

Large Language Models for Robotics: A Survey

مدل‌های زبانی بزرگ یا همان LLM ها در چند سال اخیر به یکی از مهم‌ترین ابزارهای هوش مصنوعی تبدیل شده‌اند و تأثیر آن‌ها فراتر از پردازش متن رفته و وارد رباتیک شده است. ربات‌ها برای انجام وظایف خود به مجموعه‌ای از توانایی‌ها مانند درک محیط، تحلیل وضعیت، برنامه‌ریزی مرحله‌به‌مرحله و در نهایت اجرای دقیق حرکات نیاز دارند. تا قبل از ظهور LLM ها، بیشتر این مراحل با الگوریتم‌های پیچیده، برنامه‌نویسی‌های سنگین و مدل‌های خاص انجام می‌شد. اما مدل‌های زبانی بزرگ توانسته‌اند بخش زیادی از این پیچیدگی را کاهش دهند. دلیل اصلی این پیشرفت، توانایی این مدل‌ها در فهم زبان طبیعی، تولید متن دقیق و منطقی، استدلال درباره مسائل مختلف و تبدیل توضیح انسانی به برنامه عملی است. مقاله توضیح می‌دهد که LLM ها چگونه در هر مرحله از کار ربات‌ها نقش مهمی پیدا کرده‌اند و چرا حالا تعامل انسان با ربات بسیار ساده‌تر شده است.

یکی از مهم‌ترین بخش‌های کار ربات، درک محیط است. ربات باید بداند چه چیزی در اطرافش وجود دارد، اجسام در چه موقعیتی قرار گرفته‌اند و چه رابطه‌ای با هم دارند. مدل‌های زبانی بزرگ، مخصوصاً نسخه‌های چندوجهی که می‌توانند تصویر یا ویدئو را هم تحلیل کنند، کمک می‌کنند ربات علاوه بر دیدن محیط، بتواند آن را معنی کند. به عنوان مثال اگر به ربات گفته شود «لیوان کنار سینک را بردار»، ربات باید تشخیص دهد سینک کجاست، لیوان کدام است و منظور کاربر کدام بخش از محیط است LLM. ها این توانایی را ایجاد کرده‌اند که ربات فقط به پیکسل‌ها نگاه نکند، بلکه بتواند مفهوم «کنار»، «روی»، «بین» و حتی ارتباط بین اشیاء را بفهمد. این درک مفهومی چیزی است که روش‌های سنتی دید کامپیوتری به‌سادگی در اختیار ربات نمی‌گذاشتند.

در بخش برنامه‌ریزی و استدلال، مقاله توضیح می‌دهد که LLM ها چگونه باعث تحول اساسی شده‌اند. ربات‌ها برای انجام هر کار باید بدانند که مرحله اول چیست، بعد چه کاری لازم است و اگر خطایی رخ دهد چه واکنشی باید نشان دهند. مدل‌های زبانی بزرگ می‌توانند فقط با دریافت یک فرمان کلی، آن را به مجموعه‌ای از مراحل کوچک و قابل اجرا تبدیل کنند. برای نمونه اگر گفته شود «این میز را مرتب کن»، مدل می‌تواند تصمیم بگیرد که ابتدا اشیاء را دسته‌بندی کند، سپس آن‌ها را جابه‌جا کند، و اگر مشکلی پیش آمد آن را اصلاح کند. این توانایی به دلیل روش‌هایی مانند استدلال مرحله‌ای یا Chain-of-Thought ایجاد شده است که باعث می‌شود مدل مثل انسان فکر کند LLM. ها حتی می‌توانند برای ربات کدهای اجرایی بنویسند، اسکریپت‌های ROS تولید کنند یا یک برنامه کاری جدید پیشنهاد دهند. این موضوع باعث شده ربات‌ها برای یادگیری وظایف جدید نیازی به آموزش گسترده یا داده زیاد نداشته باشند و تنها با چند مثال ساده بتوانند رفتار جدیدی یاد بگیرند.

نقش LLM ها در کنترل ربات نیز مهم است. هرچند این مدل‌ها مستقیماً موتور یا بازوی ربات را کنترل نمی‌کنند، اما می‌توانند دستورهای سطح بالا را تولید کنند و به سیستم‌های کنترل دقیق منتقل کنند. در واقع LLM مانند مغز سطح بالای ربات عمل می‌کند که تصمیم می‌گیرد چه کاری انجام شود و سیستم‌های کنترلی پایین‌دست این تصمیم را به حرکت‌های دقیق تبدیل می‌کنند. این ساختار باعث می‌شود هم دقت ربات حفظ شود و هم از توانایی زبانی و استدلالی LLM ها استفاده شود. مقاله توضیح می‌دهد که در بسیاری از ربات‌ها، LLM مسئول نوشتن یا اصلاح کدهای کنترل است و حتی می‌تواند خطاهای احتمالی را پیش‌بینی و تصحیح کند. این کار به ربات کمک می‌کند در محیط‌های واقعی بهتر سازگار شده و رفتار مطمئن‌تری نشان دهد.

مقاله همچنین نمونه‌هایی از کاربرد LLM ها در ربات‌های واقعی ارائه می‌کند. بسیاری از ربات‌های خانگی، صنعتی و انبارداری اکنون قادرند فرمان‌های پیچیده را با توضیح ساده کاربر اجرا کنند. برای مثال ربات‌هایی وجود دارند که فقط با توضیحات زبانی می‌توانند وسایل را مرتب کنند، اشیای موردنظر را پیدا کنند، دستورهای آشپزی را دنبال کنند یا وظایف صنعتی مشخصی را انجام دهند. نکته جالب این است که این ربات‌ها برای یادگیری وظایف جدید نیازی به برنامه‌نویسی دستی ندارند و بسیاری از آن‌ها می‌توانند با کمک LLM ها کدها و رفتارهای جدید تولید کنند. این موضوع نشان‌دهنده افزایش قابل توجه تطبیق‌پذیری ربات‌هاست؛ قابلیت‌هایی که برای سال‌ها یکی از چالش‌های بزرگ این حوزه بود.

با وجود این پیشرفت‌ها، مقاله به چالش‌های مهمی نیز اشاره می‌کند. یکی از بزرگ‌ترین مشکلات، پدیده «توهم» در LLM ها است. این مدل‌ها گاهی پاسخ‌های اشتباه اما ظاهراً منطقی تولید می‌کنند. این رفتار در حوزه متن شاید مشکل بزرگی ایجاد نکند، اما در رباتیک می‌تواند خطرناک باشد؛ زیرا یک دستور اشتباه ممکن است باعث برخورد ربات با انسان یا محیط شود. مشکل دیگر این است که LLM ها دنیای فیزیکی را مانند انسان نمی‌فهمند. آن‌ها از روی متن یاد گرفته‌اند و ممکن است درک درستی از وزن، نیرو، تعادل جسم یا قوانین فیزیکی نداشته باشند. ربات در جهان واقعی نیازمند دقت بالا و تصمیمات قابل اعتماد است، اما LLM ها گاهی فقط حدس می‌زنند. مسئله دیگری که مقاله مطرح می‌کند محدودیت سخت‌افزاری ربات‌هاست. اجرای مدل‌های بزرگ معمولاً به پردازشگرهای قوی نیاز دارد، در حالی که بیشتر ربات‌ها فضای محدود و توان مصرفی کم دارند.

در بخش آخر، مقاله مسیرهای آینده را بررسی می‌کند. نویسندگان معتقدند ترکیب LLM ها با مدل‌های دیداری پیشرفته، کنترل‌کننده‌های دقیق و روش‌های بازیابی اطلاعات می‌تواند نسل جدیدی از ربات‌ها را خلق کند. در این ربات‌ها، LLM نقش هسته تصمیم‌گیری را بر عهده خواهد داشت، در حالی که سیستم‌های دیگر اطلاعات واقع‌گرایانه و دقیق فراهم می‌کنند. بهبود در اتصال زبان و واقعیت فیزیکی، کاهش خطاهای منطقی، طراحی مدل‌های کوچک و تخصصی‌تر و تقویت سازوکارهای ایمنی از جمله مواردی هستند که باید در آینده مورد توجه قرار گیرند. به اعتقاد نویسندگان، ربات‌های آینده قادر خواهند بود مانند یک دستیار هوشمند و قابل اعتماد عمل کنند، فرمان‌های انسانی را به‌خوبی بفهمند و با تکیه بر استدلال و دانش خود تصمیم‌های پیچیده بگیرند.

در جمع‌بندی، مقاله توضیح می‌دهد که LLM ها باعث یک تغییر بنیادین در رباتیک شده‌اند. این مدل‌ها با فراهم کردن توانایی فهم زبان، استدلال منطقی، تولید برنامه‌های اجرایی و یادگیری سریع، ربات‌ها را به سیستم‌هایی بسیار هوشمندتر و قابل تعامل‌تر تبدیل کرده‌اند. هرچند چالش‌هایی مانند ایمنی، محدودیت سخت‌افزاری و درک فیزیکی باقی است، اما مسیر پیش رو نشان می‌دهد که ترکیب LLM با رباتیک می‌تواند آینده‌ای بسازد که در آن انسان و ربات مانند دو شریک طبیعی با هم همکاری کنند.