

## ۱. مقدمه

### ۱.۱. پیشزمینه و انگیزه

یکپارچگی مدل‌های زبانی بزرگ (LLMs) با رباتیک، نقطه عطفی در نحوه فهم، استدلال و تعامل ربات‌ها با جهان فیزیکی محسوب می‌شود. در حالی که رباتیک سنتی به برنامه‌نویسی ساخت‌یافته، سیاست‌های کنترلی تخصصی و مدل‌های دقیق مهندسی شده 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 نشان‌دهنده 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.<sup>[8]</sup> در گذشته، سامانه‌های رباتیکی نیازمند برنامه‌نویسی صریح و دانش دامنه‌ای گسترشده بودند. اما LLM‌ها همچون موتورهای استدلالی عمومی عمل می‌کنند که قادرند معنای زمینه‌ای را استخراج کنند، برنامه انجام وظایف را تولید کنند و این برنامه‌ها را به دستورهای قابل اجرا تبدیل کنند. این قابلیت امکان می‌دهد وظایف پیچیده – از ناوبری تا دستکاری – به صورت مکالمه‌ای تعریف شوند، نه با گذنویسی 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.<sup>[1]</sup> 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 قابل اعتماد در محیط‌های واقعی ضروری است. چنین توانایی‌هایی نشان می‌دهد که استدلال مبتنی بر LLM عنصر مرکزی در حرکت بهسوی خودمختاری، سازگاری و همراستایی بیشتر با انسان است.

### ۱.۲. 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 قابل اجرا است. پژوهش‌های اولیه تمرکز خود را بر خط‌لوله‌های 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 می‌کرند.<sup>[2]</sup> 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.

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 به تدریج نقش هسته شناختی را بر عهده می‌گیرند؛ هسته‌ای که ادراک، تصمیم‌گیری، کنترل و تعامل انسان – ربات را یکپارچه می‌کند.

با انگیزه پیشرفت‌های سریع این حوزه، این مقاله یک جمع‌بندی جامع از دستاوردهای اخیر در رباتیک مبتنی بر LLM ارائه می‌دهد. در این مقاله بررسی می‌کنیم که LLM‌ها چگونه برنامه‌ریزی وظایف مبتنی بر زبان طبیعی، ادراک چندوجهی، کنترل دستکاری، تصمیم‌گیری تعاملی و ساختارهای برنامه‌نویسی مازولات برای ربات‌ها را تقویت می‌کنند. همچنین به روش‌هایی مانند برنامه‌ریزی مبتنی بر گراندینگ<sup>[1]</sup> تحلیل نحوی مبتنی بر هسته‌شناسی<sup>[2]</sup> و چارچوب‌های نظاممند آموزش ربات از طریق زبان انسان<sup>[3]</sup> می‌پردازیم.

ما علاوه بر این، راهبردهای تنظیم تخصصی مدل‌های زبانی برای برنامه‌نویسی صنعتی ربات‌ها<sup>[4]</sup> و روش‌های دستکاری مبتنی بر استدلال چندوجهی مانند<sup>[5]</sup> را بررسی می‌کنیم. همچنین پیشرفت‌های اخیر در سامانه‌های کنترل ربات مبتنی بر بازیابی اطلاعات<sup>[6]</sup> و چارچوب‌های ساخت‌یافته مبتنی بر ROS<sup>[7]</sup> را مرور می‌کنیم.

با سازمان‌دهی این یافته‌ها در حوزه‌های برنامه‌ریزی، استدلال، دستکاری، ادراک، معماری‌های کنترلی و تعامل، یک نگاه یکپارچه از وضعیت کنونی هوش تجسم‌یافته مبتنی بر LLM ارائه می‌دهیم. در پایان نیز چالش‌هایی مانند اینمنی، 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

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

که LLMها از یک «پردازشگر زبان» فراتر رفته و به «عامل‌های استدلالی فعال» تبدیل شده‌اند که در هسته معماری کنترل ربات قرار می‌گیرند.

## ۱. دستکاری، ادراک و مهارت‌های تجسم‌یافته تقویت‌شده با LLMها

در حالی که برنامه‌ریزی وظایف مبتنی بر زبان یکی از برجسته‌ترین پیشرفت‌های اخیر است، دستکاری فیزیکی و کنترل تجسم‌یافته همچنان از چالش‌برانگیزترین بخش‌های رباتیک محاسب می‌شود. روش‌های سنتی برای دستکاری اجسام به دقت عددی، مدل‌سازی هندسی و خط‌لوله‌های ادراکی بسیار تنظیم‌شده مตکی هستند. اما مدل‌های زبانی بزرگ بُعد جدیدی را وارد این حوزه کرده‌اند.

سامانه‌هایی مانند VIMA و LLM-GROP نشان داده‌اند که سرنخ‌های زبانی قادرند استدلال فضایی ربات را غنی کنند. این سامانه‌ها می‌توانند تنها با تکیه بر نشانه‌های سطح‌بالا، وظایفی نظری جای‌گذاری بدون برخورد یا تقليد 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 تعیین‌پذیری 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 LLM را هدف قرار داده است. RT-Grasp با ترکیب خروجی‌های استدلالی و قالب‌های پیش‌بینی عددی ساخت‌یافته، توضیحات معنایی مدل را برای اصلاح پارامترهای دقیق گرفتن اشیا به کار می‌گیرد [5] 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 کم‌نمونه مدل‌های بینایی-زبانی مانند LIV و R3M، عملکرد ربات را در وظایف دستکاری به شکل چشم‌گیری بهبود می‌دهد. این مدل‌ها به جای 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 قابل دستکاری استفاده می‌کنند؛ امری که امکان سازگاری لحظه‌های با دستورات کاربر را فراهم می‌کند.

در مجموع، این پیشرفت‌ها نشان می‌دهند که حوزه دستکاری رباتیک از یک حوزه صرفاً عددی-هندسی به یک رویکرد ترکیبی از معنا، چندوجهی بودن و استدلال ساخت‌یافته حرکت کرده است؛ حرکتی که به ربات‌ها در کی نزدیک‌تر به انسان از تعاملات فیزیکی می‌دهد.

نسل جدید LLMها این قابلیت را به صورت چشم‌گیری گسترش داده است. مدل‌های مدرن با درک معنایی عمیق، نیت ضمنی، وابستگی‌های زمانی و منطق موردنیاز برای انجام وظیفه را شناسایی می‌کنند. برای مثال، پژوهش‌ها درباره اجرایی وظایف بر اساس زبان طبیعی نشان داده‌اند که سامانه‌ها می‌توانند عبارت «من گرسنهام» را با استناد به دانش عمومی خود به وظیفه «غذا تهیه کن» تبدیل کنند. [3] این توانایی، فاصله زیادی با سامانه‌های قدیمی مبتنی بر قواعد دارد و نشان‌دهنده جهت‌گیری به سمت 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 پرسد. این ویژگی با چارچوب‌های گفت‌وگو محور در تعامل انسان-ربات سازگار است، جایی که زبان نه تنها ابزار فرمان‌دادن، بلکه ابزاری برای مذاکره، اصلاح و روشن‌سازی است. [7] 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.

## ۲. پیشرفت‌های برنامه‌ریزی و استدلال مبتنی بر LLM

مرور ادبیات نشان می‌دهد که برنامه‌ریزی وظایف یکی از حوزه‌هایی است که LLMها بیشترین تأثیر را در آن داشته‌اند. پژوهش‌ها ثابت کرده‌اند که LLMها قادرند توالی‌های دقیق و منظم از اقدامات تولید کنند تا دقت و قابلیت اتکای برنامه‌ریزی افزایش یابد. [1] 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، از داشن پیشینی برای هدایت الگوریتم‌های جستجو ببرد.

علاوه بر این، چارچوب‌هایی مانند LLM+P و SayCan تلفیقی از راهنمایی معنایی LLM و سیاست‌های تقویتی را ارائه می‌دهند. در این رویکرد، LLM هدف‌های سطح‌بالا را مشخص می‌کند و یادگیری تقویتی عملی بودن آنها را تأیید می‌کند. [1] سامانه‌های دیگر از ساختارهایی مانند اشاره‌های خط، 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 حوزه تصمیم‌گیری و سازگاری را نیز دربرمی‌گیرد. به عنوان نمونه، LM-Nav ترکیبی از زبان، بینایی و اطلاعات ناوبری را برای دنبال‌کردن دستورهای سطح‌بالا به کار می‌گیرد، بدون آنکه نیاز به داده‌های پیمایش نشانه‌گذاری شده 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

## ۵.۱. معماری‌های مازولار و نسل مبتنی بر بازیابی (RAG)

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، NLU مجموعه‌ای از فرایندها را شامل می‌شود تحلیل و رودی زبانی، استخراج ساختار معنایی، شناسایی عناصر قابل اقدام، رفع ابهام، و تبدیل زبان طبیعی به فرمان‌های اجرایی.

رویکردهای اولیه بهشت 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]

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 ظهور LLM‌های مدرن، این قابلیت‌ها گسترش چشمگیری یافته است. مدل‌های زبانی بزرگ با تکیه بر بازنمایی‌های معنایی قوی و آموزش گسترده روی داده‌های متنه، قادرند روابط پنهان بین دستورها، محیط و نیت ضمنی را استخراج کنند. به عنوان نمونه، پژوهش‌ها نشان می‌دهد که LLM‌ها می‌توانند بدون ورودی بصری نیز یک دستور را به توالی منسجم اقدامات تبدیل کنند و هنگامی که ورودی چندوجهی اضافه شود، دقت و گراندینگ به طور قابل توجهی افزایش می‌یابد. [1] 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] این رفتار نشان‌دهنده گذار از تطبیق الگوهای دستوری به استدلال مبتنی بر دانش عمومی و زمینه محور است.

از سوی دیگر، LLM‌ها در گفت‌وگوی تعاملی برای اصلاح فرمان‌ها نیز مؤثرند. مدل قادر است هنگام نیاز به اطلاعات بیشتر، سؤال پرسید تا دستور نهایی دقیق و قابل اجرا 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]

در مجموع، NLU مبتنی 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.

## ۳. برنامه‌ریزی وظایف با استفاده از مدل‌های زبانی بزرگ

برنامه‌ریزی وظایف یکی از حوزه‌هایی است که LLM‌ها بیشترین تحول را در آن ایجاد کرده‌اند. روش‌های کلاسیک برنامه‌ریزی ربات بر برنامه‌ریزهای

با وجود افزایش توانایی LLM‌ها، این مدل‌ها همچنان ممکن است دچار توهین زایی و استدلال نادرست شوند. در حوزه 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.

سامانه ARRC نمونه بر جسته‌ای از این راهبرد است. در این سامانه، پایگاه دانشی شامل الگوهای حرکتی، راهبردهای اینمنی و قالبهای وظیفه به شکل برداری ذخیره شده و در زمان برنامه‌ریزی، بخش مرتبه از آن بازیابی می‌شود. [6] سپس LLM با اتکا بر این دانش معتبر، برنامه کنشی ساختی‌یافته را تولید می‌کند و در نهایت یک مرحله ارزیابی اینمنی، برنامه را کنترل می‌کند.

به همین ترتیب، سامانه‌های مازولار مانند ProgramPort و MetaMorph از LLM برای تبدیل دستورات زبانی به مازول‌های برنامه‌نویسی یا سیاست‌های سازگار با شکل‌بندی‌های مختلف ربات بهره می‌برند. [1] این رویکردها قابلیت مقایسه‌پذیری، 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 نه تنها به عنوان مفسر زبان، بلکه به عنوان چارچوب‌های سازمان‌دهنده رفتار ربات عمل می‌کنند.

## ۶. تعامل انسان–ربات و یادگیری تعاملی

LLM‌ها نقش اساسی در گسترش ظرفیت‌های تعامل انسان–ربات ایفا می‌کنند. برای نمونه، سامانه ROS-LLM نشان می‌دهد که ربات‌ها می‌توانند وارد گفت‌وگوهای چندمرحله‌ای شوند، بازخورد انسانی را دریافت کنند، رفتارهای سلسه‌مراتبی اجرا کنند و در طول زمان بهبود یابند [7] 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 ذخیره حافظه، یادگیری از تجربه و حتی الگوگری از سبک‌های تعامل انسانی را فراهم می‌کنند. نتیجه این پیشرفت‌ها، حرکت به سمت ربات‌هایی است که تعامل طبیعی‌تر، سازگارتر و مشارکتی‌تر دارند.

در حوزه 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.

در زمینه 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، مجموعه این پیشرفت‌ها نشان می‌دهد که LLM‌ها از یک ابزار زبانی ساده فراتر رفته و به عامل‌های تصمیم‌ساز هوشمند تبدیل شده‌اند.

## ۵. دستکاری و تعامل فیزیکی تقویت‌شده با LLM‌ها

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 عموماً بر بهینه‌سازی عددی، برآورد موقعیت اشیا و سیاست‌های کنترلی داده‌محور تکیه دارند. اما LLM‌ها 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 این که 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 زاویه گرفتن یا موقعیت گیره) تولید کند، ابتدا از یک مرحله استدلال زبانی ساخت‌یافته استفاده می‌شود. این امر باعث می‌شود مدل ابتدا ماهیت شیء (مثلاً «فنجان»، «عینک»، «توب») و اصول صحیح گرفتن آن را توضیح دهد، سپس پارامترهای دقیق را محاسبه کند. [5] این ساختار پل مهمی میان استدلال سطح بالا 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 داده‌های بسیار کمتر، مهارت‌های دستکاری را یاد بگیرند. [1]

نمادین، الگوریتم‌های جست‌وجو یا یادگیری تقویتی مبتنی بر پاداش تکیه داشتند. این روش‌ها در محیط‌های پویا، مبهم یا دارای پیچیدگی معنایی، با محدودیت‌های جدی روبرو بودند.

اما LLM‌ها این محدودیت را برطرف کرده و امکان برنامه‌ریزی سطح بالا درک زمینه‌ای و استدلال معنایی را فراهم کرده‌اند.

پژوهش‌ها نیز بر تصمیم‌گیری میان چند عامل تمرکز دارند؛ حتی بدون مشاهده کامل محیط، توالی اقدامات بلندمدت تولید کنند. [1] یکی از نقاط قوت LLM‌ها، توانایی استخراج دانش عمومی و استفاده از آن در تدوین مراحل منطقی وظایف است. برای مثال، مدل بهطور طبیعی می‌داند که برای تهییه یک نوشیدنی، ابتدا باید یک ظرف پیدا شده و سپس پر شود؛ حتی اگر این مراحل صریحاً در دستور ذکر نشده باشند.

در رویکردهای ترکیبی مانند **SayCan** و **LLM+P**، مدل‌های زبانی با برنامه‌ریزهای رسمی مبتنی بر **PDDL** ادغام می‌شوند. در این مدل‌ها، وظایف سطح بالا را تفسیر کرده و برنامه‌ریزهای سنتی تصمین 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 حوزه ناوبری، سیستم‌هایی مانند **LM-Nav** با ترکیب مدل‌های پیش‌آموزش‌دیده بینایی، مدل‌های زبان و استدلال فضایی، به ربات امکان می‌دهند که بدون نیاز به داده‌های پیمایش برچسب‌خورده، مسیرهای پیچیده را دنبال کنند. [1] مجموع این پیشرفت‌ها نشان می‌دهد که LLM‌ها پایه یک چارچوب برنامه‌ریزی انعطاف‌پذیر، قدرتمند و قابل تعمیم را تشکیل می‌دهند که می‌تواند وظایف پیچیده را تحلیل کند، نیت انسان را استخراج کند و برنامه‌های قابل اعتماد برای محیط‌های پویا تولید کند.

## ۴. استدلال پیچیده و تصمیم‌گیری

مزیت مهم دیگر LLM‌ها توانایی آن‌ها در استدلال سطح بالا و تصمیم‌گیری است. برخلاف سیستم‌های سنتی که تنها به قواعد برنامه‌ریزی شده متکی هستند، LLM‌ها قادرند بین گزینه‌ها انتخاب کنند، پیامدهای احتمالی را بستجند و در طول اجرای وظایف سازگار شوند.

پژوهش‌های مختلف نشان می‌دهد که LLM‌ها می‌توانند از حافظه معنایی بهره ببرند، میان وظایف مختلف تعمیم پیدا کنند و با استفاده از قالب‌های استدلایل ساخت‌یافته، تصمیم‌های منطقی اتخاذ کنند. برای مثال، استفاده از اشاره‌های خطأ به ربات کمک می‌کند که برنامه 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.

وظیفه پیچیده به مجموعه‌ای از مژول‌های قابل استفاده مجدد تقسیم می‌شود؛ مانند مهارت‌ها، سیاست‌های کنترلی، برنامه‌ریزی، مژول‌های ادراک و درخت‌های رفتار.

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

نمونه‌ای از این رویکرد، **PROGRAMPORT** است که زبان طبیعی را به مژول‌های برنامه‌ای برای دستکاری اشیا تبدیل می‌کند. [1] در این سامانه، به جای تولید مستقیم کد، ساختار معنایی دستورها به بلوک‌های عصی قابل استفاده مجدد تبدیل می‌شود که قابلیت تعمیم در محیط‌ها و وظایف جدید را دارد.

رویکرد دیگر، استفاده از **MetaMorph** است؛ سیستمی که با مدل‌های ترنسفورمر، سیاست‌های کنترل سازگار با شکل‌بندن‌های مختلف ربات را یاد می‌گیرد. [1] این امر باعث می‌شود ربات بتواند بدون طراحی مجدد، به سخت‌افزارهای جدید سازگار شود.

سامانه‌های مبتنی بر مدل‌های بینایی-زبانی **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 صرف‌زبانی فراتر برده و آن را به چارچوبی سازمان‌دهنده برای ادراک، کنترل و استدلال ربات تبدیل کرده‌اند؛ چارچوبی که مقیاس‌پذیری و سازگاری چشمگیری فراهم می‌کند.

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

با افزایش توانایی‌های LLM‌ها، یک چالش مهم همچنان برجا می‌ماند. در رباتیک، این خطاهای می‌توانند منجر به اقدامات خطرناک یا غیرممکن شوند. از این‌رو، رویکرد تولید مبتنی بر بازیابی (Retrieval-Augmented Generation) به عنوان یک راهکار مهم برای گراندینگ و کاهش خطا مطرح شده است.

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

- یک پایگاه دانش برداری شامل الگوهای حرکتی، قالب‌های وظیفه، راهبردهای ایمنی و رهنمودهای عملیاتی ذخیره می‌شود.
- مژول بازیابی، بخش‌های مرتبط را بر اساس دستور کاربر و وضعیت محیط استخراج می‌کند.
- LLM با تکیه بر این دانش معتبر، یک برنامه ساخت‌یافته JSON تولید می‌کند.

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

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

## ۶. 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. LLM‌ها ابزارهای بسیار قدرتمندی برای هدایت این تعامل فراهم کرده‌اند.

سامانه **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 تعامل محور 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.

سامانه **ROS-LLM** این مفهوم را گسترش می‌دهد و از LLM برای به کارگیری بازخورد انسان در چرخه 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]

افزون بر این، عامل‌های مولد (Generative Agents) امکان ذخیره تجربه، شکل‌گیری حافظه درازمدت و سازگاری رفتاری را فراهم می‌کنند. این توانایی‌ها ربات‌ها را به موجوداتی بادگیرنده، سازگار و تعاملی‌تر تبدیل می‌کند. در مجموع، رویکردهای تعاملی LLM محور باعث می‌شوند ربات‌ها بتوانند در زمان واقعی یاد بگیرند، از بازخورد انسان بهره ببرند، در محیط‌های ناپایدار یا ناشناخته عملکرد پایه‌تتری داشته باشند.

## ۷. رویکردهای مژولار برای هوش رباتیکی مقیاس‌پذیر

معماری مژولار یکی از اصول مهم طراحی در سامانه‌های رباتیکی است. LLM‌ها این رویکرد را به سطح جدیدی رسانده‌اند. در این معماری، یک

روند مهم دیگر، پیش‌آموزش بینایی- زبانی در مقیاس بزرگ است؛ مدل‌هایی مانند **LIV** و **R3M** که روی ویدئوهای انسانی آموزش دیده‌اند، ویژگی‌هایی مانند قابلیت‌های اشیا، بافت‌ها، روابط عملکردی و نشانه‌های فضایی را به خوبی 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 بر LLM‌ها تنها هندسه محیط را توصیف نمی‌کند، بلکه معنای صحنه را نیز استخراج می‌کند. توانایی‌ای که برای تعامل در محیط‌های واقعی، پویا و غیرساخت‌یافته ضروری است.

#### ۱۰. کنترل و اجرا در رباتیک مبتنی بر LLM

کنترل - یعنی ترجمه هدف به حرکت فیزیکی - جایی است که استدلال سطح بالای LLM با محدودیت‌های سخت‌افزار ربات تلاقي می‌کند.

از آنجا که LLM‌ها خروجی‌هایی با سطح انتزاع بالا تولید می‌کنند، لازم است این خروجی‌ها با مژول‌های کنترل پایین سطح و برنامه‌ریزی‌های حرکتی ادغام شوند.

در بسیاری از سامانه‌ها، خروجی LLM به صورت ساخت‌یافته ارائه می‌شود، مانند توالی‌های JSON (در ARRC)، کدهای Python (در ROS-LLM)، درخت‌های رفتاری XML، برنامه‌های شبکه (در Cap) یا هدف‌های میانی برای یادگیری تقویتی (در SayCan). [7]

در **ROS-LLM**، این خروجی‌ها مستقیماً به سرویس‌ها و اکشن‌های ROS تبدیل می‌شوند و به این ترتیب مدل می‌تواند با موتورهای مسیریابی، حل کننده‌های سینماتیک و کتابخانه‌هایی مانند MoveIt به طور کامل یکپارچه شود.

در روش‌هایی مانند SayCan، LLM وظایف سطح بالا را پیشنهاد می‌دهد و یادگیری تقویتی بررسی می‌کند که آیا این وظایف قابل اجرا هستند یا خیر؟ [1] این ساختار از امکان تولید دستورهای غیر عملی 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.

#### ۱۱. اجرای ایمن (Execution Safety-Constrained)

ایمنی یکی از مهم‌ترین دغدغه‌ها در رباتیک است، بهویژه زمانی که ربات در کنار انسان‌ها کار می‌کند یا با اشیای حساس سروکار دارد.

در نهایت، یک لایه «دروازه ایمنی» برنامه را پیش از اجرا ارزیابی می‌کند.

این pipeline باعث می‌شود احتمال تولید اقدامات اشتباه یا ناسازگار از سوی LLM به شکل چشم‌گیری کاهش یابد.

علاوه بر این، RAG نیاز به آموزش مجدد مدل را کاهش می‌دهد؛ زیرا افزودن مهارت‌های جدید تنها به اضافه کردن 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 داده عمق، موقعیت دقیق اشیا استخراج شده و به LLM داده می‌شود تا برنامه‌ریزی با شناخت واقعی تری از محیط انجام شود. [6] در مجموع، RAG سنتر قدرتمندی میان قطعیت و ایمنی در رباتیک کلاسیک و انعطاف و استدلال LLM‌ها ایجاد می‌کند. نتیجه، معماری‌ای است که برای برنامه‌ریزی خودکار، امن و مقیاس‌پذیر در دنیای واقعی بسیار مناسب است.

### ۹. 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 به صورت قابل اعتماد پردازش کند. LLM‌ها و مدل‌های بینایی - زبانی (VLMS) در این زمینه پیشرفت‌های قابل توجهی ایجاد کرده‌اند.

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

سامانه **LM-Nav** نیز با ترکیب CLIP و نقشه‌برداری embedding مبتنی بر ViNG، به ربات امکان می‌دهد که صرفاً بر اساس زبان طبیعی و بدون نیاز به داده پیمایش برچسب‌خورده، به نقاط خاصی در محیط ناوبری کند. [6]

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

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

به طور 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 بفهمند.

## ۱. تعریف وظایف از طریق مکالمه

سامانه‌هایی که از LLM استفاده می‌کنند، می‌توانند دستورهای پیجیده را از جملات کاملاً طبیعی استخراج کنند. برای مثال، کاربر می‌تواند بگوید: «میز را تمیز کن و ظرف‌ها را داخل سینک بگذار.» LLM این جمله را به مجموعه‌ای از وظایف شامل حرکت، گرفتن، جابه‌جایی و رهاسازی ترجمه می‌کند. [1]

## ۲. 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] این فرآیند مشابه روش‌های آموزشی انسانی است؛ مشاهده رفتار، ارائه اصلاح و یادگیری از تجربه.

## ۳. درک ضمنی و commonsense

سامانه‌های NLexer قادرند از جملات غیرمستقیم مانند «تشنه هستم» معنای درخواست آب را استنباط کنند. [3] این توانایی برای تعامل طبیعی و انسانی کاملاً ضروری است.

## ۴. 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.

## ۵. عامل‌های مشارکتی و حافظه‌دار

پژوهش‌های جدید درباره «عامل‌های مولد» نشان می‌دهند که ربات می‌تواند حافظه بلندمدت شکل دهد، الگوهای رفتاری پایدار ایجاد کند و تجربه گذشته را برای بهبود تعامل استفاده کند. [1] در نتیجه، HRI مبتنی بر LLM به ربات‌هایی منجر می‌شود که قابل اعتمادتر، قابل تعاملی‌تر، قابل سازگاری‌تر و برای کاربران غیرمتخصص بسیار قابل استفاده‌تر هستند.

## ۱۳. محدودیتها، چالش‌ها و مسیرهای آینده

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

## ۱. 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.

مدل‌های زبانی بزرگ، با وجود توانایی‌های چشمگیرشان، ماهیتی احتمالی دارند و ممکن است خروجی‌هایی تولید کنند که از نظر فیزیکی نادرست یا خطروناک باشند. به همین دلیل، بسیاری از سامانه‌ها لایه‌های اضافی برای تضمین اینمی در نظر گرفته‌اند.

در سامانه ARRC، هر اقدام پیشنهادی از سوی LLM باید از یک مجموعه چندلایه از «دوازه‌های اینمی» زیر عبور کند: [6]

اعتبارسنجی محدوده کاری (Workspace Validation): محل‌های هدف بررسی می‌شوند تا خارج از محدوده قابل دسترس ربات نباشند.

محدودیت‌های سرعت و شتاب: حرکات از سقف‌های تعیین شده 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.

حالاتی عقب‌نشینی اضطراری: در صورت بروز خطاها مکرر، ربات به حالت امن بازمی‌گردد.

تعداد تکرار محدود برای تلاش مجدد.

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 لایه اینمی نظری مکمل ارائه می‌دهد. این مدل با برآورد عدم‌قطعیت خروجی، از انجام اقداماتی جلوگیری می‌کند که مدل نسبت به آن‌ها اعتماد کافی ندارد. [1] سامانه ROS-LLM نیز از «پرچم خطای» در ساختار MDP استفاده می‌کند تا ربات بتواند شکست اجرای وظایف را تشخیص داده و درخواست بازخورد یا اصلاح از کاربر داشته باشد. [7]

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

## ۱۴. تعامل انسان–ربات و هوش مشارکتی

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

## ۲. دقت عددی و محدودیت‌های فیزیکی

مدل‌های زبانی در استدلال و معنا بسیار توانمندند، اما در محاسبات دقیق، هندسه سه‌بعدی، پیش‌بینی پیوسته حرکت ضعیف‌تر عمل می‌کنند. از همین رو روش‌هایی مانند **RT-Grasp** ضروری‌اند.<sup>[5]</sup>

## ۳. نیاز به داده و مشکلات انتقال دامنه

تنظیم تخصصی مدل‌ها برای کارهای صنعتی نیازمند منابع محاسباتی و داده زیاد است، حتی با وجود روش‌های بهینه مانند **QLORA**.<sup>[4]</sup>

## ۴. ایمنی و اعتبارسنجی

از آنجا که LLM‌ها ذاتاً احتمالی هستند، نمی‌توان ایمنی کامل را تنها با اتکا به آن‌ها تضمین کرد. به همین دلیل، اجرای امن نیازمند محدود کننده‌های سخت‌افزاری و اعتبارسنجی نمایدین است.

## ۵. 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.

## ۶. 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.

## ۷. وظایف بلندمدت

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

## ۸. ملاحظات اخلاقی و اجتماعی

با افزایش خودمختاری ربات‌ها، نگرانی‌هایی مانند مسئولیت، حریم خصوصی، اثرات شغلی و امنیت انسان‌ها اهمیت بیشتری می‌یابد.

## ۱۴. نتیجه‌گیری

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 که LLM‌ها دیگر صرفاً نقش مکمل ندارند، بلکه به هسته 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<sup>[2]</sup> 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 زمینه برنامه‌ریزی، مدل‌های زبانی اکنون قادرند برنامه‌های بلندمدت، زمینه‌محور و قابل اجرا تولید کنند و آن‌ها را با کمک مکانیزم‌های گراندینگ و برنامه‌ریزهای کلاسیک تکمیل نمایند.<sup>[1]</sup> ادغام ورودی‌های چندوجهی و دانش پیشین نشان داده است که چشم‌انداز LLM‌ها به عنوان «برنامه‌ریزهای عمومی» روز بروز قابل دسترس‌تر می‌شود.

در حوزه دستکاری، روش‌هایی مانند **RT-Grasp**، **LLM-GROP**، **VIMA** و **MetaMorph** 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 مازلولار مانند **PROGRAMPORT** و **MetaMorph** نشان می‌دهند که LLM‌ها قادرند ساختارهای رفتاری مقیاس‌پذیر ایجاد کنند. در کنار آن، روش‌های **RAG** مانند **ARRC**<sup>[6]</sup> ثابت کردند که با تکیه بر پایگاه دانش واقعی، می‌توان خطأ و توهمندی مدل را کنترل کرده و برنامه‌ریزی قابل اعتماد ارائه داد.

در حوزه تعامل انسان–ربات، سامانه‌هایی نظیر **ROS-LLM**<sup>[7]</sup> نشان داده‌اند که زبان طبیعی می‌تواند واسطه اصلی تعامل باشد. این امکان به کاربران غیرمتخصص اجازه می‌دهد ربات‌ها را اصلاح، هدایت یا برنامه‌ریزی کنند - آن‌ها هم بدون نیاز به دانش فنی.

با وجود این پیشرفت‌ها، چالش‌هایی همچون دقت عددی، ایمنی، تعمیم، پردازش زمان واقعی و مسائل اخلاقی همچنان پا بر جاست. LLM‌ها ممکن است در برخی شرایط رفتار غیرقابل اعتماد 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 مطالعه 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

[8] F. Zeng, W. Gan, Y. Wang, N. Liu and P. S. Yu, "Large Language Models for Robotics: A Survey," p. 19, 2023.

نسل جدیدی از هوش تجسمی‌یافته باز می‌کند. ربات‌هایی که نه تنها 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 به تدریج 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.

#### منابع :

- [1] W. Jiaqi, W. Zihao, L. Yiwei, J. Hanqi and S. Peng, "Large Language Models for Robotics: Opportunities, Challenges, and Perspectives," p. 18, 2021.
- [2] B. Gianluca, . D. Danilo, M. Enrico and O. Francesco, "Parsing Natural Language Sentences into Robot," p. 4, 2019.
- [3] L. Rui, G. Yibei, J. Runxiang and Z. Xiaoli, "A Review of Natural-Language-Instructed Robot," p. 42, 2024.
- [4] A. Benjamin, K. Urs, T. Aleksandar, K. Darko and H. Andreas, "Domain-Specific Fine-Tuning of Large Language," p. 5, 2021.
- [5] X. Jinxuan, J. Shiyu, L. Yutian, Z. Yuqian and Z. Liangjun, "RT-Grasp: Reasoning Tuning Robotic Grasping via," p. 8, 2024.
- [6] V. Eugene, M. Ammar Jaleel, . R. Salim and C. Robin, "ARRC: Advanced Reasoning Robot Control—Knowledge-Driven Autonomous Manipulation Using Retrieval-Augmented Generation," p. 8, 2025.
- [7] E. M. Christopher, W. Yuhui, Hongzhan, G. Antoine, J. Gonzalez-Billandon, M. Zimmer and J. Wang, "ROS-LLM: A ROS framework for embodied AI with task feedback and structured reasoning," p. 26, 2024.