

# گزارش فاز یک (سامانه کنترل ربات مبتنی بر مدل‌های زبانی بزرگ)

محمد مهدی اسدی، امیرحسین سلیمانی

استاد راهنما: دکتر مازیار پالهنگ

نسخه ۱.۰

آذر ۱۴۰۴

## چکیده

با سازمان‌دهی این یافته‌ها در حوزه‌های برنامه‌ریزی، استدلال، دستکاری، ادراک، معماری‌های کنترلی، یک نگاه یکپارچه از وضعیت کنونی هوش تجسم‌یافته<sup>۷</sup> مبتنی بر مدل زبانی بزرگ 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 را بحث می‌کنیم.

مدل‌های زبانی بزرگ<sup>۱</sup> 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<sup>۲</sup> استدلال کنند و مهارت‌های دستکاری<sup>۳</sup> اشیاء را در محیط‌های پویا اجرا کنند. با حرکت رباتیک به سمت سطوح بالاتری از خودمختاری و هوش تجسم‌یافته، سامانه‌های مبتنی بر مدل‌های زبانی بزرگ 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 مدل‌های زبانی بزرگ ارائه می‌دهد. در این مقاله بررسی می‌کنیم که مدل‌های زبانی بزرگ چگونه برنامه‌ریزی وظایف مبتنی بر زبان طبیعی، تصمیم‌گیری تعاملی و ساختارهای برنامه‌نویسی ماژولار برای ربات‌ها را تقویت می‌کنند. همچنین به روش‌هایی مانند برنامه‌ریزی مبتنی بر ثابت‌سازی<sup>۴</sup> [1] و تحلیل نحوی مبتنی بر هستی‌شناسی<sup>۵</sup> [2] و چارچوب‌های نظام‌مند آموزش ربات از طریق زبان انسان [3] می‌پردازیم.

ما علاوه بر این، راهبردهای تنظیم تخصصی مدل‌های زبانی برای برنامه‌نویسی صنعتی ربات‌ها [4] و روش‌های دستکاری مبتنی بر استدلال چندوجهی [5] را بررسی می‌کنیم. همچنین پیشرفت‌های اخیر در سامانه‌های کنترل ربات مبتنی بر بازایی اطلاعات<sup>۶</sup> [6] و چارچوب‌های ساخت‌یافته مبتنی بر ROS [7] را مرور می‌کنیم.

## واژگان کلیدی

مدل‌های زبانی بزرگ، رباتیک، هوش تجسم‌یافته، درک زبان طبیعی، برنامه‌ریزی وظایف، دستکاری اشیاء، استدلال چندوجهی، تولید مبتنی بر بازایی اطلاعات، تعامل انسان-ربات، سامانه‌های کنترل ربات

## ۱. مقدمه

### ۱.۱. پیش‌زمینه و انگیزه

یکپارچگی مدل‌های زبانی بزرگ با رباتیک، نقطه عطفی در نحوه فهم، استدلال و تعامل ربات‌ها با جهان فیزیکی محسوب می‌شود. در حالی که رباتیک سنتی به برنامه‌نویسی ساخت‌یافته<sup>۸</sup>، 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

<sup>5</sup> ontology-based action parsing

<sup>6</sup> retrieval-augmented robotic

<sup>7</sup> embodied intelligence

<sup>8</sup> structured

<sup>1</sup> Large language model

<sup>2</sup> Multimodal inputs

<sup>3</sup> manipulation

<sup>4</sup> Grounded planning

نشان‌دهنده گذار به سمت «هوش تجسم‌یافته» است. جایی که ربات‌ها با استفاده از بازنمایی‌های شهودی و غنی‌شده از نظر شناختی یاد می‌گیرند، عمل می‌کنند و همکاری می‌کنند. [8]

در گذشته، سامانه‌های رباتیکی نیازمند برنامه‌نویسی صریح و دانش دامنه‌ای گسترده بودند. اما مدل‌های زبانی بزرگ همچون موتورهای استدلالی عمومی<sup>9</sup> عمل می‌کنند؛ که قادرند معنای زمینه‌ای<sup>10</sup> را استخراج کنند، برنامه انجام وظایف را تولید کنند و این برنامه‌ها را به دستورهای قابل‌اجرا تبدیل کنند. این قابلیت امکان می‌دهد وظایف پیچیده -از ناوبری<sup>11</sup> تا دستکاری- به صورت مکالمه‌ای تعریف شوند، نه با کدنویسی تخصصی یا الگوهای نمادین پیچیده. برای مثال، سامانه‌های برنامه‌ریزی مبتنی بر مدل‌های زبانی بزرگ 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]

مدل‌های چندوجهی نیز با ترکیب ورودی‌های بصری، زبانی و زمینه‌ای، بازنمایی‌های واقعیت محور تولید می‌کنند که پل میان ارتباط انسانی و کنترل ربات است. این مدل‌ها توانایی استدلال پیچیده، درک قابلیت‌های اشیاء و آگاهی فضایی را فراهم می‌کنند؛ که برای عملکرد قابل‌اعتماد در محیط‌های واقعی ضروری است. چنین توانایی‌هایی نشان می‌دهد که استدلال مبتنی بر مدل‌های زبانی بزرگ عنصر مرکزی در حرکت به سوی خودمختاری، سازگاری و هم‌راستایی بیشتر با انسان است.

## ۲.۱. از زبان تا عمل: نقش درک زبان طبیعی

یکی از کارکردهای اساسی مدل‌های زبانی بزرگ در رباتیک، تبدیل زبان انسان به بازنمایی<sup>۱۲</sup>‌های 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] 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 جدید مدل‌های زبانی بزرگ 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 اجرای 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.

مدل‌های زبانی بزرگ 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 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.

## ۳.۱. پیشرفت‌های برنامه‌ریزی و استدلال مبتنی بر مدل‌های زبانی بزرگ

مرور ادبیات نشان می‌دهد که برنامه‌ریزی وظایف، یکی از حوزه‌هایی است که مدل‌های زبانی بزرگ بیشترین تأثیر را در آن داشته‌اند. پژوهش‌ها ثابت کرده‌اند که مدل‌های زبانی بزرگ 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] این هم‌افزایی به ربات اجازه می‌دهد که دستورهای بلندمدت را به گام‌های ساخت‌یافته تقسیم کند، پیش‌شرط‌ها<sup>۱۳</sup> و روابط فضایی<sup>۱۴</sup> را تحلیل کند، از مکانیزم‌های ثابت‌سازی 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 مدل‌های زبانی بزرگ و سیاست‌های تقویتی را ارائه می‌دهند. در این رویکرد، مدل‌های زبانی بزرگ هدف‌های سطح‌بالا را مشخص می‌کند و یادگیری تقویتی عملی‌بودن آنها را تأیید می‌کند. [1] سامانه‌های دیگر از ساختارهایی مانند اشاره‌های خطا<sup>۱۵</sup>، 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

<sup>13</sup> preconditions

<sup>14</sup> spatial relationships

<sup>15</sup> error cues

<sup>9</sup> generalized reasoning engines

<sup>10</sup> contextual meaning

<sup>11</sup> Navigation

<sup>12</sup> representations

و اطلاعات ناوبری را برای دنبال کردن دستورهای سطح بالا به کار می‌گیرد. بدون آنکه نیاز به داده‌های پیمایش نشانه‌گذاری شده<sup>۱۶</sup> 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 مدل‌های زبانی بزرگ از یک «پردازشگر زبان» فراتر رفته و به «عامل‌های استدلالی فعال» تبدیل شده‌اند که در هسته معماری کنترل ربات قرار می‌گیرند.

#### ۴.۱. دستکاری، ادراک و مهارت‌های تجسم‌یافته تقویت‌شده با مدل‌های زبانی بزرگ

در حالی که 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 دستکاری اجسام به دقت عددی، مدل‌سازی هندسی و خط لوله‌های ادراکی بسیار تنظیم‌شده متکی هستند. اما مدل‌های زبانی بزرگ بعد جدیدی را وارد این حوزه کرده‌اند.

سامانه‌هایی مانند **LLM-GROP** و **VIMA** نشان داده‌اند که سرخ‌های زبانی قادرند استدلال فضایی ربات را غنی کنند. این سامانه‌ها می‌توانند تنها با تکیه بر نشانه‌های سطح بالا، وظایفی نظیر جای‌گذاری بدون برخورد یا تقلید چندوجهی را انجام دهند. [1] چارچوب‌های چندوجهی با یکپارچه‌سازی بینایی، زبان و حرکت، این امکان را می‌دهند که 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»<sup>۱۷</sup> تکیه دارند، یکی از موانع اصلی در دستکاری، یعنی دقت عددی را هدف قرار داده است. **RT-Grasp** با ترکیب خروجی‌های استدلالی مدل‌های زبانی بزرگ و قالب‌های پیش‌بینی عددی ساخت‌یافته، توضیحات معنایی مدل را برای اصلاح پارامترهای دقیق گرفتن اشیاء به کار می‌گیرد. [5] این کار پلی میان فهم زبانی-مفهومی و کنترل هندسی ایجاد می‌کند؛ پلی که پیش از این بیشتر بر دوش روش‌های تحلیلی یا یادگیری مبتنی بر داده بود.

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

با وجود افزایش توانایی مدل‌های زبانی بزرگ، این مدل‌ها همچنان ممکن است دچار توهم زایی و استدلال نادرست شوند. در حوزه رباتیک، چنین خطاهایی می‌توانند خطرناک باشند. از همین رو، تولید مبتنی بر بازیابی

**Retrieval-augmented generation (RAG)** به یک رویکرد مهم برای ثابت سازی و اعتبارسنجی تبدیل شده است.

سامانه **ARRC** نمونه برجسته‌ای از این راهبرد است. در این سامانه، پایگاه دانشی شامل الگوهای حرکتی، راهبردهای ایمنی و قالب‌های وظیفه به شکل برداری ذخیره شده و در زمان برنامه‌ریزی، بخش مرتبط از آن بازیابی می‌شود. [6] سپس مدل‌های زبانی بزرگ با اتکا بر این دانش معتبر، برنامه کنشی ساخت‌یافته را تولید می‌کند و در نهایت یک مرحله ارزیابی ایمنی، برنامه را کنترل می‌کند.

به همین ترتیب، سامانه‌های ماژولار مانند **ProgramPort** و **MetaMorph** از مدل‌های زبانی بزرگ 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 مدل‌های زبانی بزرگ نه تنها به عنوان مفسر زبان، بلکه به عنوان چارچوب‌های سازمان‌دهنده رفتار ربات عمل می‌کنند.

#### ۶.۱. تعامل انسان-ربات و یادگیری تعاملی

مدل‌های زبانی بزرگ 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 سلسله‌مراتبی 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»<sup>۱۸</sup> امکان ذخیره حافظه، یادگیری از تجربه و حتی الگوگیری از سبک‌های تعامل انسانی را فراهم می‌کنند. نتیجه این پیشرفت‌ها، حرکت به سمت ربات‌هایی است که تعامل طبیعی‌تر، سازگارتر و مشارکتی‌تر دارند.

<sup>18</sup> generative agents

<sup>16</sup> trajectory annotations

<sup>17</sup> reasoning-tuned

## ۲. درک زبان طبیعی برای سامانه‌های رباتیکی

### درک زبان طبیعی (NLU) Natural language understanding

سنگ‌بنای توانایی ربات‌های مبتنی بر مدل‌های زبانی بزرگ برای تفسیر نیت انسان و تبدیل آن به رفتارهای قابل اجرا است. در ادبیات پژوهش، NLU مجموعه‌ای از فرایندها را شامل می‌شود. تحلیل ورودی زبانی، استخراج ساختار معنایی، شناسایی عناصر قابل اقدام، رفع ابهام و تبدیل زبان طبیعی به فرمان‌های اجرایی.

رویکردهای اولیه به شدت 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 مدل‌های زبانی بزرگ 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. به عنوان نمونه پژوهش‌ها نشان می‌دهد که مدل‌های زبانی بزرگ می‌توانند بدون ورودی بصری نیز یک دستور را به توالی منسجم اقدامات تبدیل کنند و هنگامی که ورودی چندوجهی اضافه شود، دقت و ثابت‌سازی 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 زمینه‌محور است.

از سوی دیگر، مدل‌های زبانی بزرگ در گفت‌وگوی تعاملی برای اصلاح فرمان‌ها نیز مؤثرند. مدل قادر است هنگام نیاز به اطلاعات بیشتر، سؤال بپرسد 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]

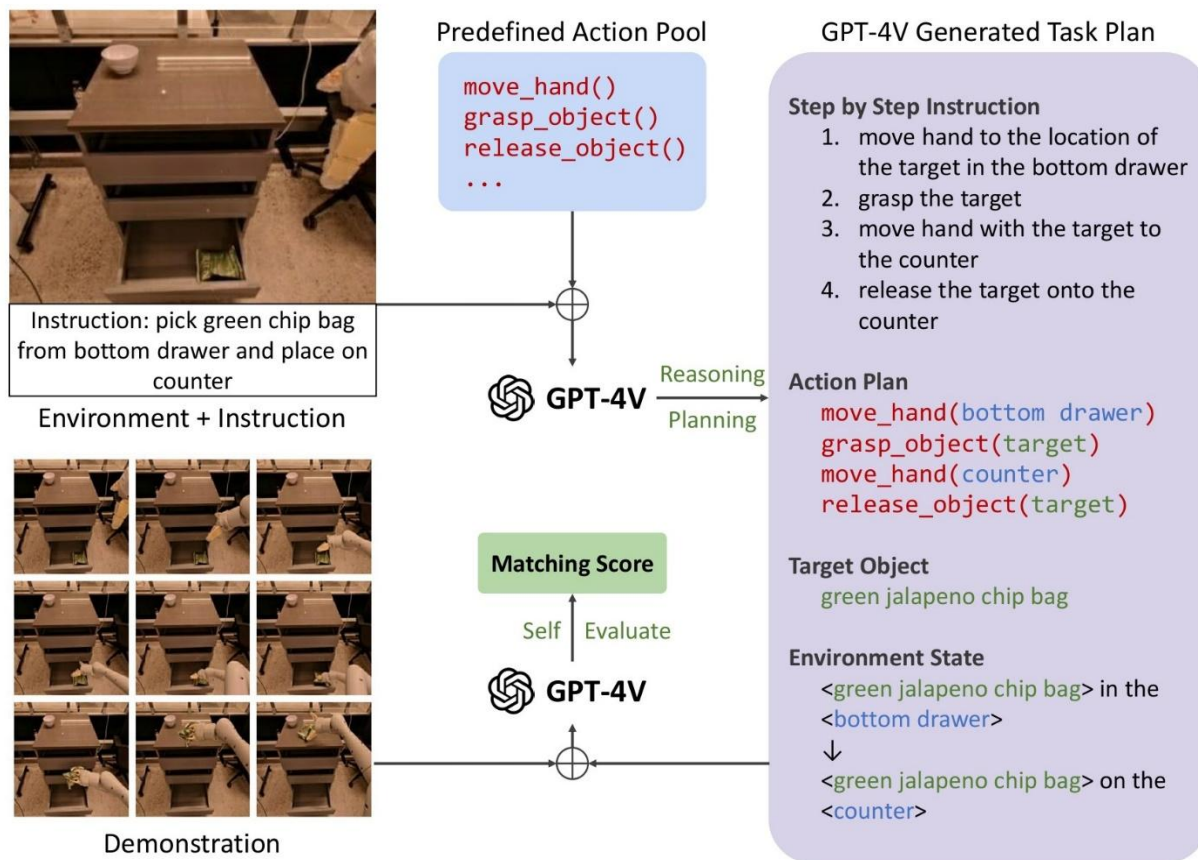
در مجموع، NLU مبتنی بر مدل‌های زبانی بزرگ زیربنای شناختی سامانه‌های رباتیکی را شکل می‌دهد و آنها را قادر می‌سازد تا دستورهای پیچیده انسانی را تفسیر کرده و مطابق آن عمل کنند.

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

برنامه‌ریزی وظایف یکی از حوزه‌هایی است که مدل‌های زبانی بزرگ 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. اما مدل‌های زبانی بزرگ این محدودیت را برطرف کرده و امکان برنامه‌ریزی سطح بالا با درک زمینه‌ای و استدلال معنایی را فراهم کرده‌اند.

پژوهش‌ها نشان می‌دهد که مدل‌های زبانی بزرگ قادرند از روی دستورهای زبانی، حتی بدون مشاهده کامل محیط، توالی اقدامات بلندمدت تولید کنند. [1] یکی از نقاط قوت مدل‌های زبانی بزرگ، توانایی استخراج دانش عمومی و استفاده از آن در تدوین مراحل منطقی وظایف است. برای مثال، مدل به طور 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.

در رویکردهای ترکیبی مانند LLM+P و SayCan، مدل‌های زبانی با برنامه‌ریزهای رسمی مبتنی بر PDDL ادغام می‌شوند. در این مدل‌ها، مدل زبانی بزرگ وظایف سطح بالا را تفسیر کرده و برنامه‌ریزهای سنتی تضمین می‌کنند که این عملیات با محدودیت‌های جهان واقعی سازگار باشد. [1] علاوه بر این، روش‌هایی مانند رمزگشایی‌های واقعیت‌محور<sup>۱۹</sup> با هم‌تراز کردن خروجی مدل زبانی بزرگ 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 کند. (تصویر 1) [1] مجموع این پیشرفت‌ها نشان می‌دهد که مدل‌های زبانی بزرگ 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.



تصویر 1. چارچوب روش پیشنهادی برای برنامه‌ریزی وظایف تجسیدی با GPT-4V. فریم ابتدایی و متن دستورها به مدل داده می‌شود. GPT-4V دستور را به مجموعه‌ای از گام‌های عملیاتی تجزیه کرده و از میان یک مخزن از اعمال، نمایه‌های مناسب را برمی‌گزیند. هم‌زمان، شیء هدف و تغییرات محیطی قبل و بعد از انجام دستور در تصاویر تحلیل می‌شود. در پایان، GPT-4V طرح تولیدشده را با طرح مرجع مقایسه کرده و امتیاز می‌دهد.

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

#### 5. دستکاری و تعامل فیزیکی تقویت‌شده با LLMها

وظایف دستکاری -مانند گرفتن، جابه‌جایی و چیدمان اشیاء- 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 اشیاء و سیاست‌های کنترلی داده‌محور تکیه دارند. اما مدل‌های زبانی بزرگ 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

#### 4. استدلال پیچیده و تصمیم‌گیری

پژوهش‌های مختلف نشان می‌دهد که مدل‌های زبانی بزرگ 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 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.

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.

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»، «توپ») 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]

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 کنند.

مجموع 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. مدل‌های زبانی بزرگ 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 مدل زبانی بزرگ 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 (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 مدل زبانی بزرگ برای به‌کارگیری 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.
- برنامه را بازسازی یا بهبود دهد. [7]

افزون بر این، عامل‌های مولد امکان ذخیره تجربه، شکل‌گیری حافظه درازمدت و سازگاری رفتاری را فراهم می‌کنند. این توانایی‌ها ربات‌ها را به موجوداتی یادگیرنده، سازگار و تعاملی‌تر تبدیل می‌کند.

در مجموع، رویکردهای تعاملی مدل زبانی بزرگ محور باعث می‌شوند ربات‌ها بتوانند در زمان واقعی یاد بگیرند، از بازخورد انسان بهره ببرند، در محیط‌های ناپایدار یا ناشناخته عملکرد باثبات‌تری داشته باشند.

## ۷. رویکردهای ماژولار برای هوش رباتیکی مقیاس پذیر

معماری ماژولار یکی از اصول مهم طراحی در سامانه‌های رباتیکی است. مدل‌های زبانی بزرگ این رویکرد را به سطح جدیدی رسانده‌اند. در این معماری، یک وظیفه پیچیده به مجموعه‌ای از ماژول‌های قابل‌استفاده مجدد تقسیم می‌شود. مانند مهارت‌ها، سیاست‌های کنترلی، برنامه‌ریزها، ماژول‌های ادراک و درخت‌های رفتار.

مدل‌های زبانی بزرگ 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 قابل‌استفاده 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.

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]



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

با افزایش توانایی‌های مدل‌های زبانی بزرگ، یک چالش مهم همچنان برجا می‌ماند. در رباتیک، این خطاها می‌توانند منجر به اقدامات خطرناک یا غیرممکن شوند. از این رو، رویکرد تولید مبتنی بر بازیابی به عنوان یک راهکار مهم برای ثابت‌سازی و کاهش خطا مطرح شده است.

سامانه **ARRC** نمونه‌ای برجسته 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.
- ماژول بازیابی، بخش‌های مرتبط را بر اساس دستور کاربر و وضعیت محیط استخراج می‌کند.
- مدل زبانی بزرگ با تکیه بر این دانش معتبر، یک برنامه ساخت‌یافته **JSON** تولید می‌کند.
- در نهایت، یک لایه «دروازه ایمنی» برنامه را پیش از اجرا ارزیابی می‌کند.

این خط لوله باعث می‌شود احتمال تولید اقدامات اشتباه یا ناسازگار از سوی مدل زبانی بزرگ 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 اشیاء استخراج شده و به مدل زبانی بزرگ داده می‌شود تا برنامه‌ریزی با شناخت واقعی‌تری از محیط انجام شود. [6] در مجموع، **RAG** آمیزش قدرتمندی میان قطعیت و ایمنی در رباتیک کلاسیک و انعطاف و استدلال مدل زبانی بزرگ ایجاد می‌کند.

## ۹. 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. مدل‌های زبانی بزرگ و مدل‌های بینایی-زبانی<sup>۲۰</sup> 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 کفش چه چیزی است؟» یا «فنجان سبز کجاست؟» پاسخ مناسب بدهد.

سامانه **LM-Nav** نیز با ترکیب جایگذاری‌های **CLIP** و نقشه‌برداری 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 اشیاء را استخراج می‌کند و این اطلاعات برای اجرای صحیح وظایف **pick & place** ضروری است. [6]

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.

این بازنمایی‌ها موجب افزایش کارایی نمونه<sup>۲۱</sup> و تعمیم بدون آموزش اضافی در ربات‌های واقعی می‌شود.

در نهایت، می‌توان گفت ادراک مبتنی بر مدل‌های زبانی بزرگ تنها هندسه محیط را توصیف نمی‌کند، بلکه معنای صحنه را نیز استخراج می‌کند. توانایی‌ای که برای تعامل در محیط‌های واقعی، پویا و غیرساخت‌یافته<sup>۲۲</sup> ضروری است.

### ۱۰. کنترل و اجرا در رباتیک مبتنی بر مدل زبانی بزرگ

کنترل -یعنی ترجمه هدف به حرکت فیزیکی- جایی است که استدلال سطح‌بالای مدل زبانی بزرگ با محدودیت‌های سخت‌افزار ربات تلاقی می‌کند.

از آنجا که مدل‌های زبانی بزرگ 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 مدل زبانی بزرگ به صورت ساخت‌یافته ارائه می‌شود، مانند توالی‌های **JSON** (در **ARRC**)، کدهای **Python** (در **ROS-LLM**)، درخت‌های رفتاری **XML**، برنامه‌های شبه‌کد (در **Cap**) یا هدف‌های میانی برای یادگیری تقویتی (در **SayCan**). (تصویر 2)

<sup>22</sup> unstructured

<sup>20</sup> Vision Language Models

<sup>21</sup> sample efficiency



تصویر 2. معماری در سطح بالا: مازول ادراک، مشاهدات شیء محور تولید می کند. برنامه ریز RAG دانشی مربوط به کار را بازایی کرده و یک طرح JSON می سازد. اجراکننده نیز با بررسی های ایمنی، دستورها را از طریق XArm SDK اعتبارسنجی و اجرا می کند.

خطا» در ساختار MDP استفاده می کند تا ربات بتواند شکست اجرای وظایف را تشخیص داده و درخواست بازخورد یا اصلاح از کاربر داشته باشد. [7]

## ۱۲. تعامل انسان-ربات و هوش مشارکتی

تعامل انسان-ربات (HRI) Human-robot interaction یکی از حوزه هایی است که مدل های زبانی بزرگ بیشترین اثرگذاری را در آن داشته اند. مشکل قدیمی رباتیک این بود که کاربران غیرمتخصص نمی توانستند ربات ها را به آسانی کنترل یا برنامه ریزی کنند، اما مدل های زبانی این شکاف را پر کرده اند.

### ۱. تعریف وظایف از طریق مکالمه

در ROS-LLM، این خروجی ها مستقیماً به سرویس ها و اکشن های ROS تبدیل می شوند و به این ترتیب مدل می تواند با موتورهای مسیریابی، حل کننده های سینماتیک و کتابخانه هایی مانند MoveIt به طور کامل یکپارچه شود. [7]

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

ایمنی یکی از مهم ترین دغدغه ها در رباتیک است، به ویژه زمانی که ربات در کنار انسان ها کار می کند یا با اشیای حساس سروکار دارد.

مدل های زبانی بزرگ، با وجود توانایی های چشمگیرشان، ماهیتی احتمالی دارند و ممکن است خروجی هایی تولید کنند که از نظر فیزیکی نادرست یا خطرناک باشند. به همین دلیل، بسیاری از سامانه ها لایه های اضافی برای تضمین ایمنی در نظر گرفته اند.

در سامانه ARRC، هر اقدام پیشنهادی از سوی مدل زبانی بزرگ باید از یک مجموعه چندلایه از «دروازه های ایمنی»<sup>۲۳</sup> زیر عبور کند. شامل: [6]

اعتبارسنجی محدوده کاری (Workspace Validation): محل های هدف بررسی می شوند تا خارج از محدوده قابل دسترس ربات نباشند.

محدودیت های سرعت و شتاب: حرکات از سقف های تعیین شده فراتر نمی روند.

نظارت بر گشتاور و نیروی گریپر: در صورت افزایش نیروی نامعمول، عملیات قطع می شود.

زمان سنج برای هر مرحله: اگر یک حرکت بیش از حد طولانی شود، عملیات متوقف می شود.

حالت های عقب نشینی اضطراری: در صورت بروز خطاهای مکرر، ربات به حالت امن بازمی گردد.

تعداد تکرار محدود برای تلاش مجدد.

این لایه ها تضمین می کنند که حتی اگر مدل زبانی بزرگ برنامه ای غیرایمن یا ناسازگار تولید کند، اجرا هرگز بدون کنترل و اعتبارسنجی صورت نمی گیرد.

رویکرد KnowNo نیز یک لایه ایمنی نظری مکمل ارائه می دهد. این مدل با برآورد عدم قطعیت خروجی، از انجام اقداماتی جلوگیری می کند که مدل نسبت به آن ها اعتماد کافی ندارد. [1] سامانه ROS-LLM نیز از «پرچم



### 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**. [4]

### ۴. ایمنی و اعتبارسنجی

از آنجا که مدل زبانی بزرگ 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.

### 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 مدل زبانی بزرگ 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، چالش برانگیز 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 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 مدل‌های زبانی بزرگ 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 [2] ربات‌ها را به مجموعه محدودی از الگوهای

سامانه‌هایی که از مدل‌های زبانی بزرگ 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.» مدل‌های زبانی بزرگ 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

سامانه‌های **NLex** 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 مدل‌های زبانی بزرگ به ربات‌هایی منجر می‌شود که قابل‌اعتمادتر، قابل‌تعامل‌تر، قابل‌سازی‌تر و برای کاربران غیرمتخصص بسیار قابل‌استفاده‌تر هستند.

### ۱۳. محدودیت‌ها، چالش‌ها

#### ۱. ثابت‌سازی و هم‌ترازی با جهان واقعی

مدل‌های زبانی بزرگ 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.

#### 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. [5]

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

دستوری مقید می‌کردند. اما مدل‌های زبانی بزرگ 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 ثابت‌سازی و برنامه‌ریزی‌های کلاسیک تکمیل نمایند. [1] ادغام ورودی‌های چندوجهی و دانش پیشین نشان داده است که چشم‌انداز مدل‌های زبانی بزرگ به‌عنوان «برنامه‌ریزهای عمومی» روزبه‌روز قابل‌دسترس‌تر می‌شود.

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

معماری‌های ماژولار مانند **MetaMorph** و **PROGRAMPORT** می‌دهند که مدل‌های زبانی بزرگ قادرند ساختارهای رفتاری مقیاس‌پذیر ایجاد کنند. در کنار آن، رویکردهای **RAG** مانند **ARRC** [6] ثابت کرده‌اند که با تکیه بر پایگاه دانش واقعی، می‌توان خطا و توهم‌زایی مدل را کنترل کرده و برنامه‌ریزی قابل‌اعتماد ارائه داد.

در حوزه تعامل انسان-ربات، سامانه‌هایی نظیر **ROS-LLM** [7] نشان داده‌اند که زبان طبیعی می‌تواند واسطه 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 پابرجاست. مدل‌های زبانی بزرگ ممکن است در برخی شرایط رفتار غیرقابل‌اعتماد نشان دهند، بنابراین ترکیب آن‌ها با مکانیزم‌های محافظتی و مدل‌های مبتنی بر فیزیک ضروری است.

## 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.