چکیده(Abstract)

مدل‌های زبانی بزرگ (LLMs) به یکی از عناصر تحول‌آفرین در رباتیک مدرن تبدیل شده‌اند. این مدل‌ها به ربات‌ها امکان می‌دهند که زبان طبیعی را تفسیر کنند، وظایف چندمرحله‌ای را برنامه‌ریزی کنند، بر اساس ورودی‌های چندوجهی استدلال کنند و مهارت‌های دستکاری اشیا را در محیط‌های پویا اجرا کنند. با حرکت رباتیک به سمت سطوح بالاتری از خودمختاری و هوش تجسم‌یافته، سامانه‌های مبتنی بر LLM به‌تدریج نقش هستهٔ شناختی را بر عهده می‌گیرند؛ هسته‌ای که ادراک، تصمیم‌گیری، کنترل و تعامل انسان–ربات را یکپارچه می‌کند.

با انگیزهٔ پیشرفت‌های سریع این حوزه، این مقاله یک جمع‌بندی جامع از دستاوردهای اخیر در رباتیک مبتنی بر LLM ارائه می‌دهد؛ دستاوردهایی که بر تحلیل مجموعه‌ای از مطالعات بنیادین و درک یکپارچهٔ ما از Summary.pdfاستوار است. در این مقاله بررسی می‌کنیم کهLLM ها چگونه برنامه‌ریزی وظایف مبتنی بر زبان طبیعی، ادراک چندوجهی، کنترل دستکاری، تصمیم‌گیری تعاملی و ساختارهای برنامه‌نویسی ماژولار برای ربات‌ها را تقویت می‌کنند. همچنین به روش‌هایی مانند برنامه‌ریزی مبتنی بر گراندینگ [1.Large Language Models for Robotics: Opportunities, Challenges, and Perspectives]، تحلیل نحوی مبتنی بر هستی‌شناسی [2.Parsing Natural Language Sentences into Robot Actions] و چارچوب‌های نظام‌مند آموزش ربات از طریق زبان انسان [3.A Review of Natural-Language-Instructed Robot Execution Systems] می‌پردازیم.

ما علاوه بر این، راهبردهای تنظیم تخصصی مدل‌های زبانی برای برنامه‌نویسی صنعتی ربات‌ها [4.Domain-Specific Fine-Tuning...] و روش‌های دستکاری مبتنی بر استدلال چندوجهی مانندRT-Grasp [5.RT-Grasp...] را بررسی می‌کنیم. همچنین پیشرفت‌های اخیر در سامانه‌های کنترل ربات مبتنی بر بازیابی اطلاعات [6.ARRC...] و چارچوب‌های ساخت‌یافتهٔ مبتنی برROS [7.ROS-LLM...] را مرور می‌کنیم.

با سازمان‌دهی این یافته‌ها در حوزه‌های برنامه‌ریزی، استدلال، دستکاری، ادراک، معماری‌های کنترلی و تعامل، یک نگاه یکپارچه از وضعیت کنونی هوش تجسم‌یافتهٔ مبتنی بر LLM ارائه می‌دهیم. در پایان نیز چالش‌هایی مانند **ایمنی، پایداری گراندینگ، دقت عددی و تعمیم در دنیای واقعی** را بیان کرده و مسیرهای امیدبخش آینده برای ربات‌های خودکار، همسو با انسان و مقیاس‌پذیر در محیط‌های پیچیده را بحث می‌کنیم.

واژگان کلیدی(Keywords)

مدل‌های زبانی بزرگ(LLMs) ؛ رباتیک؛ هوش تجسم‌یافته؛ درک زبان طبیعی؛ برنامه‌ریزی وظایف؛ دستکاری اشیا؛ استدلال چندوجهی؛ تولید مبتنی بر بازیابی اطلاعات؛ تعامل انسان–ربات؛ سامانه‌های کنترل ربات.

1.مقدمه(Introduction)

1.1.پیش‌زمینه و انگیزه

یکپارچگی مدل‌های زبانی بزرگ (LLMs) با رباتیک، نقطهٔ عطفی در نحوهٔ فهم، استدلال و تعامل ربات‌ها با جهان فیزیکی محسوب می‌شود. در حالی که رباتیک سنتی به برنامه‌نویسی ساخت‌یافته، سیاست‌های کنترلی تخصصی و مدل‌های دقیق مهندسی‌شده متکی بود، پیشرفت‌های اخیر در پردازش زبان طبیعی، امکان چارچوب‌هایی را فراهم کرده است که در آنها ربات می‌تواند **مقصود انسان را مستقیماً از طریق زبان طبیعی** درک کند. این تغییر نشان‌دهندهٔ گذار به‌سمت «هوش تجسم‌یافته» است؛ جایی که ربات‌ها با استفاده از بازنمایی‌های شهودی و غنی‌شده از نظر شناختی، یاد می‌گیرند، عمل می‌کنند و همکاری می‌کنند [8.Large Language Models for Robotics: A Survey].

در گذشته، سامانه‌های رباتیکی نیازمند برنامه‌نویسی صریح و دانش دامنه‌ای گسترده بودند. امالLM ها همچون موتورهای استدلالی عمومی عمل می‌کنند که قادرند معنای زمینه‌ای را استخراج کنند، برنامهٔ انجام وظایف را تولید کنند و این برنامه‌ها را به دستورهای قابل‌اجرا تبدیل کنند. این قابلیت امکان می‌دهد وظایف پیچیده – از ناوبری تا دستکاری – به صورت مکالمه‌ای تعریف شوند، نه با کدنویسی تخصصی یا الگوهای نمادین پیچیده. برای مثال، سامانه‌های برنامه‌ریزی مبتنی بر LLM قادرند حتی بدون ورودی بصری، دستورهای زبانی پیچیده را به توالی اقدامات تبدیل کنند و هنگامی که ورودی چندوجهی در اختیارشان قرار گیرد، عملکرد بسیار بهتری نشان می‌دهند [1.Large Language Models for Robotics: Opportunities, Challenges, and Perspectives].

مدل‌های چندوجهی LLM نیز با ترکیب ورودی‌های بصری، زبانی و زمینه‌ای، بازنمایی‌های «گراندشده» تولید می‌کنند که پل میان ارتباط انسانی و کنترل ربات است. این مدل‌ها توانایی استدلال پیچیده، درک قابلیت‌های اشیا و آگاهی فضایی را فراهم می‌کنند که برای عملکرد قابل‌اعتماد در محیط‌های واقعی ضروری است. چنین توانایی‌هایی نشان می‌دهد که استدلال مبتنی بر LLM

قادرند توالی‌های دقیق و منظم از اقدامات تولید کنند، از دانش عمومی بهره بگیرند و حتی با برنامه‌ریزهای کلاسیک همکاری کنند تا دقت و قابلیت اتکای برنامه‌ریزی افزایش یابد [1.Large Language Models for Robotics: Opportunities, Challenges, and Perspectives].

این هم‌افزایی به ربات اجازه می‌دهد که:

- دستورهای بلندمدت را به گام‌های ساخت‌یافته تقسیم کند،
- پیش‌شرط‌ها و روابط فضایی را تحلیل کند،
- از مکانیزم‌های گراندینگ احتمالاتی استفاده کند،
- و از دانش پیشینی برای هدایت الگوریتم‌های جست‌وجو بهره ببرد.

علاوه بر این، چارچوب‌هایی مانند **SayCan** و **LLM+P** تلفیقی از راهنمایی معنایی LLM و سیاست‌های تقویتی را ارائه می‌دهند. در این رویکرد، LLM هدف‌های سطح‌بالا را مشخص می‌کند و یادگیری تقویتی عملی‌بودن آنها را تأیید می‌کند [1...].

سامانه‌های دیگر از ساختارهایی مانند **اشاره‌های خطا**، بازخورد چندوجهی و مدل‌های **استدلال–عمل مانند ReAct** بهره می‌برند. این روش‌ها موجب ثبات بیشتر ربات در محیط‌های پویا می‌شوند.

استدلال پیچیده تنها به برنامه‌ریزی محدود نمی‌شود و حوزهٔ تصمیم‌گیری و سازگاری را نیز دربرمی‌گیرد. به‌عنوان نمونه، **LM-Nav** ترکیبی از زبان، بینایی و اطلاعات ناوبری را برای دنبال‌کردن دستورهای سطح‌بالا به‌کار می‌گیرد، بدون آنکه نیاز به داده‌های پیمایش نشانه‌گذاری‌شده باشد [1...].

در کنار آن، مدل‌های **آگاهی از عدم قطعیت** مانند KnowNo به ربات کمک می‌کنند از تصمیمات اشتباه یا خطرناک اجتناب کند و قابلیت اطمینان خود را در وظایف طولانی افزایش دهد [1...].

این مجموعه از روش‌ها بیانگر آن است که LLM ها از یک «پردازشگر زبان» فراتر رفته و به «عامل‌های استدلالی فعال» تبدیل شده‌اند که در هستهٔ معماری کنترل ربات قرار می‌گیرند.

1.4.دستکاری، ادراک و مهارت‌های تجسم‌یافته تقویت‌شده باLLM ها

در حالی‌که برنامه‌ریزی وظایف مبتنی بر زبان یکی از برجسته‌ترین پیشرفت‌های اخیر است، **دستکاری فیزیکی و کنترل تجسم‌یافته** همچنان از چالش‌برانگیزترین بخش‌های رباتیک محسوب می‌شود. روش‌های سنتی برای عنصر مرکزی در حرکت به‌سوی خودمختاری، سازگاری و هم‌راستایی بیشتر با انسان است.

1.2.از زبان تا عمل: نقش درک زبان طبیعی

یکی از کارکردهای اساسی مدل‌های زبانی بزرگ در رباتیک، **تبدیل زبان انسان به بازنمایی‌های نمادین، ساخت‌یافته یا قابل‌اجرا** است. پژوهش‌های اولیه تمرکز خود را بر خط لوله‌های تحلیل زبان طبیعی قرار داده بودند؛ برای مثال، استفاده از سامانه‌های مبتنی بر هستی‌شناسی که با تحلیل وابستگی‌های نحوی، **افعال، اعضای بدن ربات و محدودیت‌ها** را استخراج می‌کردند [2.Parsing Natural Language Sentences into Robot Actions].

این سامانه‌ها نشان دادند که ربات می‌تواند به شکل هوشمندانه به دستورهای انسانی واکنش نشان دهد و در عین حال، سازگاری اقدامات پیشنهادی را با وضعیت فعلی خود بررسی کند؛ برای نمونه، رد کردن دستوری که تعادل ربات را مختل می‌کند.

نسل جدیدLLM ها این قابلیت را به‌صورت چشم‌گیری گسترش داده است. مدل‌های مدرن با درک معنایی عمیق، **نیت ضمنی، وابستگی‌های زمانی و منطق موردنیاز برای انجام وظیفــه** را شناسایی می‌کنند. برای مثال، پژوهش‌ها دربارهٔ اجرای وظایف بر اساس زبان طبیعی نشان داده‌اند که سامانه‌ها می‌توانند عبارت «من گرسنه‌ام» را با استناد به دانش عمومی خود به وظیفهٔ «غذا تهیه کن» تبدیل کنند [3.A Review of Natural-Language-Instructed Robot Execution Systems].

این توانایی، فاصله زیادی با سامانه‌های قدیمی مبتنی بر قواعد دارد و نشان‌دهندهٔ جهت‌گیری به‌سمت تفسیر **سازگار و وابسته به زمینه** است.

LLMها همچنین در **رفع ابهام‌های زبانی** نقش مهمی دارند؛ زمانی که دستور ناقص است، مدل می‌تواند از کاربر سؤال بپرسد. این ویژگی با چارچوب‌های گفت‌وگومحور در تعامل انسان–ربات سازگار است، جایی که زبان نه‌تنها ابزار فرمان‌دادن، بلکه ابزاری برای **مذاکره، اصلاح و روشن‌سازی** است [7.ROS-LLM...].

به این ترتیب، زبان طبیعی به بستری برای تعامل دوطرفه تبدیل می‌شود؛ بستری که هم برای صدور دستور و هم برای تصمیم‌گیری مشترک کاربرد دارد.

1.3.پیشرفت‌های برنامه‌ریزی و استدلال مبتنی برLLM

مرور ادبیات نشان می‌دهد که **برنامه‌ریزی وظایف** یکی از حوزه‌هایی است که LLMها بیشترین تأثیر را در آن داشته‌اند. پژوهش‌ها ثابت کرده‌اند کهLLM ها

با وجود افزایش توانایی LLM ها، این مدل‌ها همچنان ممکن است دچار **توهم زایی** و استدلال نادرست شوند. در حوزهٔ رباتیک، چنین خطاهایی می‌توانند خطرناک باشند. از همین رو، **تولید مبتنی بر بازیابی (RAG)** به یک رویکرد مهم برای **گراندینگ و اعتبارسنجی** تبدیل شده است.

سامانهٔ **ARRC** نمونهٔ برجسته‌ای از این راهبرد است. در این سامانه، پایگاه دانشی شامل الگوهای حرکتی، راهبردهای ایمنی و قالب‌های وظیفه به‌شکل برداری ذخیره شده و در زمان برنامه‌ریزی، بخش مرتبط از آن بازیابی می‌شود [6.ARRC...].

سپس LLM با اتکا بر این دانش معتبر، برنامهٔ کنشی ساخت‌یافته را تولید می‌کند و در نهایت یک مرحلهٔ ارزیابی ایمنی، برنامه را کنترل می‌کند.

به همین ترتیب، سامانه‌های ماژولار مانند **ProgramPort** و **MetaMorph** از LLM برای تبدیل دستورات زبانی به ماژول‌های برنامه‌نویسی یا سیاست‌های سازگار با شکل‌بدن‌های مختلف ربات بهره می‌برند [1...].

این رویکردها قابلیت **مقیاس‌پذیری، تعمیم و ترکیب‌پذیری** را افزایش می‌دهند و به ربات‌ها اجازه می‌دهند مهارت‌ها را ترکیب کرده، به ابزارهای جدید سازگار شوند یا در محیط‌های متنوع وظایف پیچیده را انجام دهند.

در نهایت می‌توان گفت معماری‌های مبتنی بر LLM نه‌تنها به‌عنوان مفسر زبان، بلکه به‌عنوان **چارچوب‌های سازمان‌دهندهٔ رفتار ربات** عمل می‌کنند.

1.6. تعامل انسان–ربات و یادگیری تعاملی

LLMها نقش اساسی در گسترش ظرفیت‌های تعامل انسان–ربات ایفا می‌کنند. برای نمونه، سامانهٔ **ROS-LLM** نشان می‌دهد که ربات‌ها می‌توانند وارد گفت‌وگوهای چندمرحله‌ای شوند، بازخورد انسانی را دریافت کنند، رفتارهای سلسله‌مراتبی اجرا کنند و در طول زمان بهبود یابند [7.ROS-LLM...].

در این چارچوب‌ها، ورودی‌های چندوجهی مانند متن، تصویر یا داده‌های حسگر به ربات کمک می‌کند تا ابهام را کاهش دهد، خطا را اصلاح کند و به تغییرات محیط سازگار شود.

از سوی دیگر، مدل‌های مبتنی بر «عامل‌های مولد» امکان ذخیرهٔ حافظه، یادگیری از تجربه و حتی الگوگیری از سبک‌های تعامل انسانی را فراهم می‌کنند. نتیجهٔ این پیشرفت‌ها، حرکت به‌سمت ربات‌هایی است که تعامل **طبیعی‌تر، سازگارتر و مشارکتی‌تر** دارند.

1.7. دستاوردهای این مقاله

دستکاری اجسام به دقت عددی، مدل‌سازی هندسی و خط لوله‌های ادراکی بسیار تنظیم‌شده متکی هستند. اما مدل‌های زبانی بزرگ بُعد جدیدی را وارد این حوزه کرده‌اند :**استدلال معنایی**.

سامانه‌هایی مانند **LLM-GROP** و **VIMA** نشان داده‌اند که سرنخ‌های زبانی قادرند استدلال فضایی ربات را غنی کنند. این سامانه‌ها می‌توانند تنها با تکیه بر نشانه‌های سطح‌بالا، وظایفی نظیر جای‌گذاری بدون برخورد یا تقلید چندوجهی را انجام دهند [1...].

چارچوب‌های چندوجهی با یکپارچه‌سازی بینایی، زبان و حرکت، امکان می‌دهند ربات‌ها بدون نیاز به داده‌های اختصاصی، میان دسته‌های مختلفی از وظایف **تعمیم‌پذیری** پیدا کنند.

از سوی دیگر، ورود روش‌هایی مانند **RT-Grasp** که بر «استدلال تنظیم‌شده» تکیه دارند، یکی از موانع اصلی در دستکاری یعنی **دقت عددی** را هدف قرار داده است RT-Grasp .با ترکیب خروجی‌های استدلالی LLM و قالب‌های پیش‌بینی عددی ساخت‌یافته، توضیحات معنایی مدل را برای اصلاح پارامترهای دقیق گرفتن اشیا به‌کار می‌گیرد [5.RT-Grasp...].

این کار پلی میان **فهم زبانی–مفهومی** و **کنترل هندسی و دقیق** ایجاد می‌کند؛ پلی که پیش از این بیشتر بر دوش روش‌های تحلیلی یا یادگیری مبتنی بر داده بود.

پژوهش‌های دیگری نیز نشان داده‌اند که تنظیم کم‌نمونهٔ مدل‌های بینایی–زبانی مانند R3M و LIV ، عملکرد ربات را در وظایف دستکاری به‌شکل چشم‌گیری بهبود می‌دهد. این مدل‌ها به‌جای نیاز به داده‌های گسترده ربات، از ویدئوهای انسانی برای یادگیری بازنمایی‌های غنی از **قابلیت‌ها، ویژگی‌های سطحی و روابط بین اشیا** بهره می‌برند [1...].

در کنار این‌ها، سامانه‌هایی مانند **VoxPoser** از زبان برای استخراج محدودیت‌های قابل‌دستکاری استفاده می‌کنند؛ امری که امکان سازگاری لحظه‌ای با دستورات کاربر را فراهم می‌کند.

در مجموع، این پیشرفت‌ها نشان می‌دهند که حوزهٔ دستکاری رباتیک از یک حوزهٔ صرفاً عددی–هندسی به یک رویکرد **ترکیبی از معنا، چندوجهی‌بودن و استدلال ساخت‌یافته** حرکت کرده است؛ حرکتی که به ربات‌ها درکی نزدیک‌تر به انسان از تعاملات فیزیکی می‌دهد.

1.5. معماری‌های ماژولار و نسل مبتنی بر بازیابی (RAG)

تبدیل کنند و هنگامی‌که ورودی چندوجهی اضافه شود، دقت و گراندینگ به‌طور قابل‌توجهی افزایش می‌یابد [1.Large Language Models for Robotics: Opportunities, Challenges, and Perspectives].

پیشرفت در سامانه‌های اجرای وظایف از طریق زبان نیز نشان می‌دهد که ربات‌ها توانایی درک **مقصود غیرمستقیم** را پیدا کرده‌اند. برای مثال، می‌توانند دستور ضمنی را از جملاتی مانند «هوا سرد است» استنتاج کنند [3.A Review of Natural-Language-Instructed Robot Execution Systems].

این رفتار نشان‌دهندهٔ گذار از تطبیق الگوهای دستوری به **استدلال مبتنی بر دانش عمومی و زمینه‌محور** است.

از سوی دیگر، LLMها در **گفت‌وگوی تعاملی** برای اصلاح فرمان‌ها نیز مؤثرند. مدل قادر است هنگام نیاز به اطلاعات بیشتر، سؤال بپرسد تا دستور نهایی دقیق و قابل‌اجرا شود. چنین تعاملی با رویکردهای مدرن تعامل انسان–ربات سازگار است؛ جایی که زبان علاوه بر نقش فرمان‌دهی، نقش **ابزار مذاکره و رفع ابهام** را نیز دارد [7.ROS-LLM...].

در مجموع، NLUمبتنی بر LLM زیربنای شناختی سامانه‌های رباتیکی را شکل می‌دهد و آنها را قادر می‌سازد تا دستورهای پیچیده انسانی را تفسیر کرده و مطابق آن عمل کنند.

2.2.برنامه‌ریزی وظایف با استفاده از مدل‌های زبانی بزرگ

برنامه‌ریزی وظایف یکی از حوزه‌هایی است کهLLM ها بیشترین تحول را در آن ایجاد کرده‌اند. روش‌های کلاسیک برنامه‌ریزی ربات بر **برنامه‌ریزهای نمادین، الگوریتم‌های جست‌وجو یا یادگیری تقویتی مبتنی بر پاداش** تکیه داشتند. این روش‌ها در محیط‌های پویا، مبهم یا دارای پیچیدگی معنایی، با محدودیت‌های جدی روبه‌رو بودند.

امLLM ها این محدودیت را برطرف کرده و امکان برنامه‌ریزی سطح‌بالا با **درک زمینه‌ای و استدلال معنایی** را فراهم کرده‌اند.

پژوهش‌ها نشان می‌دهد کهLLM ها قادرند از روی دستورهای زبانی抽象، حتی بدون مشاهده کامل محیط، **توالی اقدامات بلندمدت** تولید کنند [1...].

یکی از نقاط قوتLLM ها، توانایی استخراج **دانش عمومی** و استفاده از آن در تدوین مراحل منطقی وظایف است. برای مثال، مدل به‌طور طبیعی می‌داند که برای تهیه یک نوشیدنی، ابتدا باید یک ظرف پیدا شده و سپس پر شود؛ حتی اگر این مراحل صریحاً در دستور ذکر نشده باشند.

بر اساس تلفیق تحلیل Summary.pdf و هشت مقالهٔ اصلی حوزهٔ LLM–رباتیک، این مقاله:

1. یک جمع‌بندی آکادمیک و یکپارچه از **درک زبان طبیعی، برنامه‌ریزی، استدلال، دستکاری و تعامل** در رباتیک مبتنی بر LLMارائه می‌دهد.
2. یافته‌های مربوط به **تنظیم تخصصی صنعتی، دستکاری چندوجهی، برنامه‌ریزی مبتنی بر بازیابی و چارچوب‌های تعاملی ROS** را ادغام می‌کند.
3. محدودیت‌های کنونی شامل **پایداری گراندینگ، ایمنی، و تعمیم در محیط واقعی** را تحلیل می‌کند.
4. چشم‌اندازی جامع برای مسیر آیندهٔ **هوش تجسم‌یافته مبتنی بر LLM** ترسیم می‌کند.

2.بدنه اصلی(Main Body)

2.1.درک زبان طبیعی برای سامانه‌های رباتیکی

درک زبان طبیعی (NLU) سنگ‌بنای توانایی ربات‌های مبتنی بر LLM برای **تفسیر نیت انسان** و تبدیل آن به رفتارهای قابل اجرا است. در ادبیات پژوهش، NLUمجموعه‌ای از فرایندها را شامل می‌شود: تحلیل ورودی زبانی، استخراج ساختار معنایی، شناسایی عناصر قابل‌اقدام، رفع ابهام، و تبدیل زبان طبیعی به فرمان‌های اجرایی.

رویکردهای اولیه به‌شدت بر **خط لوله‌های زبانی و بازنمایی‌های نمادین** تکیه داشتند. برای مثال، چارچوب‌های مبتنی بر هستی‌شناسی با استفاده از وابستگی‌های نحوی، افعال، اعضای بدن و محدودیت‌های حرکتی را استخراج کرده و از آنها برای ایجاد فرمان‌های ربات استفاده می‌کردند [2.Parsing Natural Language Sentences into Robot Actions].

این سامانه‌ها قادر بودند در صورت ناقص بودن اطلاعات (برای نمونه، «دستت را بالا ببر» بدون ذکر راست یا چپ)، از کاربر سؤال پرسیده و مسیر صحیح را مشخص کنند. همچنین این چارچوب‌ها برای بررسی **سازگاری فرمان** با وضعیت فیزیکی فعلی ربات ضروری بودند.

با ظهورLLM های مدرن، این قابلیت‌ها گسترش چشمگیری یافته است. مدل‌های زبانی بزرگ با تکیه بر بازنمایی‌های معنایی قوی و آموزش گسترده روی داده‌های متنی، قادرند **روابط پنهان بین دستورها، محیط و نیت ضمنی** را استخراج کنند. به‌عنوان نمونه، پژوهش‌ها نشان می‌دهد کهLLM ها می‌توانند بدون ورودی بصری نیز یک دستور 抽象را به توالی منسجم اقدامات

در زمینهٔ چندوجهی، مدل **LM-Nav**نشان می‌دهد که چگونه می‌توان زبان، ادراک بصری و مدل‌های ناوبری را برای تصمیم‌گیری هماهنگ ادغام کرد.

برخی پژوهش‌ها نیز بر **تصمیم‌گیری میان چند عامل** تمرکز دارند؛ جایی که LLMها قادرند تعامل میان چند ربات یا تعامل میان ربات و انسان را هماهنگ کنند. این توانایی برای محیط‌های مشارکتی ضروری است.

در مجموع، مجموعهٔ این پیشرفت‌ها نشان می‌دهد کهLLM ها از یک ابزار زبانی ساده فراتر رفته و به **عامل‌های تصمیم‌ساز هوشمند** تبدیل شده‌اند.

2.4.دستکاری و تعامل فیزیکی تقویت‌شده باLLM ها

وظایف دستکاری—مانند گرفتن، جابه‌جایی و چیدمان اشیا—به دلیل نیاز به **دقت هندسی، کنترل پیوسته و استدلال فضایی** از چالش‌برانگیزترین حوزه‌های رباتیک محسوب می‌شوند. روش‌های سنتی برای این نوع وظایف معمولاً بر بهینه‌سازی عددی، برآورد موقعیت اشیا و سیاست‌های کنترلی داده‌محور تکیه دارند. اماLLM ها امکان یک رویکرد **مفهومی–معنایی** را برای هدایت این فرآیند فراهم می‌کنند.

برای مثال، پژوهش‌هایی مانند **LLM-GROP**نشان می‌دهند که مدل‌های زبانی قادرند سرنخ‌های معنایی را استخراج کرده و به ربات کمک کنند تا **راهکارهای منطقی برای دستکاری** انتخاب کند؛ مانند این‌که اشیای شکننده را زیر اجسام سنگین قرار ندهد [1...].

چارچوب‌های چندوجهی مانند **VIMA**و **TIP**نیز با ترکیب متن و تصویر، به ربات‌ها امکان می‌دهند **دستکاری را از طریق نشانه‌های زبانی و بصری** یاد بگیرند و مهارت‌ها را به مواردی تعمیم دهند که پیش‌تر ندیده‌اند.

پیشرفت مهم دیگری در این حوزه، روش **RT-Grasp**است. در این رویکرد، قبل از آنکه مدل خروجی عددی دقیق (مثل زاویهٔ گرفتن یا موقعیت گیره) تولید کند، ابتدا از یک مرحلهٔ **استدلال زبانی ساخت‌یافته** استفاده می‌شود. این امر باعث می‌شود مدل ابتدا ماهیت شیء (مثلاً «فنجان»، «عینک»، «توپ») و اصول صحیح گرفتن آن را توضیح دهد، سپس پارامترهای دقیق را محاسبه کند
[5.RT-Grasp...].

این ساختار پل مهمی میان **استدلال سطح‌بالا** و **کنترل فیزیکی دقیق** ایجاد می‌کند؛ شکافی که سال‌ها چالشی اساسی در رباتیک بود.

روش‌های دیگری مانند **R3M**و **LIV**نشان داده‌اند که آموزش مدل‌های بینایی–زبانی با داده‌های ویدئویی انسان، بازنمایی‌هایی غنی از **قابلیت‌ها، روابط اشیا و ویژگی‌های محیط** ایجاد می‌کند. این باعث می‌شود ربات‌ها با

در رویکردهای ترکیبی مانند **LLM+P**و **SayCan**، مدل‌های زبانی با برنامه‌ریزهای رسمی مبتنی بر PDDL ادغام می‌شوند. در این مدل‌ها، LLM وظایف سطح‌بالا را تفسیر کرده و برنامه‌ریزهای سنتی تضمین می‌کنند که این عملیات با محدودیت‌های جهان واقعی سازگار باشد [1...].

علاوه بر این، روش‌هایی مانند **grounded decoding**با هم‌تراز کردن خروجی LLM با مدل‌های فیزیکی محیط، قابلیت اتکا و سازگاری برنامه‌ریزی را افزایش می‌دهند. این هم‌ترازی احتمال تولید اقدامات غیرممکن یا ناسازگار را کاهش می‌دهد [1...].

در حوزهٔ ناوبری، سیستم‌هایی مانند **LM-Nav**با ترکیب مدل‌های پیش‌آموزش‌دیدهٔ بینایی، مدل‌های زبان و استدلال فضایی، به ربات امکان می‌دهند که بدون نیاز به داده‌های پیمایش برچسب‌خورده، مسیرهای پیچیده را دنبال کند [1...].

مجموع این پیشرفت‌ها نشان می‌دهد کهLLM ها پایهٔ یک **چارچوب برنامه‌ریزی انعطاف‌پذیر، قدرتمند و قابل‌تعمیم** را تشکیل می‌دهند که می‌تواند وظایف پیچیده را تحلیل کند، نیت انسان را استخراج کند و برنامه‌های قابل‌اعتماد برای محیط‌های پویا تولید کند.

2.3.استدلال پیچیده و تصمیم‌گیری

مزیت مهم دیگرLLM ها توانایی آن‌ها در **استدلال سطح‌بالا** و تصمیم‌گیری است. برخلاف سیستم‌های سنتی که تنها به قواعد برنامه‌ریزی‌شده متکی هستند، LLMها قادرند بین گزینه‌ها انتخاب کنند، پیامدهای احتمالی را بسنجند و در طول اجرای وظایف سازگار شوند.

پژوهش‌های مختلف نشان می‌دهد کهLLM ها می‌توانند از **حافظه معنایی** بهره ببرند، میان وظایف مختلف تعمیم پیدا کنند و با استفاده از قالب‌های استدلالی ساخت‌یافته، تصمیم‌های منطقی اتخاذ کنند. برای مثال، استفاده از **اشاره‌های خطا** به ربات کمک می‌کند که برنامهٔ تولیدشده را اصلاح و بهینه کند [1...].

مدل‌هایی مانند **ReAct**استدلال و عمل را در یک چرخه ترکیب می‌کنند؛ در نتیجه ربات می‌تواند حالت درونی خود را حفظ کند، پاسخ‌های نادرست را فیلتر کند و از تصمیم‌های متناقض جلوگیری کند.

در حوزهٔ ایمنی، روش **KnowNo**با تخمین عدم قطعیت، مانع از انجام اقداماتی می‌شود که مدل نسبت به آنها اعتماد کافی ندارد [1...].

این ویژگی در وظایف چندمرحله‌ای و حساس، اهمیت ویژه‌ای دارد.

افزون بر این، **عامل‌های مولد (Generative Agents)** امکان ذخیرهٔ تجربه، شکل‌گیری حافظهٔ درازمدت و سازگاری رفتاری را فراهم می‌کنند. این توانایی‌ها ربات‌ها را به موجوداتی **یادگیرنده، سازگار و تعاملی‌تر** تبدیل می‌کند.

در مجموع، رویکردهای تعاملیLLM محور باعث می‌شوند ربات‌ها بتوانند:

- در زمان واقعی یاد بگیرند،
- از بازخورد انسان بهره ببرند،
- و در محیط‌های ناپایدار یا ناشناخته عملکرد باثبات‌تری داشته باشند.

2.6.رویکردهای ماژولار برای هوش رباتیکی مقیاس‌پذیر

معماری ماژولار یکی از اصول مهم طراحی در سامانه‌های رباتیکی استLLM .ها این رویکرد را به سطح جدیدی رسانده‌اند. در این معماری، یک وظیفهٔ پیچیده به مجموعه‌ای از **ماژول‌های قابل‌استفادهٔ مجدد** تقسیم می‌شود؛ مانند:

- مهارت‌ها
- سیاست‌های کنترلی
- برنامه‌ریزها
- ماژول‌های ادراک
- درخت‌های رفتار

LLMها با تبدیل توصیف‌های زبانی به این ماژول‌ها، امکان **ساخت سازه‌های رفتاری پیچیده** را بدون نیاز به برنامه‌نویسی دستی فراهم می‌کنند.

نمونه‌ای از این رویکرد، **PROGRAMPORT**است که زبان طبیعی را به **ماژول‌های برنامه‌ای** برای دستکاری اشیا تبدیل می‌کند [1...].

در این سامانه، به‌جای تولید مستقیم کد، ساختار معنایی دستورها به بلوک‌های عصبی قابل‌استفادهٔ مجدد تبدیل می‌شود که قابلیت تعمیم در محیط‌ها و وظایف جدید را دارند.

رویکرد دیگر، استفاده از **MetaMorph**است؛ سیستمی که با مدل‌های ترنسفورمر، سیاست‌های کنترل سازگار با شکل‌بدن‌های مختلف ربات را یاد می‌گیرد
[1...].

این امر باعث می‌شود ربات بتواند بدون طراحی مجدد، به سخت‌افزارهای جدید سازگار شود.

داده‌های بسیار کمتر، مهارت‌های دستکاری را یاد بگیرند [1...].

در سمت دیگر، سامانه‌هایی مانند **VoxPoser**به ربات امکان می‌دهند تا محدودیت‌های دستکاری را مستقیماً از متن استخراج کرده و رفتار خود را به‌طور پویا تنظیم کند.

مجموع این پیشرفت‌ها نشان می‌دهد که دستکاری رباتیک در حال حرکت از یک مسئلهٔ صرفاً هندسی به یک مسئلهٔ **ترکیبی از معنا، ادراک و استدلال چندوجهی** است—رویکردی که به مهارت‌هایی نزدیک به انسان منجر می‌شود.

2.5.راهبردهای تعاملی و ادغام بازخورد انسانی

ربات‌ها در محیط‌های واقعی باید **عملکرد خود را بر اساس بازخورد انسان، خطاهای گذشته یا تغییرات محیط اصلاح کنند**LLM .ها ابزارهای بسیار قدرتمندی برای هدایت این تعامل فراهم کرده‌اند.

سامانهٔ **TEXT2REWARD**با تبدیل بازخورد زبانی انسان به **کدهای پاداش**، به عامل‌های یادگیری تقویتی امکان می‌دهد اصلاحات انسانی را در مسیر یادگیری خود اعمال کنند [1...].

رویکرد **InstructRL**نیز از LLM برای تولید سیاست‌های ابتدایی استفاده می‌کند و سپس به عامل اجازه می‌دهد با تکیه بر بازخورد انسان، رفتار خود را اصلاح کند.

چارچوب‌های تعامل‌محور مانند **LILAC**به کاربران اجازه می‌دهند تنها با زبان طبیعی و (در صورت نیاز) تصاویر، مسیر ربات یا پارامترهای کنترل را تنظیم کنند
[1...].

این روش‌ها باعث می‌شوند ربات‌ها سریع‌تر و طبیعی‌تر به خواسته‌های کاربر سازگار شوند.

سامانهٔ **ROS-LLM**این مفهوم را گسترش می‌دهد و از LLM برای به‌کارگیری بازخورد انسان در چرخهٔ برنامه‌ریزی و اجرا استفاده می‌کند. در این سامانه، ربات می‌تواند:

- خطاهای اجرا را تشخیص دهد،
- پیشنهادهای اصلاحی کاربر را تفسیر کند،
- و برنامه را بازسازی یا بهبود دهد [7.ROS-LLM...].

از محیط انجام شود [6...].

در مجموع، RAGسنتز قدرتمندی میان **قطعیت و ایمنی در رباتیک کلاسیک** و **انعطاف و استدلالLLM ها** ایجاد می‌کند. نتیجه، معماری‌ای است که برای برنامه‌ریزی خودکار، امن و مقیاس‌پذیر در دنیای واقعی بسیار مناسب است.

2.8.ادراک، درک صحنه و ادغام چندوجهی

ادراک، رکن اساسی هوش تجسم‌یافته است. ربات برای تعامل با جهان باید بتواند اطلاعات بصری، فضایی و معنایی را به‌صورت قابل‌اعتماد پردازش کند . LLMها و مدل‌های بینایی–زبانی (VLMs) در این زمینه پیشرفت‌های قابل‌توجهی ایجاد کرده‌اند.

سیستم **PaLM-E**نمونه‌ای مهم از ادغام ورودی‌های چندوجهی است. در این مدل، **تصاویر، بردارهای حالت و متن** به‌صورت مشترک پردازش می‌شوند و مدل قادر است روی مجموعهٔ گسترده‌ای از وظایف بینایی و دستکاری عمل کند
[8.Large Language Models for Robotics: A Survey].
این یکپارچگی به ربات اجازه می‌دهد به پرسش‌هایی مانند «داخل کشو چه چیزی است؟» یا «فنجان سبز کجاست؟» پاسخ مناسب بدهد.

سامانهٔ **LM-Nav**نیز با ترکیبembedding های CLIP و نقشه‌برداری مبتنی برViNG ، به ربات امکان می‌دهد که صرفاً بر اساس زبان طبیعی و بدون نیاز به دادهٔ پیمایش برچسب‌خورده، به نقاط خاصی در محیط ناوبری کند [8...].

این ترکیب، هم ادراک معنایی و هم استدلال فضایی را به‌شکلی یکپارچه فراهم می‌کند.

در حوزهٔ دستکاری، ARRCبا استفاده از **AprilTags +**عمق موقعیت سه‌بعدی دقیق اشیا را استخراج می‌کند و این اطلاعات برای اجرای صحیح وظایف pick & place ضروری است [6.ARRC...].

به‌طور مشابه، مدل‌های چندوجهی مانند **VIMA**و **TIP**از زبان و تصویر برای تعریف محدودیت‌ها و اهداف دستکاری استفاده می‌کنند و به ربات امکان می‌دهند روابطی مثل «بلوک را سمت چپ مخروط قرار بده» را بفهمد.

روند مهم دیگر، **پیش‌آموزش بینایی–زبانی در مقیاس بزرگ** است؛ مدل‌هایی مانند **R3M**و **LIV**که روی ویدئوهای انسانی آموزش دیده‌اند، ویژگی‌هایی مانند:

- قابلیت‌های اشیا،

سامانه‌های مبتنی بر مدل‌های بینایی–زبانی مانند **NLMap**نیز برای ایجاد نقشه‌های معنایی و انعطاف‌پذیر مورد استفاده قرار می‌گیرند و ربات را قادر می‌کنند دستورهای پیچیده و محیط‌های باز را بهتر درک کند [1...].

در مجموع، این رویکردها LLM را از یک ابزار صرفاً زبانی فراتر برده و آن را به **چارچوبی سازمان‌دهنده برای ادراک، کنترل و استدلال ربات** تبدیل کرده‌اند؛ چارچوبی که مقیاس‌پذیری و سازگاری چشمگیری فراهم می‌کند.

2.7.تولید مبتنی بر بازیابی (RAG) برای برنامه‌ریزی مطمئن در رباتیک

با افزایش توانایی‌هایLLM ها، یک چالش مهم همچنان برجا می‌ماند : **توهم‌زایی و استدلال نادرست** .در رباتیک، این خطاها می‌توانند منجر به اقدامات خطرناک یا غیرممکن شوند. ازاین‌رو، رویکرد **تولید مبتنی بر بازیابی (Retrieval-Augmented Generation)**به عنوان یک راهکار مهم برای **گراندینگ و کاهش خطا** مطرح شده است.

سامانهٔ **ARRC**نمونه‌ای برجسته از کاربرد RAG در رباتیک است [6.ARRC...].
در این سامانه:

- یک پایگاه دانش برداری شامل **الگوهای حرکتی، قالب‌های وظیفه، راهبردهای ایمنی و رهنمودهای عملیاتی** ذخیره می‌شود.
- ماژول بازیابی، بخش‌های مرتبط را بر اساس دستور کاربر و وضعیت محیط استخراج می‌کند.
- LLMبا تکیه بر این دانش معتبر، **یک برنامهٔ ساخت‌یافتهٔ JSON**تولید می‌کند.
- در نهایت، یک لایهٔ «دروازهٔ ایمنی» برنامه را پیش از اجرا ارزیابی می‌کند.

این pipeline باعث می‌شود احتمال تولید اقدامات اشتباه یا ناسازگار از سوی LLMبه‌شکل چشم‌گیری کاهش یابد.
علاوه بر این، RAGنیاز به آموزش مجدد مدل را کاهش می‌دهد؛ زیرا افزودن مهارت‌های جدید تنها به اضافه‌کردن اطلاعات به پایگاه دانش نیاز دارد.

مزیت دیگر این رویکرد، امکان **استدلال چندوجهی مبتنی بر محیط** است. برای نمونه، در ARRC با استفاده از AprilTags و دادهٔ عمق، موقعیت دقیق اشیا استخراج شده و به LLM داده می‌شود تا برنامه‌ریزی با شناخت واقعی‌تری

در روش‌هایی مانندSayCan ، LLMوظایف سطح‌بالا را پیشنهاد می‌دهد و یادگیری تقویتی بررسی می‌کند که آیا این وظایف قابل اجرا هستند یا خیر [1...].

این ساختار از امکان تولید دستورهای غیرعملی جلوگیری می‌کند.

مدل‌هایی مانند **VIMA**، **TIP**و **MetaMorph** نیز با رمزگذاری حرکات و سیاست‌های چندوجهی، پل میان **زبان و کنترل پیوسته** را ایجاد می‌کنند.

در مجموع، کنترل در نسل جدید ربات‌ها دیگر صرفاً مبتنی بر منطق دست‌نویس نیست؛ بلکه بر **ساختارهای زبانی قابل تفسیر** استوار است که در لایه‌های پایین‌تر توسط روش‌های کلاسیک رباتیک پشتیبانی و اجرا می‌شوند.

2.10.اجرای مقید به ایمنی(Safety-Constrained Execution)

ایمنی یکی از مهم‌ترین دغدغه‌ها در رباتیک است، به‌ویژه زمانی که ربات در کنار انسان‌ها کار می‌کند یا با اشیای حساس سروکار دارد. مدل‌های زبانی بزرگ، با وجود توانایی‌های چشمگیرشان، ماهیتی **احتمالی** دارند و ممکن است خروجی‌هایی تولید کنند که از نظر فیزیکی نادرست یا خطرناک باشند.
به همین دلیل، بسیاری از سامانه‌ها لایه‌های اضافی برای تضمین ایمنی در نظر گرفته‌اند.

در سامانهٔ **ARRC**، هر اقدام پیشنهادی از سوی LLM باید از یک مجموعهٔ چندلایه از «دروازه‌های ایمنی» عبور کند [6.ARRC...]:

- **اعتبارسنجی محدودهٔ کاری :(Workspace Validation)** محل‌های هدف بررسی می‌شوند تا خارج از محدودهٔ قابل‌دسترس ربات نباشند.
- **محدودیت‌های سرعت و شتاب :**حرکات از سقف‌های تعیین‌شده فراتر نمی‌روند.
- **نظارت بر گشتاور و نیروی گریپر :**در صورت افزایش نیروی نامعمول، عملیات قطع می‌شود.
- **زمان‌سنج برای هر مرحله :**اگر یک حرکت بیش از حد طولانی شود، عملیات متوقف می‌شود.
- **حالت‌های عقب‌نشینی اضطراری :**در صورت بروز خطاهای مکرر، ربات به حالت امن بازمی‌گردد.
- **تعداد تکرار محدود برای تلاش مجدد.**

- بافت‌ها،
- روابط عملکردی،
- و نشانه‌های فضایی

را به‌خوبی یاد می‌گیرند [1...].

این بازنمایی‌ها موجب افزایش **کارایی نمونه (sample efficiency)** و **تعمیم بدون آموزش اضافی** در ربات‌های واقعی می‌شود.

در نهایت، می‌توان گفت ادراک مبتنی برLLM ها تنها هندسهٔ محیط را توصیف نمی‌کند، بلکه **معنای صحنه** را نیز استخراج می‌کند—توانایی‌ای که برای تعامل در محیط‌های واقعی، پویا و غیرساخت‌یافته ضروری است.

2.9.کنترل و اجرا در رباتیک مبتنی برLLM

کنترل—یعنی ترجمهٔ هدف به حرکت فیزیکی—جایی است که **استدلال سطح‌بالای LLM** با **محدودیت‌های سخت‌افزار ربات** تلاقی می‌کند. از آنجا کهLLM ها خروجی‌هایی با سطح انتزاع بالا تولید می‌کنند، لازم است این خروجی‌ها با ماژول‌های کنترل پایین‌سطح و برنامه‌ریزهای حرکتی ادغام شوند.

در بسیاری از سامانه‌ها، خروجی LLM به‌صورت ساخت‌یافته ارائه می‌شود، مانند:

- توالی‌های) JSON در(ARRC ،
- کدهای) Python در(ROS-LLM ،
- درخت‌های رفتاریXML ،
- برنامه‌های شبه‌کد) در(Cap ،
- یا هدف‌های میانی برای یادگیری تقویتی) در.(SayCan

در **ROS-LLM**، این خروجی‌ها مستقیماً به سرویس‌ها و اکشن‌های ROS تبدیل می‌شوند و به این ترتیب مدل می‌تواند با:

- موتورهای مسیریابی،
- حل‌کننده‌های سینماتیک،
- و کتابخانه‌هاییماندMoveIt

به‌طور کامل یکپارچه شود [7.ROS-LLM...].

- و یادگیری از تجربه.

3.درک ضمنی وcommonsense

سامانه‌های NLexe قادرند از جملات غیرمستقیم مانند «تشنگی‌ام شده» معنای **درخواست آب** را استنباط کنند [3...].

این توانایی برای تعامل طبیعی و انسانی کاملاً ضروری است.

4.تعامل چندوجهی

برخی سامانه‌ها علاوه بر متن، از تصویر یا داده‌های حسگر نیز استفاده می‌کنند. برای مثال، کاربر می‌تواند تصویر را نشان دهد و بگوید:

»لیوان را کنار این قرار بده«.

و ربات موقعیت اشاره‌شده را تفسیر می‌کند.

5.عامل‌های مشارکتی و حافظه‌دار

پژوهش‌های جدید دربارهٔ «عامل‌های مولد» نشان می‌دهند که ربات می‌تواند حافظهٔ بلندمدت شکل دهد، الگوهای رفتاری پایدار ایجاد کند و تجربهٔ گذشته را برای بهبود تعامل استفاده کند [1...].

در نتیجه، HRIمبتنی بر LLM به ربات‌هایی منجر می‌شود که:

- قابل‌اعتمادتر،
- قابل‌تعاملی‌تر،
- قابل‌سازگاری‌تر،
- و برای کاربران غیرمتخصص بسیار قابل‌استفاده‌تر هستند.

2.12.محدودیت‌ها، چالش‌ها و مسیرهای آینده

با وجود پیشرفت‌های چشمگیر، رباتیک مبتنی بر LLM هنوز با چالش‌های مهمی روبه‌روست.

1.گراندینگ و هم‌ترازی با جهان واقعی

LLMها گاهی اقداماتی پیشنهاد می‌دهند که از نظر فیزیکی ممکن نیست. سامانه‌هایی مانند ARRC و SayCan این مشکل را کاهش می‌دهند، اما **گراندینگ کامل چندوجهی** همچنان دشوار است.

2.دقت عددی و محدودیت‌های فیزیکی

مدل‌های زبانی در استدلال و معنا بسیار توانمندند، اما در:

این لایه‌ها تضمین می‌کنند که حتی اگر LLM برنامه‌ای غیرایمن یا ناسازگار تولید کند، **اجرا هرگز بدون کنترل و اعتبارسنجی صورت نمی‌گیرد**.

رویکرد **KnowNo**نیز یک لایهٔ ایمنی نظری مکمل ارائه می‌دهد. این مدل با برآورد عدم‌قطعیت خروجی، از انجام اقداماتی جلوگیری می‌کند که مدل نسبت به آن‌ها اعتماد کافی ندارد [1...].

سامانهٔ **ROS-LLM**نیز از «پرچم خطا» در ساختار MDP استفاده می‌کند تا ربات بتواند شکست اجرای وظایف را تشخیص داده و درخواست بازخورد یا اصلاح از کاربر داشته باشد [7.ROS-LLM...].

در مجموع، می‌توان گفت نسل جدید سامانه‌های رباتیکی مبتنی بر LLM به یک **معماری دوگانه** نیاز دارد:
LLMبرای استدلال، و سامانه‌های نمادین/مهندسی‌شده برای ایمنی. این ترکیب کلید ورود ربات‌ها به محیط‌های واقعی و انسانی است.

2.11.تعامل انسان–ربات و هوش مشارکتی

تعامل انسان–ربات (HRI) یکی از حوزه‌هایی است کهLLM ها بیشترین اثرگذاری را در آن داشته‌اند.
مشکل قدیمی رباتیک این بود که کاربران غیرمتخصص نمی‌توانستند ربات‌ها را به آسانی کنترل یا برنامه‌ریزی کنند، اما مدل‌های زبانی این شکاف را پر کرده‌اند.

1.تعریف وظایف از طریق مکالمه

سامانه‌هایی که از LLM استفاده می‌کنند، می‌توانند دستورهای پیچیده را از جملات کاملاً طبیعی استخراج کنند.
برای مثال، کاربر می‌تواند بگوید:

»میز را تمیز کن و ظرف‌ها را داخل سینک بگذار«.

LLMاین جمله را به مجموعه‌ای از وظایف شامل حرکت، گرفتن، جابه‌جایی و رهاسازی ترجمه می‌کند [1...].

2.بازخورد و اصلاح تعاملی

در **ROS-LLM**، ربات پس از هر مرحله می‌تواند بازخورد کاربر را دریافت کرده و رفتار خود را اصلاح کند [7.ROS-LLM...].
این فرآیند مشابه روش‌های آموزشی انسانی است:

- مشاهدهٔ رفتار،
- ارائهٔ اصلاح،

- و امنیت انسان‌ها

اهمیت بیشتری می‌یابد.

3.نتیجه‌گیری(Conclusion)

ظهور مدل‌های زبانی بزرگ (LLMs) بنیان‌های رباتیک مدرن را دگرگون کرده است. این مدل‌ها امکان سطحی از **انتزاع، استدلال و تعامل طبیعی** را فراهم کرده‌اند که پیش از این در سامانه‌های رباتیکی وجود نداشت. مرور انجام‌شده— که حوزه‌هایی همچون برنامه‌ریزی چندوجهی، دستکاری مبتنی بر استدلال، تنظیم تخصصی صنعتی، کنترل مبتنی بر بازیابی و چارچوب‌های-ROS محور را دربر می‌گیرد—نشان می‌دهد کهLLM ها دیگر صرفاً نقش مکمل ندارند، بلکه به **هستهٔ شناختی سیستم‌های رباتیکی** تبدیل شده‌اند.

روند غالب در تمام پژوهش‌های بررسی‌شده، گذار از پردازش زبان مبتنی بر قواعد به سوی **درک معنایی غنی و انعطاف‌پذیر** است. در گذشته، سامانه‌های مبتنی بر هستی‌شناسی [2.Parsing Natural Language Sentences into Robot Actions] ربات‌ها را به مجموعه محدودی از الگوهای دستوری مقید می‌کردند. اما LLMهای امروزی با تکیه بر دانش جهان‌واقعی و توانایی استدلال مبتنی بر زمینه، قادرند دستورات **مبهم، غیرمستقیم یا ناقص** را تحلیل کنند، نیت انسان را تشخیص دهند و با تغییرات محیطی سازگار شوند. این توانایی گذار ربات‌ها را از «پیرو دستور» به «همکار هوشمند» ممکن کرده است.

در زمینهٔ برنامه‌ریزی، مدل‌های زبانی اکنون قادرند **برنامه‌های بلندمدت، زمینه‌محور و قابل‌اجرا** تولید کنند و آن‌ها را با کمک مکانیزم‌های گراندینگ و برنامه‌ریزهای کلاسیک تکمیل نمایند [1.Large Language Models for Robotics: Opportunities, Challenges, and Perspectives]. ادغام ورودی‌های چندوجهی و دانش پیشین نشان داده است که چشم‌انداز LLMها به‌عنوان «برنامه‌ریزهای عمومی» روزبه‌روز قابل‌دسترس‌تر می‌شود.

در حوزهٔ دستکاری، روش‌هایی مانند **VIMA**، **LLM-GROP**و **RT-Grasp** [5.RT-Grasp...] نشان داده‌اند که زبان طبیعی می‌تواند راهنمایی برای درک ویژگی‌های اشیا، قیود ایمنی و راهبردهای دستکاری فراهم کند. ادغام استدلال زبانی با کنترل پیوسته، یکی از شکاف‌های دیرینه در رباتیک را پُر کرده است.

معماری‌های ماژولار مانند **PROGRAMPORT**و **MetaMorph**نشان می‌دهند کهLLM ها قادرند **ساختارهای رفتاری مقیاس‌پذیر** ایجاد کنند. در کنار آن، رویکردهای RAG مانند **ARRC** [6.ARRC...]

- محاسبات دقیق،
- هندسهٔ سه‌بعدی،
- پیش‌بینی پیوستهٔ حرکت

ضعیف‌تر عمل می‌کنند. از همین رو روش‌هایی مانند **RT-Grasp**ضروری‌اند [5.RT-Grasp...].

3.نیاز به داده و مشکلات انتقال دامنه

تنظیم تخصصی مدل‌ها برای کارهای صنعتی نیازمند منابع محاسباتی و دادهٔ زیاد است، حتی با وجود روش‌های بهینه مانندQLoRA [4.Domain-Specific Fine-Tuning...].

4.ایمنی و اعتبارسنجی

از آنجا کهLLM ها ذاتاً احتمالی هستند، نمی‌توان ایمنی کامل را تنها با اتکا به آن‌ها تضمین کرد.
به همین دلیل، اجرای امن نیازمند **محدودکننده‌های سخت‌افزاری** و **اعتبارسنجی نمادین** است.

5.تعمیم در دنیای واقعی

ربات‌ها اغلب در محیط‌های واقعی نسبت به شبیه‌سازی عملکرد ضعیف‌تری دارند.
مدل‌های چندوجهی کمک می‌کنند، اما مسئله هنوز حل نشده است.

6.تأخیر و نیازهای زمان واقعی

استنتاج LLM ممکن است کند باشد و برای وظایف حساس به زمان مناسب نباشد.
بهینه‌سازی مدل‌ها یا اجرای روی سخت‌افزار محلی می‌تواند این مشکل را کاهش دهد.

7.وظایف بلندمدت

برنامه‌های چندمرحله‌ای طولانی، هنوز به دلیل خطاهای تجمعی و محدودیت پنجرهٔ زمینه، چالش‌برانگیز هستند.

8.ملاحظات اخلاقی و اجتماعی

با افزایش خودمختاری ربات‌ها، نگرانی‌هایی مانند:

- مسئولیت،
- حریم خصوصی،
- اثرات شغلی،

ثابت کرده‌اند که با تکیه بر پایگاه دانش واقعی، می‌توان خطا و توهم‌زایی مدل را کنترل کرده و برنامه‌ریزی قابل‌اعتماد ارائه داد.

در حوزهٔ تعامل انسان–ربات، سامانه‌هایی نظیر **ROS-LLM** [7.ROS-LLM...] نشان داده‌اند که زبان طبیعی می‌تواند **واسط اصلی تعامل** باشد. این امکان به کاربران غیرمتخصص اجازه می‌دهد ربات‌ها را اصلاح، هدایت یا برنامه‌ریزی کنند—آن هم بدون نیاز به دانش فنی.

با وجود این پیشرفت‌ها، چالش‌هایی همچون **دقت عددی، ایمنی، تعمیم، پردازش زمان واقعی و مسائل اخلاقی** همچنان پابرجاستLLM .ها ممکن است در برخی شرایط رفتار غیرقابل‌اعتماد نشان دهند، بنابراین ترکیب آن‌ها با مکانیزم‌های محافظتی و مدل‌های مبتنی بر فیزیک ضروری است. علاوه بر این، چالش‌های حقوقی و اجتماعی مرتبط با ربات‌های خودمختار همچنان نیازمند مطالعهٔ عمیق هستند.

با نگاه به آینده، روشن است که همگرایی مدل‌های چندوجهی، ادراک مبتنی بر حسگر، استدلال مبتنی بر بازیابی و کنترل آگاه از فیزیک، مسیر را به سمت **نسل جدیدی از هوش تجسم‌یافته** باز می‌کند. ربات‌هایی که نه‌تنها می‌توانند وظایف را اجرا کنند، بلکه قادرند **اهداف را بفهمند، نیازها را پیش‌بینی کنند و با انسان‌ها همکاری طبیعی‌تری داشته باشند.**

مسیر شکل‌گرفته در این مقاله نشان می‌دهد که رباتیک مبتنی بر LLM به‌تدریج از یک نوآوری پژوهشی به سمت **ستون فقرات سامانه‌های خودمختار آینده** حرکت می‌کند. پیشرفت در زمینه‌های گراندینگ، ایمنی و ادغام چندوجهی، کلید گشودن ظرفیت کامل ربات‌های هوشمند مبتنی بر زبان خواهد بود.

4.منابع(References)

در ادامه، منابع مطابق قالبی که در فایل اصلی استفاده شده، فهرست می‌شود:

1. **Large Language Models for Robotics: Opportunities, Challenges, and Perspectives**
2. **Parsing Natural Language Sentences into Robot Actions**
3. **A Review of Natural-Language-Instructed Robot Execution Systems**
4. **Domain-Specific Fine-Tuning of Large Language Models for Interactive Robot Programming**
5. **RT-Grasp: Reasoning Tuning Robotic Grasping via Multi-modal Large Language Model**
6. **ARRC: Advanced Reasoning Robot Control—Knowledge-Driven Autonomous Manipulation Using Retrieval-Augmented Generation**
7. **ROS-LLM: A ROS Framework for Embodied AI with Task Feedback and Structured Reasoning**
8. **Large Language Models for Robotics: A Survey**