

Large Language Models for Robotics: Opportunities, Challenges, and Perspectives

مقاله حاضر یک روش جدید برای استفاده از مدل‌های زبانی بزرگ در رباتیک ارائه می‌کند و هدف آن ساخت سیستمی است که ربات بتواند تنها با اتکا به زبان طبیعی و بدون نیاز به برنامه‌نویسی سنگین، وظایف پیچیده و متغیر را در محیط واقعی انجام دهد. نویسندهای مقاله باور دارند که مدل‌های زبانی بزرگ به دلیل توانایی درک جمله‌ها، داشتن دانش گسترده و قابلیت استدلال مرحله به مرحله، می‌توانند نقش لایه هوشی ربات را بر عهده بگیرند. این لایه هوشی بین فرمان کاربر، داده‌های حسگری و توانایی‌های حرکتی رابط برقرار می‌کند و به ربات اجازه می‌دهد وظایفی را انجام دهد که قبلًا نیاز به مدل‌سازی دقیق یا داده‌آموزی گسترده داشت.

مقاله کار خود را با بررسی مشکل اصلی شروع می‌کند. پیش از این، ربات‌ها برای درک فرمان‌های انسانی و تبدیل آن‌ها به اقدامات مناسب به زیرسیستم‌های مختلف وابسته بودند. به عنوان مثال، اگر کاربر دستور می‌داد که «جعبه کوچک را از میز بردار و کنار دیوار بگذار»، ربات باید ابتدا اجسام را تشخیص می‌داد، سپس موقعیت‌ها را تحلیل می‌کرد، بعد برنامه‌ای برای حرکت مناسب می‌ساخت و در نهایت حرکات دقیق را اجرا می‌کرد. این مراحل نیازمند مدل‌های جداگانه، الگوریتم‌های پیچیده و برنامه‌نویسی‌های خاص بودند. مدل‌های زبانی بزرگ با اتصال این مراحل و کاهش وابستگی به کدهای دستی، این فرآیند را ساده کرده‌اند. نویسندهای مقاله تلاش کرده‌اند نشان دهنده که چگونه یک مدل زبانی می‌تواند این مراحل را یکپارچه کند و ربات را قادر سازد تا تنها با استفاده از یک توضیح متنی، وظیفه را از ابتداء تا انتها درک و اجرا کند.

در گام بعدی مقاله توضیح می‌دهد که مدل‌های زبانی چگونه می‌توانند دستورهای انسان را تفسیر و ساختار آن‌ها را استخراج کنند. این مدل‌ها با استفاده از دانش وسیعی که از میلیاردها جمله و متن فرا گرفته‌اند، قادر هستند نیت کاربر را حدس بزنند، ابهام‌های موجود در دستور را برطرف کنند و معنای دقیق‌تری از جمله ارائه دهنند. برای مثال اگر کاربر بگوید «کتاب را از آنجا بردار»، مدل زبانی می‌تواند با توجه به متن و اطلاعات محیطی که از ربات دریافت می‌کند، «آن‌جا» را مشخص کند یا حتی اگر لازم باشد سوال تکمیلی بپرسد. این توانایی تفسیر انعطاف‌پذیر یکی از علت‌هایی است که باعث شده مدل‌های زبانی برای رباتیک بسیار مناسب باشند.

یکی دیگر از بخش‌های مهم مقاله، نقش مدل زبانی در استدلال و برنامه‌ریزی است. ربات برای انجام یک وظیفه باید بتواند آن را به مراحل کوچکتر تبدیل کند. این کار در گذشته با استفاده از الگوریتم‌های ثابت انجام می‌شد، اما مدل‌های زبانی اکنون قادرند با استفاده از استدلال زبانی، خودشان مراحل را طراحی کنند. اگر ربات به این صورت فرمان بگیرد که «آشپزخانه را مرتب کن»، مدل زبانی می‌تواند نتیجه بگیرد که ابتدا باید ظرف‌ها را جمع کند، سپس وسایل اضافه را از روی پیشخوان بردارد، بعد زیاله‌ها را جدا کند و در نهایت کف را پاکسازی کند. نکته مهم این است که این مدل‌ها مراحل را نه با استفاده از قوانین ثابت، بلکه با استفاده از دانش عمومی و استدلال منطقی طراحی می‌کنند. این فرآیند باعث می‌شود ربات‌ها در برخورد با موقعیت‌های جدید عملکرد بهتری داشته باشند.

در بخشی از مقاله نویسندهای توضیح می‌دهند که مدل زبانی چگونه با سیستم کنترل ربات ارتباط برقرار می‌کند. اگرچه مدل زبانی نمی‌تواند سیگنال‌های دقیق حرکتی تولید کند، اما می‌تواند وظایف را به شکل قابل فهم برای کنترلر پایین‌دست ارائه دهد. به عبارت

دیگر مدل زبانی می‌تواند بگوید چه کار باید انجام شود و سیستم کنترلی تشخیص می‌دهد چگونه باید آن کار انجام شود. این ساختار از بروز خطاها خطرناک جلوگیری می‌کند، زیرا وظایف پیچیده به اقدامات ساده‌تر و منطقی تقسیم شده‌اند. مقاله اشاره می‌کند که در بسیاری از آزمایش‌ها، ربات‌ها توانسته‌اند با این روش کارهایی مانند جایه‌جایی اجسام، باز کردن درها، مرتب‌سازی وسایل و تعامل با محیط خانه را بدون برنامه‌نویسی سنگین انجام دهند.

بخش مهم دیگر مقاله مربوط به استفاده از دانش گسترده مدل است. مدل‌های زبانی بزرگ فقط اطلاعات مرحله‌ای ارائه نمی‌دهند، بلکه می‌توانند از دانش عمومی خود نیز کمک بگیرند. وقتی ربات با یک شرایط ناشناخته مواجه شود، مدل زبانی می‌تواند با انجام استدلال زبانی به یک راه حل مناسب برسد. نویسنده‌گان توضیح می‌دهند که این مدل‌ها وقتی با داده‌های حسگری ربات ترکیب می‌شوند، می‌توانند تفسیر دقیق‌تری از وضعیت ارائه دهند. گاهی مدل زبانی از طریق استناد، متن‌ها یا دانش گذشته خود توضیحاتی برای تصمیم‌ها می‌دهد که برای مهندسین رباتیک بسیار مفید است. این توضیحات باعث می‌شود رفتار ربات قابل پیش‌بینی‌تر و قابل اعتماد‌تر باشد.

مقاله سپس به چالش‌های موجود در استفاده از مدل‌های زبانی در رباتیک می‌پردازد. یکی از بزرگ‌ترین چالش‌ها مسئله توهم یا اشتباه‌های منطقی است. مدل‌ها گاهی جمله‌هایی تولید می‌کنند که ظاهراً منطقی هستند اما در واقعیت درست نیستند. در رباتیک یک اشتباه کوچک می‌تواند به آسیب فیزیکی، خرابی تجهیزات یا رفتار ناخواسته منجر شود. مسئله دیگر این است که مدل‌های زبانی دریافت عمیقی از قوانین حقیقی فیزیک ندارند. آن‌ها ممکن است درباره مسیر حرکت، وزن یک جسم یا خطرات احتمالی اشتباه کنند. مقاله تاکید می‌کند که استفاده از مدل زبانی باید با سیستم‌های کنترل دقیق و نظارت مستمر همراه باشد تا از رفتار غلط جلوگیری شود.

یکی دیگر از چالش‌های مهم، وابستگی مدل‌ها به سخت‌افزار قوی است. ربات‌ها معمولاً محدودیت توان پردازشی دارند و نمی‌توانند مدل‌های بزرگ را مستقیماً اجرا کنند. نویسنده‌گان چند راه حل برای این موضوع ارائه می‌کنند، مانند استفاده از مدل‌های فشرده‌تر، انجام پردازش‌های سنگین در یک سرور خارجی، یا استفاده از ساختارهای ترکیبی که بخش‌هایی از پردازش را بین ربات و یک سیستم ابری تقسیم می‌کند. همچنین به این نکته اشاره می‌شود که استفاده از روش‌های ارزیابی لحظه‌ای و بازخورد در حلقه می‌تواند دقیق مدل‌ها را در شرایط واقعی بالا ببرد.

در پایان مقاله، نویسنده‌گان آینده این حوزه را تحلیل می‌کنند. به اعتقاد آن‌ها مدل‌های زبانی بزرگ فقط یک ابزار نیستند بلکه مسیر تازه‌ای برای طراحی ربات‌های نسل جدید به وجود آورده‌اند. ربات‌ها در آینده نه تنها قادر خواهند بود فرمان‌ها را بفهمند بلکه می‌توانند تصمیم‌های پیچیده و چندمرحله‌ای بگیرند، شرایط جدید را تحلیل کنند و بخشی از خطاها احتمالی را خودشان برطرف کنند. روند آینده به سمت ساخت مدل‌های کوچک‌تر اما دقیق‌تر، اتصال بهتر بین زبان و داده‌های حسگری، افزایش ایمنی سیستم‌ها و کاهش وابستگی به داده‌های آموزشی گسترده پیش می‌رود. مقاله نتیجه می‌گیرد که این فناوری می‌تواند فصل جدیدی در تعامل انسان و ربات ایجاد کند؛ فصلی که در آن ربات نه فقط یک ابزار قابل برنامه‌ریزی، بلکه یک همکار هوشمند است که می‌تواند زبان انسان را بفهمد و به شکل منطقی و انعطاف‌پذیر پاسخ دهد.