیم مشکل قابل حل است.

وی افزود: این همراه با آماری که نشان می دهد توسعه آبرهوش مصنوعی یک رویداد تقریباً تضمین شده است، نشان می دهد که ما باید از تلاش های ایمنی قابل توجهی در زمینه هوش مصنوعی حمایت کنیم.

توسعه ابرهوش مصنوعی اجتناب ناپذیر است

از آنجایی که هوش مصنوعی در حالت ابرهوشمندی می تواند به تنهایی یاد بگیرد، تطبیق یابد و به صورت نیمه مستقل عمل کند، اطمینان از ایمن بودن آن، به ویژه با وجود افزایش قابلیت های آن به طور فزاینده ای چالش برانگیز می شود.

می توان گفت که هوش مصنوعی فوق هوشمند «ذهن» خود را خواهد داشت. پس چگونه آن را کنترل کنیم؟ آیا قوانین رباتیک آیزاک آسیموف که رفتار اخلاقی بین انسان و ربات را برقرار می کند، هنوز در دنیای امروز اعمال می شود؟

دکتر یامپولسکی که حوزه اصلی مورد علاقه او ایمنی هوش مصنوعی است، توضیح داد: ما با یک رویداد تقریبا تضمین شده با پتانسیل ایجاد یک فاجعه وجودی روبرو هستیم.

وی افزود: جای تعجب نیست که بسیاری این را مهم ترین مشکلی که بشریت تا به حال با آن روبرو بوده است، توصیف می کنند. نتیجه می تواند رفاه یا انقراض بشر باشد.

یکی از چالش های اصلی، تصمیم گیری های بالقوه و شکست موجودات فوق هوشمند است که پیش بینی و تهدید آنها را دشوار می کند.

علاوه بر این، فقدان قابلیت توضیح در تصمیمات هوش مصنوعی نگرانی هایی را ایجاد می کند، زیرا درک این تصمیمات برای کاهش تصادفات و تضمین نتایج بدون سوگیری، به ویژه در زمینه های حیاتی مانند مراقبت های بهداشتی و مالی که هوش مصنوعی در آن نفوذ کرده است، بسیار مهم است.

با افزایش استقلال هوش مصنوعی، کنترل انسان روی آن کاهش می یابد که منجر به نگرانی های ایمنی می شود. دکتر یامپولسکی ایجاد تعادل بین توانایی و کنترل هوش مصنوعی را پیشنهاد و اذعان می کند که ابرهوش ذاتاً فاقد قابلیت کنترل است.

وی همچنین درباره اهمیت همسویی هوش مصنوعی با ارزش های انسانی بحث می کند و روش هایی را برای به حداقل رساندن خطرها مانند اصلاح پذیری، شفافیت هوش مصنوعی و طبقه بندی به عنوان قابل کنترل یا غیرقابل کنترل پیشنهاد می

کند.
با وجود این چالش ها، دکتر یامپولسکی ادامه تحقیقات و سرمایه گذاری در ایمنی و امنیت هوش مصنوعی را تشویق و تأکید می کند که اگرچه دستیابی به هوش مصنوعی صد درصد ایمن ممکن است دست نیافتنی باشد، اما تلاش برای بهبود ایمنی آن همواره ارزشمند است.