

به نام خدا



گزارش فاز اول پروژه: سامانه کنترل ربات مبتنی بر مدل‌های زبانی بزرگ

محمد مهدی اسدی

امیر حسین سلیمانی

استاد راهنما: دکتر مازیار پاله‌نگ

۱۴۰۴ آذر

فهرست

3.....	چکیده.....
4.....	فصل 1
4.....	مقدمه.....
6.....	فصل 2
6.....	رویکردهای استفاده از مدل‌های زبانی در رباتیک
6.....	1-2 مقدمه.....
6.....	2-2 تغییر پارادایم در رباتیک مبتنی بر مدل‌های زبانی بزرگ
9.....	3-2 در ک ک زبان طبیعی و تعامل انسان-ربات به عنوان زیربنای شناختی
12.....	4-2 برنامه‌ریزی وظیفه و اجرای بلندمدت با هدایت زبان.....
15.....	5-2 دستکاری فیزیکی، استدلال چندوجهی و پیوند معنا با کنترل
18.....	6-2 معماری‌های مازولار، بازیابی دانش و ادغام با پشههای رباتیکی
22.....	7-2 اینمی، قابلیت اطمینان و تعامل انسان در حلقه
25.....	8-2 محدودیت‌ها، چالش‌های باز و مسیرهای آینده
27.....	فصل 3
27.....	نتیجه‌گیری
29.....	منابع

چکیده

مدل‌های زبانی بزرگ^۱ 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 جندوچهی استدلال کنند و مهارت‌های دستکاری^۲ اشیا را در محیط‌های پویا اجرا کنند. با حرکت رباتیک به سمت سطوح بالاتری از خودمحختاری و هوش تجسم یافته، سامانه‌های مبتنی بر مدل‌های زبانی بزرگ به تدریج نقش هسته شناختی را بر عهده می‌گیرند؛ هسته‌ای که ادراک، تصمیم‌گیری، کنترل و تعامل انسان‌ربات را یکپارچه می‌کند.

این گزارش یک جمع‌بندی جامع از دستاوردهای اخیر در رباتیک مبتنی بر مدل‌های زبانی بزرگ ارائه می‌دهد. در این گزارش 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 تحلیل نحوی مبتنی بر هستی‌شناسی و چارچوب‌های نظاممند آموزش ربات از طریق زبان انسان می‌پردازیم.

ما علاوه بر این، راهبردهای تنظیم تخصصی مدل‌های زبانی برای برنامه‌نویسی صنعتی ربات‌ها و روش‌های دستکاری مبتنی بر استدلال چندوچهی را بررسی می‌کنیم. همچنین پیشرفت‌های اخیر در سامانه‌های کنترل ربات مبتنی بر بازیابی اطلاعات و چارچوب‌های ساخت یافته مبتنی بر ROS را مرور می‌کنیم.

با سازماندهی این یافته‌ها در حوزه‌های برنامه‌ریزی، استدلال، دستکاری، ادراک، معماری‌های کنترلی، یک نگاه یکپارچه از وضعیت کنونی هوش تجسم یافته مبتنی بر مدل زبانی بزرگ 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 را بحث می‌کنیم.

¹ Large language model

² manipulation

فصل ۱

مقدمه

پیشرفت‌های اخیر در حوزه‌ی هوش مصنوعی، به‌ویژه ظهور مدل‌های زبانی بزرگ، موجب تحول بنیادین در بسیاری از شاخه‌های مهندسی و علوم کامپیوتر شده است. یکی از حوزه‌هایی که به‌طور مستقیم از این پیشرفت‌ها تأثیر پذیرفته، علم رباتیک است. در رویکردهای کلاسیک رباتیک، سامانه‌های رباتیکی عمده‌ای بر پایه‌ی برنامه‌نویسی صریح، قوانین از پیش تعريف شده و مدل‌های کنترلی سخت‌گیرانه طراحی می‌شوند. این رویکرد اگرچه در محیط‌های کنترل شده عملکرد قابل قبولی داشت، اما در مواجهه با محیط‌های پویا، دستورات مبهم انسانی و وظایف پیچیده و چندمرحله‌ای با محدودیت‌های جدی روبرو بود.

در سال‌های اخیر، مدل‌های زبانی بزرگ به‌عنوان یک ابزار شناختی قدرتمند، امکان تعامل طبیعی تر میان انسان و ماشین را فراهم کرده‌اند. این مدل‌ها قادرند زبان طبیعی را نه تنها به‌عنوان یک ورودی متنی، بلکه به‌عنوان بستری برای استدلال، برنامه‌ریزی و تصمیم‌گیری به کار گیرند. استفاده از این قابلیت‌ها در رباتیک، مسیر جدیدی را تحت عنوان «رباتیک مبتنی بر مدل‌های زبانی بزرگ» گشوده است؛ مسیری که در آن زبان طبیعی به یک رابط اصلی میان انسان و ربات تبدیل می‌شود و ربات‌ها می‌توانند نیت انسانی را در سطحی فراتر از دستورات صریح درک کنند.

در این گزارش، این مبحث به صورت گام‌به‌گام و ساختاریافته در فصل‌های اول تا هفتم مورد بررسی قرار گرفته است. در فصل اول، تغییر پارادایم در رباتیک و گذار از رویکردهای کلاسیک به سامانه‌های مبتنی بر مدل‌های زبانی بزرگ تشریح شده و مفاهیمی نظری هوش تجسم یافته و نقش زبان در ادراک و کنش ربات معرفی گردیده است. فصل دوم به درک زبان طبیعی و تعامل انسان–ربات به‌عنوان زیربنای شناختی این سامانه‌ها می‌پردازد و نشان می‌دهد که چگونه مدل‌های زبانی بزرگ می‌توانند ابعام، نیت ضمنی و گفت‌وگوی تعاملی را مدیریت کنند.

در ادامه، فصل سوم برنامه‌ریزی وظایف پیچیده و اجرای بلندمدت با هدایت زبان طبیعی را بررسی می‌کند و نقش مدل‌های زبانی را به‌عنوان برنامه‌ریزهای سطح بالا توضیح می‌دهد. فصل چهارم بر چالش‌های دستکاری فیزیکی و استدلال چندوجهی تمرکز دارد و نحوه‌ی

پیوند معنا، ادراک بصری و کنترل فیزیکی را مورد بحث قرار می‌دهد. در فصل پنجم، معماری‌های مژولار و شیوه‌های ادغام مدل‌های زبانی با پشته‌های رباتیکی موجود مانند ROS معرفی می‌شوند. سپس در فصل ششم، مسائل اینمنی، قابلیت اطمینان و نقش انسان در حلقه‌ی تصمیم‌گیری بررسی شده و اهمیت طراحی سازوکارهای کنترلی و نظارتی بر جسته می‌گردد. در نهایت، فصل هفتم به محدودیت‌ها، چالش‌های باز و مسیرهای آینده‌ی این حوزه اختصاص دارد.

هدف از ارائه‌ی این گزارش، فراهم کردن دیدی جامع و تحلیلی نسبت به نقش مدل‌های زبانی بزرگ در سامانه‌های رباتیکی و بررسی مزایا، چالش‌ها و چشم‌اندازهای پیش‌رو است؛ به گونه‌ای که خواننده بتواند تصویری روشن از وضعیت فعلی و مسیر توسعه‌ی این فناوری به دست آورد.

فصل 2

رویکردهای استفاده از مدل‌های زبانی در رباتیک

1-2 مقدمه

2-2 تغییر پارادایم در رباتیک مبتنی بر مدل‌های زبانی بزرگ

در دهه‌های گذشته، رباتیک عمدهاً بر پایه‌ی برنامه‌نویسی صریح، مدل‌های نمادین سخت‌گیرانه و زنجیره‌های کنترلی از پیش تعریف شده توسعه یافته است. در این رویکرد کلاسیک، هر رفتار ربات باید به صورت مستقیم و با جزئیات دقیق توسط مهندسان طراحی می‌شد؛ از تعریف سناریوهای ممکن گرفته تا مدیریت خطاهای و تعامل با انسان. این شیوه اگرچه در محیط‌های صنعتی کنترل شده موفق بوده است، اما در مواجهه با محیط‌های پویا، دستورات مبهم انسانی و ظایف چند مرحله‌ای پیچیده به سرعت به بن‌بست می‌رسد. محدودیت اصلی این نمونه ناتوانی آن در درک و استفاده از زبان طبیعی به عنوان یک رابط عمومی شناختی میان انسان و ربات است.

ظهور مدل‌های زبانی بزرگ³ این وضعیت را به طور بنیادین تغییر داده است. مدل‌های زبانی بزرگ که بر حجم عظیمی از داده‌های متنه آموزش دیده‌اند، قادرند ساختارهای زبانی پیچیده، روابط معنایی پنهان و حتی نیت ضمنی انسان را استخراج کنند. زبان طبیعی از یک ابزار جانی به یک مؤلفه‌ی مرکزی در ادراک، تصمیم‌گیری، برنامه‌ریزی و کنترل ربات تبدیل می‌شود^[1]. در این چارچوب جدید، ربات دیگر صرفاً یک ماشین اجراکننده‌ی دستورات از پیش تعریف شده نیست، بلکه به سامانه‌ای با سطحی از استدلال شناختی و توانایی تفسیر هدف انسانی بدل می‌شود.

یکی از مفاهیم کلیدی که این گذار را توضیح می‌دهد، «هوش تجسم یافته» است. هوش تجسم یافته⁴ بر این ایده تأکید دارد که هوش واقعی تنها در سطح محاسبات نمادین شکل نمی‌گیرد، بلکه در تعامل مستمر میان ادراک، بدن فیزیکی و محیط معنا می‌یابد. مدل‌های زبانی بزرگ، با پیوند دادن زبان به ادراک چندوجهی⁵ و کنش فیزیکی، امکان شکل‌گیری چنین هوشی را فراهم می‌کنند. در سامانه‌های مدرن، مدل‌های زبانی بزرگ به عنوان لایه‌ی شناختی عمل می‌کنند که اطلاعات زبانی، بصری و زمینه‌ای را یکپارچه کرده و آن را به تصمیم‌های قابل اجرا در جهان واقعی تبدیل می‌سازد^[2].

³ Large Language Models

⁴ Embodied intelligence

⁵ Multimodal perception

اهمیت این تغییر مدل زمانی روش‌نتر می‌شود که به نمونه‌های عملی نگاه کنیم. برای مثال، سامانه‌ی **SayCan** که توسط گوگل توسعه داده شده است، نشان می‌دهد چگونه می‌توان استدلال زبانی سطح بالا را با مهارت‌های کنترلی از پیش‌آموخته‌ی ربات ترکیب کرد. در این سامانه، ربات دستورات باز و غیرساخت یافته‌ای مانند «یک نوشیدنی برایم آماده کن» را دریافت کرده و آن‌ها را به توالی‌ای از اقدامات فیزیکی مانند حرکت در محیط، باز کردن کشوها، گرفتن اشیا و قرار دادن آن‌ها در مکان مناسب تبدیل می‌کند. [1] نکته‌ی مهم این است که کاربر نیازی به دانستن جزئیات کنترلی یا زبان برنامه‌نویسی ربات ندارد؛ زبان طبیعی به تنها‌یی به رابط اصلی تعامل تبدیل شده است.

نمونه‌ی برجسته‌ی دیگر، **PaLM-E** است که گامی فراتر از برنامه‌ریزی وظیفه بر می‌دارد و مدل زبانی را مستقیماً در حلقه‌ی ادراک و تصمیم‌گیری ربات قرار می‌دهد. در این سامانه، مدل زبانی نه تنها متن، بلکه تصاویر و بردارهای حالت ربات را نیز به عنوان ورودی دریافت می‌کند و می‌تواند به پرسش‌هایی درباره‌ی محیط پاسخ دهد یا تصمیم‌های کنترلی اتخاذ کند. [1] این رویکرد نشان می‌دهد که مدل‌های زبانی بزرگ می‌توانند از یک «مولد برنامه⁶» به یک «موز تجسم یافته⁷» برای ربات تبدیل شوند؛ مغزی که به‌طور پیوسته با محیط فیزیکی در تعامل است. تصویر 1)

آنچه این مثال‌ها را به هم پیوند می‌دهد، حرکت از رباتیک مبتنی بر دستورالعمل‌های صریح به رباتیک مبتنی بر نیت و معنا است. در این نمونه جدید، مهندسی رباتیک بیش از آنکه به نوشتمن کدهای کنترلی محدود شود، به طراحی معماری‌های می‌پردازد که بتوانند خروجی‌های احتمالی و انعطاف‌پذیر مدل‌های زبانی بزرگ را به رفتارهای ایمن و قابل اعتماد تبدیل کنند. [1]

در سال‌های اخیر، دامنه کاربرد مدل‌های زبانی بزرگ از سامانه‌های رباتیک عمومی فراتر رفته و به‌طور فزاینده‌ای وارد حوزه تولید صنعتی، کارخانه‌های هوشمند و سیستم‌های تصمیم‌یار مهندسی شده است. این روند را می‌توان به عنوان یک امتداد طبیعی از پژوهش‌های LLM-محور در رباتیک دانست که در آن، ربات‌ها نه تنها به عنوان عامل‌های فیزیکی، بلکه به عنوان اجزای یک اکوسیستم تصمیم‌گیری صنعتی در نظر گرفته می‌شوند.

مطالعات اخیر نشان می‌دهند که LLM‌ها می‌توانند نقش «لایه شناختی» را در سیستم‌های تولید ایفا کنند؛ لایه‌ای که توانایی تفسیر زبان طبیعی، دانش فرایندی، داده‌های تاریخی و تعامل انسان-ماشین را به صورت یکپارچه فراهم می‌سازد. برای مثال، چارچوب-**MANUF** یک معماری تصمیم‌گیری صنعتی مبتنی بر چند LLM تنظیم شده را معرفی می‌کند که با ترکیب رتبه‌بندی و هم‌جوشی خروجی‌ها، سوگیری تصمیم‌گیری در سناریوهای عملیاتی تولید را کاهش می‌دهد. [3] این رویکرد نشان می‌دهد که LLM‌ها می‌توانند از سطح «تولید متن» به سطح تصمیم‌سازی قابل اتکا در محیط‌های صنعتی واقعی ارتقا یابند.

⁶ Program generator

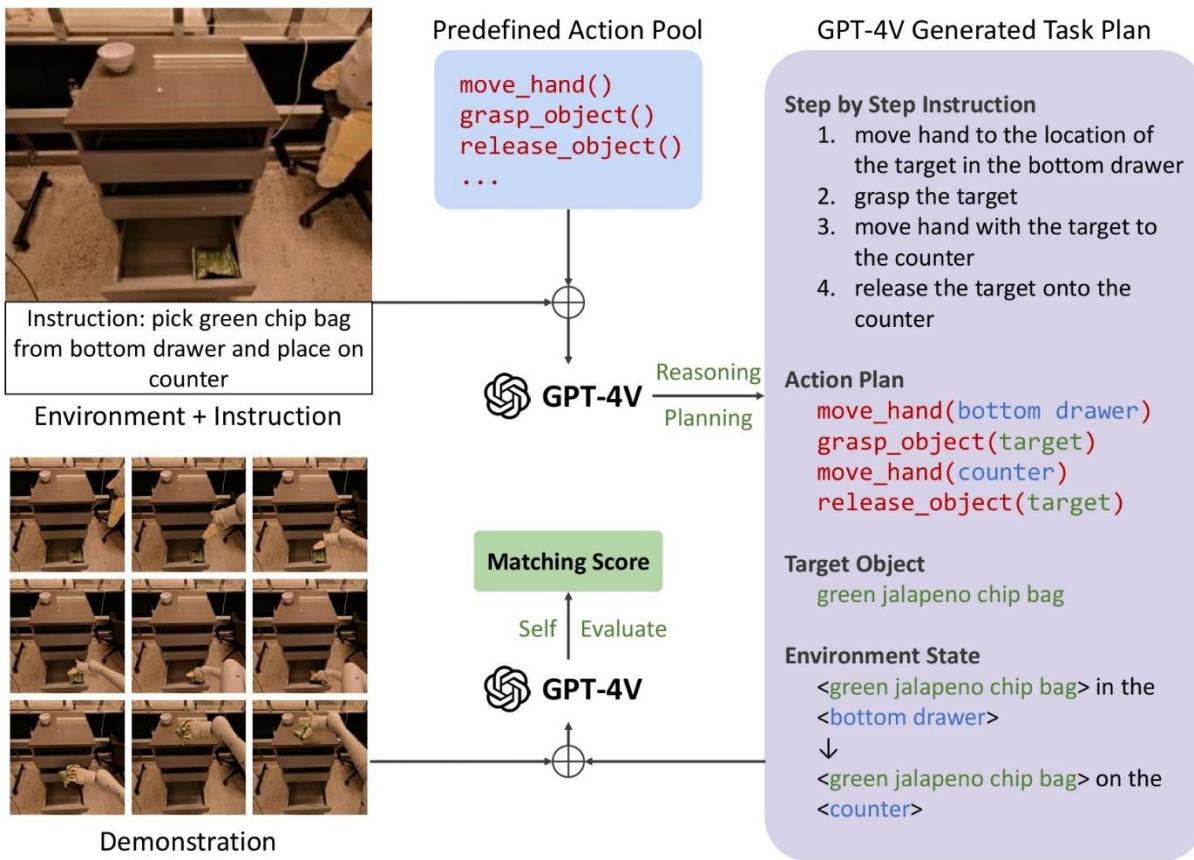
⁷ The embodied brain

از منظر معماری سیستم، پژوهش‌های جدید بر یکپارچه‌سازی LLM‌ها با زیرساخت‌های Industry 4.0 و 5.0 تأکید دارند. چارچوب «مدل دانش صنعتی بزرگ» نشان می‌دهد که LLM‌ها چگونه می‌توانند با گراف‌های دانش صنعتی، داده‌های سایبر-فیزیکی و مدل‌های دیجیتال دوقلو ترکیب شوند تا در کم معنایی عمیق‌تری از فرایندهای تولیدی ایجاد کنند. [4] این موضوع بهویژه در صنایع سنگین مانند فولاد، که داده‌های ناهمگن و وابستگی‌های پیچیده فرایندی وجود دارد، اهمیت بالایی دارد.

در سطح عملیاتی کارخانه، استفاده از عامل‌های LLM برای کنترل ماژولار، برنامه‌ریزی تولید و تطبیق پذیری سیستم‌ها مورد توجه قرار گرفته است. در این راستا، پژوهش «سیستم تولید ماژولار انعطاف‌پذیر تقویت‌شده با عامل‌های LLM» نشان می‌دهد که چگونه LLM‌ها، در ترکیب با Digital Twin و اتوماسیون صنعتی، قادر به برنامه‌ریزی و اجرای وظایف تولیدی غیر از پیش‌تعریف‌شده هستند. [5] این رویکرد، پیوند مستقیمی میان پژوهش‌های رباتیک شناختی و کارخانه‌های هوشمند برقرار می‌کند.

از منظر اینمنی و قابلیت اطمینان، یکی از چالش‌های کلیدی در به کارگیری LLM‌ها در صنعت، تولید خروجی‌های قابل اعتماد برای سیستم‌های کنترلی است. پژوهش LLM4PLC یک خط لوله تکرارشونده برای تولید کد قابل تأیید PLC مبتنی بر LLM ارائه می‌دهد که از ابزارهای راستی‌آزمایی صوری و بازخورد انسانی بهره می‌گیرد. [6] این کار نشان می‌دهد که ادغام LLM‌ها با کنترل صنعتی می‌تواند از سطح پیشنهاد متنی فراتر رفته و به تولید کد اینمن و اجرایی منجر شود.

در سطح کلان‌تر، مروههای جامع اخیر بر نقش LLM‌ها در نسل آینده تولید هوشمند و 5.0 تأکید دارند. این مطالعات، LLM‌ها را به عنوان یک فناوری توانمندساز برای تولید انسان‌محور، همکاری انسان-ماشین و تصمیم‌گیری دانش‌محور معرفی می‌کنند. [7] افرون بر این، پژوهش‌های حوزه طراحی و ساخت نشان داده‌اند که LLM‌ها می‌توانند کل زنجیره طراحی تا تولید، از مشخصات اولیه تا دستورالعمل‌های ساخت، را پشتیبانی کنند و تعامل انسان با سیستم‌های پیچیده مهندسی را ساده‌سازی نمایند. [8]



تصویر ۱. چارچوب روش پیشنهادی برای برنامه‌ریزی وظایف تجسسی 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 برمی‌گیرند. 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 مرجح مقایسه کرده و امتیاز می‌دهد. [1]

2-3 درک زبان طبیعی و تعامل انسان-ربات به عنوان زیربنای شناختی

درک زبان طبیعی سنگ بنای رباتیک مبتنی بر مدل‌های زبانی بزرگ است. اگر ربات نتواند زبان انسان را به درستی تفسیر کند، تمامی قابلیت‌های بعدی مانند برنامه‌ریزی، تضمیم‌گیری و کنترل فیزیکی با اختلال مواجه می‌شوند. در رباتیک کلاسیک، درک زبان عمدتاً به نگاشت‌های ساده‌ی نحوی یا قواعد از پیش تعریف شده محدود بود. این سامانه‌ها معمولاً بر هستی‌شناسی‌های⁸ دستی، وابستگی‌شناسی نحوی⁹ و استخراج افعال و اشیای صریح تکیه داشتند و تنها در دامنه‌های بسیار محدود عملکرد قابل قبولی نشان می‌دادند. [4] نتیجه‌ی

⁸ Based Ontologies

⁹ Syntactic dependency theory

چنین رویکردی، سامانه‌هایی شکننده بود که در مواجهه با ابهام، بیان غیرمستقیم یا تغییرات جزئی در جمله به سرعت دچار خطا می‌شدند.

[5]

مدل‌های زبانی بزرگ این وضعیت را به طور اساسی تغییر داده‌اند. LLM‌ها به جای تکیه بر قواعد صریح، از بازنمایی‌های معنایی غنی استفاده می‌کنند که از داده‌های عظیم متی آموخته شده‌اند. این بازنمایی‌ها به مدل اجازه می‌دهند روابط ضمنی، دانش عمومی و زمینه‌ی گفتار را در نظر بگیرد. به عنوان مثال، یک مدل زبانی بزرگ می‌تواند از جمله‌ای مانند «هوا سرد است» نیت ضمنی «پنجره را بیند» یا «بخاری را روشن کن» را استنتاج کند، بدون آنکه این عمل به طور مستقیم بیان شده باشد. [4] چنین قابلیتی برای تعامل طبیعی انسان–ربات حیاتی است، زیرا زبان انسانی ذاتاً سرشار از ابهام و ارجاعات ضمنی است. [4]

علاوه بر استخراج نیت، مدل‌های زبانی بزرگ نقش مهمی در گفت‌و‌گوی تعاملی ایفا می‌کنند. در سامانه‌های مدرن، اگر دستور کاربر ناقص یا ناسازگار با وضعیت فیزیکی ربات باشد، مدل می‌تواند به جای شکست خاموش، پرسش تکمیلی مطرح کند یا گزینه‌های ممکن را پیشنهاد دهد. این رفتار گفت‌و‌گو محور، زبان را از یک کانال یک‌طرفه فرماندهی به یک ابزار مذاکره و اصلاح مشترک میان انسان و ربات تبدیل می‌کند. [2] در نتیجه، تعامل انسان–ربات به جای آنکه شبیه برنامه‌نویسی باشد، به مکالمه‌ای هدف محور نزدیک می‌شود.

پیاده‌سازی عملی این مفاهیم را می‌توان در سامانه‌های کنترل زبانی ربات‌های واقعی مشاهده کرد. یکی از نمونه‌های کلاسیک، سیستم Parsing Natural Language Sentences into Robot Actions مانند Zora یا NAO پیاده‌سازی شده است. در این سامانه، دستورات گفتاری کاربر مانند «دست را بالا ببر» یا «سرت را به چپ بچرخان» به اعمال فیزیکی متناظر ترجمه می‌شوند. [5] نکته‌ی مهم این است که کاربر نیازی به دانستن نام موتورها، مفاصل یا توالی‌های کنترلی ندارد؛ زبان طبیعی به‌نهایی برای هدایت ربات کافی است. این مثال نشان می‌دهد که چگونه حتی پیش از ظهور مدل‌های زبانی بزرگی مدرن، زبان به عنوان رابط کنترل مطرح بوده، اما با دامنه‌ای محدود و نیازمند مهندسی دستی گسترده.

توانایی رفع ابهام نیز در همین سامانه به صورت عملی نمایش داده شده است. زمانی که کاربر دستور ناقصی مانند «دست را بالا ببر» صادر می‌کند، ربات به طور فعال می‌پرسد «دست چپ یا راست؟» و تنها پس از دریافت پاسخ، عمل را اجرا می‌کند. [5] این رفتار، نمونه‌ای ساده اما بسیار مهم از گفت‌و‌گوی تعاملی است که در ربات‌های واقعی آزمایش شده و نشان می‌دهد چرا در ک زبان طبیعی نمی‌تواند به یک نگاشت ایستا محدود شود. LLM‌ها این ایده را گسترش می‌دهند و امکان پرسش‌های پیچیده‌تر، پیشنهاد گزینه‌ها و حتی مذاکره درباره‌ی هدف را فراهم می‌کنند.

جنبه‌ی دیگر در ک زبان طبیعی، آگاهی از وضعیت درونی و فیزیکی ربات است. در همان سامانه‌ی انسان‌نمای، دو حالت اجرای «stateless» و «stateful» تعریف شده است؛ در حالت اول، ربات وضعیت بدنی خود را میان دستورات حفظ می‌کند و در حالت دوم، پس از هر عمل به وضعیت پیش‌فرض بازمی‌گردد. [5] این تمایز نشان می‌دهد که فهم زبان به‌نهایی کافی نیست، بلکه تفسیر

دستور باید در زمینه‌ی حالت فعلی ربات انجام شود. LLM‌ها با توانایی نگهداری زمینه‌ی مکالمه و تاریخچه‌ی تعامل، این مسئله را در مقیاس بسیار وسیع‌تری حل می‌کنند.

ایمنی نیز به‌طور مستقیم با درک زبان طبیعی گره خورده است. در مثال‌های واقعی، ربات انسان‌نما دستورات ناسازگار یا خطرناک، مانند بلند کردن یک پا در حالی که روی پای دیگر ایستاده است، را رد می‌کند و دلیل آن را به زبان طبیعی توضیح می‌دهد. [5] این رفتار نشان می‌دهد که نگاشت زبان به عمل باید همواره با دانش محدودیت‌های فیزیکی همراه باشد. در سامانه‌های مبتنی بر LLM، این دانش می‌تواند به‌صورت ضمنی از داده‌های آموزشی یا به‌صورت صریح از طریق لایه‌های ایمنی به مدل تزریق شود. [6]

با گسترش پژوهش‌ها در زمینه درک زبان طبیعی در سامانه‌های رباتیک، توجه پژوهشگران به این نکته جلب شده است که بسیاری از چالش‌های نگاشت زبان به عمل، در مقیاس صنعتی پیچیده‌تر و چندلازه‌تر ظاهر می‌شوند. در محیط‌های تولیدی، زبان طبیعی نه تنها برای صدور فرمان‌های سطح بالا، بلکه برای توصیف فرایندها، قیود تولید، دانش اپراتوری و تعامل میان واحدهای مختلف کارخانه به کار می‌رود. از این منظر، درک زبان طبیعی در سیستم‌های صنعتی را می‌توان امدادی از مسئله درک زبان در رباتیک دانست، اما با الزامات سخت‌گیرانه‌تر از نظر دقت، سازگاری و قابلیت تفسیر.

مطالعات اخیر نشان می‌دهند که LLM‌ها می‌توانند نقش واسطی میان زبان انسانی و نمایش‌های رسمی مورد استفاده در تولید صنعتی ایفا کنند. برای مثال، چارچوب‌های پیشنهادی برای کاربرد LLM در صنعت نشان می‌دهند که این مدل‌ها قادرند توصیفات زبانی را به ساختارهای قابل پردازش مانند گراف‌های فرایندی، دستورالعمل‌های اجرایی یا نمایش‌های نمادین تبدیل کنند. [12] این توانایی، شکاف میان زبان غیررسمی اپراتورها و نمایش‌های سخت‌ساخت سیستم‌های کنترلی را کاهش می‌دهد.

در سطحی عمیق‌تر، پژوهش‌های مرتبه با «مدل‌های دانش صنعتی بزرگ» نشان می‌دهند که درک زبان طبیعی در صنعت، زمانی بیشترین اثربخشی را دارد که با دانش زمینه‌ای غنی ترکیب شود. در این رویکرد، LLM‌ها به عنوان هسته‌ی استنتاج زبانی عمل کرده و با گراف‌های دانش صنعتی، داده‌های تاریخی تولید و مدل‌های فرایندی تعامل می‌کنند. [4] چنین ترکیبی امکان تفسیر دقیق‌تر عبارات زبانی مبهم، وابسته به زمینه و مبتنی بر تجربه عملی را فراهم می‌سازد.

از سوی دیگر، در حوزه طراحی و ساخت، پژوهش‌ها نشان داده‌اند که LLM‌ها می‌توانند توصیف‌های زبانی را به‌صورت گام‌به‌گام به مشخصات طراحی، قیود ساخت و در نهایت دستورالعمل‌های تولید تبدیل کنند. این توانایی به‌ویژه در محیط‌هایی که دانش طراحی و ساخت به‌صورت ضمنی در زبان مهندسان نهفته است، اهمیت دارد. [8] در چنین سناریوهایی، درک زبان طبیعی نه تنها برای اجرای دستور، بلکه برای استخراج دانش مهندسی پنهان ضروری است.

در محیط‌های تولید هوشمند، درک زبان طبیعی همچنین به تعامل میان چند عامل (انسان، ربات و سیستم‌های نرم‌افزاری) گسترش می‌یابد. پژوهش‌های مربوط به سامانه‌های چند‌عامله مبتنی بر LLM نشان می‌دهند که زبان طبیعی می‌تواند به عنوان یک پروتکل مشترک برای هماهنگی میان عامل‌های ناهمگون در کارخانه عمل کند. [13] در این چارچوب، درک زبان دیگر صرفاً به نگاشت فرمان به عمل محدود نیست، بلکه به ابزاری برای هماهنگی، مذاکره و تطبیق پویا میان اجزای سیستم تبدیل می‌شود.

در نهایت، با ورود LLM ها به لایه‌های اجرایی صنعت، مسئله در ک زبان طبیعی به سطوح پایین تر کنترل نیز نفوذ کرده است. پژوهش‌هایی مانند LLM4PLC نشان می‌دهند که توصیفات زبانی می‌توانند به صورت ساخت‌یافته به منطق کنترلی رسمی ترجمه شوند، مشروط بر آنکه فرایندهای اعتبارسنجی و بازخورد مناسب در حلقه قرار گیرند. [6] این امر نشان می‌دهد که در ک زبان طبیعی در سیستم‌های صنعتی می‌تواند از سطح تعامل کاربر فراتر رفته و به بخشی از زنجیره تولید منطق کنترلی تبدیل شود.

4-2 برنامه‌ریزی وظیفه و اجرای بلندمدت با هدایت زبان

پس از آنکه زبان طبیعی به عنوان یک رابط شناختی قابل اعتماد میان انسان و ربات ثبت شد، گام منطقی بعدی استفاده از این توانمندی برای برنامه‌ریزی وظایف پیچیده و چندمرحله‌ای است. برنامه‌ریزی وظیفه¹⁰ در رباتیک به معنای تبدیل یک هدف سطح بالا، که اغلب به صورت زبانی بیان می‌شود، به دنباله‌ای از اعمال اجرایی است که ربات بتواند آنها را در دنیای فیزیکی انجام دهد. در رباتیک کلاسیک، این فرایند معمولاً به روش‌های نمادین برنامه‌ریزی مبتنی بر حالت، درخت جست‌وجو یا زبان‌هایی نظیر PDDL محدود می‌شد. این روش‌ها اگرچه از نظر صوری دقیق هستند، اما نیازمند مدل‌سازی کامل محیط و تعریف صریح تمامی پیش‌شرط‌ها و اثرات اعمال‌اند؛ امری که در محیط‌های واقعی بهندرت امکان‌پذیر است. [1]

مدل‌های زبانی بزرگ رویکردی متفاوت به این مسئله ارائه می‌دهند. مدل‌های زبانی بزرگ قادرند از داشش عمومی و ساختارهای روایی زبان برای تجزیه‌ی یک هدف کلی به زیروظایف معنادار استفاده کنند، حتی در شرایطی که مدل صریحی از محیط در اختیار ندارند. به بیان دیگر، آن‌ها می‌توانند نقش یک «برنامه‌ریز سطح بالا»¹¹ را ایفا کنند که خروجی آن نه دستورات کنترلی دقیق، بلکه یک توالی مفهومی از اقدامات است. [4]

با این حال، استفاده‌ی مستقیم از خروجی زبانی LLM ها برای کنترل ربات با چالش‌های جدی مواجه است. خروجی‌های مدل‌های زبانی ذاتاً احتمالی‌اند و تضمین رسمی درباره‌ی صحت یا ایمنی آن‌ها وجود ندارد. به همین دلیل، معماری‌های عملی معمولاً از یک ساختار ترکیبی استفاده می‌کنند که در آن LLM وظیفه‌ی تولید یا ارزیابی برنامه‌ی سطح بالا را بر عهده دارد و اجرای واقعی توسط مأذول‌های کنترلی قابل اعتماد انجام می‌شود. این رویکرد ترکیبی، که گاه با عنوان «برنامه‌ریزی گراندشه»¹² شناخته می‌شود، یکی از محورهای اصلی پژوهش در رباتیک مبتنی بر LLM . [2]

¹⁰ Task programming

¹¹ High level programming

¹² Grounded planning

پیاده‌سازی عملی این ایده به طور شاخص در سامانه‌ی SayCan دیده می‌شود. در SayCan، مدل زبانی بزرگ وظیفه دارد بر اساس دستور زبانی کاربر، مجموعه‌ای از اقدامات ممکن را ارزیابی کند و احتمال مناسب بودن هر اقدام برای رسیدن به هدف را تخمین بزنند. این اقدامات از پیش به صورت مهارت‌های قابل اجرا روی ریات تعریف شده‌اند، مانند «حرکت به سمت میز»، «باز کردن کشو» یا «گرفتن فنجان» [1] سپس، این امتیازدهی زبانی با احتمال موفقیت هر مهارت که از داده‌های تجربی بدست آمده است ترکیب می‌شود. نتیجه، انتخاب عملی است که هم از نظر معنایی با هدف کاربر همخوانی دارد و هم از نظر فیزیکی قابل اجراست.

این معماری نشان می‌دهد که مدل‌های زبانی بزرگ به جای جایگزینی کامل برنامه‌ریزی کلاسیک، می‌توانند آن را تکمیل کنند. در SayCan، زبان به عنوان راهنمای انتخاب در فضای اعمال عمل می‌کند، نه به عنوان کنترل کننده‌ی مستقیم موتورها. برای مثال، زمانی که کاربر می‌گوید «برایم یک نوشیدنی آماده کن»، مدل زبانی می‌داند که ابتدا باید به آشپزخانه رفت، سپس یک لیوان پیدا کرده و در نهایت آن را پر کرد، حتی اگر این مراحل به طور صریح در دستور ذکر نشده باشند. [1] این توانایی استخراج ساختار وظیفه از زبان طبیعی دقیقاً همان چیزی است که در برنامه‌ریزی کلاسیک به سختی و با هزینه‌ی مهندسی بالا بدست می‌آید.

نمونه‌ی دیگر از این رویکرد، سامانه‌های مبتنی بر ROS-LLM هستند که در آن‌ها مدل زبانی بزرگ به عنوان یک لایه‌ی برنامه‌ریز روی پشتی استاندارد رباتیکی ROS قرار می‌گیرد. در این سامانه‌ها، دستورات زبانی کاربر به توالی‌ای از فراخوانی سرویس‌ها و اکشن‌های ROS تبدیل می‌شود. [2] مزیت این روش در آن است که می‌توان بدون تغییر در زیرساخت کنترلی موجود، قابلیت برنامه‌ریزی زبانی را به ربات افزود. به علاوه، ROS-LLM نشان می‌دهد که چگونه مدل‌های زبانی بزرگ می‌توانند با ابزارهای مهندسی جافتاده ادغام شوند، نه اینکه به عنوان یک سامانه‌ی جداگانه عمل کنند.

اجرای بلندمدت وظایف یکی دیگر از جنبه‌های مهم برنامه‌ریزی زبانی است. بسیاری از وظایف دنیای واقعی، مانند تمیز کردن یک اتاق یا آماده‌سازی یک وعده‌ی غذایی، شامل ده‌ها مرحله هستند و ممکن است در طول اجرا با خطا یا تغییر شرایط مواجه شوند. مدل‌های زبانی بزرگ به دلیل توانایی نگهداری زمینه و باز تفسیر هدف، برای مدیریت چنین وظایفی مناسب‌اند. [1] در عمل، این ویژگی به ربات اجازه می‌دهد پس از یک شکست جزئی، برنامه‌ی خود را اصلاح کند یا از کاربر راهنمایی بگیرد، به جای آنکه کل فرایند را متوقف سازد.

با این حال، مثال‌های عملی نشان می‌دهند که این توانمندی هنوز محدودیت‌هایی دارد. در سامانه‌های واقعی، اجرای بلندمدت اغلب نیازمند نظارت انسانی یا محدود کردن دامنه‌ی تصمیم‌گیری LLM است تا از انحراف از هدف اصلی جلوگیری شود. [2] این موضوع بار دیگر اهمیت معماری‌های ترکیبی را برجسته می‌کند؛ معماری‌هایی که در آن‌ها زبان به عنوان ابزار استدلال و هدایت به کار می‌رود، اما کنترل نهایی همچنان در چارچوب‌های ایمن و قابل پیش‌بینی انجام می‌شود.

با بلوغ معماری‌های مبتنی بر مدل‌های زبانی بزرگ در سامانه‌های رباتیک، این معماری‌ها به تدریج از چارچوب‌های آزمایشگاهی به ساختارهای مقیاس‌پذیر و چندلایه در محیط‌های صنعتی توسعه یافته‌اند. در چنین محیط‌هایی، معماری LLM محور ناگزیر باید هم‌زمان

با لایه‌های ادراکی، برنامه‌ریزی، کنترل و دانش فرایندی تعامل داشته باشد؛ امری که نیازمند باز تعریف نقش LLM در کل زنجیره تصمیم‌گیری است.

مطالعات اخیر در حوزه تولید هوشمند نشان می‌دهند که LLM‌ها اغلب به عنوان یک لایه میان‌افزار شناختی (cognitive middleware) در معماری سیستم قرار می‌گیرند. این لایه وظیفه دارد ورودی‌های زبانی انسان، داده‌های ساخت‌یافته صنعتی و بازخورد سیستم‌های فیزیکی را به یک نمایش معنایی مشترک تبدیل کند. چارچوب‌های پیشنهادی برای کاربرد LLM در صنعت نشان می‌دهند که این مدل‌ها می‌توانند به صورت ماژولار در کنار سیستم‌های موجود مانند MES، ERP و Digital Twin قرار گیرند، بدون آنکه نیاز به جایگزینی کامل زیرساخت‌های سنتی باشد. [12]

در سطح بالاتر معماری، پژوهش‌های مبتنی بر «مدل دانش صنعتی بزرگ» نشان می‌دهند که ترکیب LLM با گراف‌های دانش و مدل‌های فرایندی، امکان استنتاج چندمرحله‌ای و وابسته به زمینه را فراهم می‌سازد. [4] در این معماری‌ها، LLM تنها یک مولد پاسخ نیست، بلکه به عنوان موتور استنتاجی عمل می‌کند که بر روی دانش ساخت‌یافته صنعت اجرا می‌شود. این رویکرد، شباهت مفهومی قابل توجهی با معماری‌های بازیابی-افزوده در رباتیک دارد، اما با دامنه و پیچیدگی بالاتر داده‌های صنعتی.

از منظر معماری عامل‌محور، پژوهش‌های اخیر نشان داده‌اند که استفاده از چند LLM یا چند نمونه عامل LLM می‌تواند به تفکیک مسئولیت‌ها در سیستم‌های تولیدی منجر شود. برای مثال، در سامانه‌های چندعامله تولید، هر عامل می‌تواند مسئول یک زیرسیستم (برنامه‌ریزی، تخصیص منابع، پایش کیفیت یا تعامل انسانی) باشد و هماهنگی میان آن‌ها از طریق زبان طبیعی یا پیام‌های ساخت‌یافته انجام گیرد. [13] این معماری‌ها به ویژه در کارخانه‌های ماژولار و خطوط تولید انعطاف‌پذیر کارآمد هستند.

در لایه اجرا، یکی از جالش‌های کلیدی معماری‌های LLM محور، اتصال این و قابل اعتماد به سیستم‌های کنترلی پایین‌دست است. پژوهش LLM4PLC نشان می‌دهد که می‌توان معماری‌هایی طراحی کرد که در آن‌ها LLM نقش تولید‌کننده منطق سطح بالا را بر عهده دارد، در حالی که لایه‌های اعتبارسنجی صوری و ابزارهای کنترلی کلاسیک، صحت و ایمنی خروجی را تضمین می‌کنند. [6] چنین معماری‌هایی، مرز میان تولید زبان‌محور و اجرای فیزیکی را به صورت نظاممند مدیریت می‌کنند.

در نهایت، مرورهای جامع حوزه تولید هوشمند نشان می‌دهند که معماری‌های مبتنی بر LLM در حال حرکت به سوی الگوهای انسان‌محور و تطبیق‌پذیر هستند، جایی که تعامل انسان، ربات و سیستم‌های نرم‌افزاری در یک چارچوب مشترک زبانی و شناختی صورت می‌گیرد. [7] این تحول معماری، پیوند مستقیمی میان پژوهش‌های رباتیک شناختی و الزامات عملی صنعت برقرار می‌کند.

2-5 دستکاری فیزیکی، استدلال چندوجهی و پیوند معنا با کنترل

اگرچه در فصل پیش نشان داده شد که مدل‌های زبانی بزرگ می‌توانند اهداف سطح بالا را به برنامه‌های اجرایی معنادار تبدیل کنند، اما یکی از دشوارترین چالش‌ها در رباتیک همچنان باقی می‌ماند: دستکاری فیزیکی اشیا در دنیای واقعی. دستکاری نیازمند دقت عددی بالا، ادراک پیوسته و واکنش سریع به عدم قطعیت‌های محیطی است. به طور سنتی، این حوزه بر مدل‌های هندسی، برنامه‌ریزی مسیر و کنترل بازخوردی تکیه داشته و فاصله‌ی قابل توجهی میان استدلال نمادین و اجرای فیزیکی وجود داشته است. [7]

مدل‌های زبانی بزرگ، به تهایی قادر به حل مستقیم مسائل کنترل پیوسته نیستند، اما می‌توانند نقش مهمی در راهنمایی معنایی فرایند دستکاری ایفا کنند. ایده‌ی اصلی آن است که مدل‌های زبانی بزرگ دانش سطح بالا درباره اشیا، کار کرد آن‌ها و راهبردهای متداول دستکاری را فراهم می‌کنند، در حالی که جزئیات عددی همچنان به ماژول‌های کنترلی تخصصی سپرده می‌شود. برای مثال، دانستن اینکه «فنجان را باید از دسته گرفت» یا «بطری را باید عمودی نگه داشت» نوعی دانش معنایی است که به راحتی در زبان بیان می‌شود اما به سختی در قالب قواعد هندسی صریح مدل‌سازی می‌شود. [8]

این رویکرد به ویژه در معماری‌های چندوجهی اهمیت می‌یابد. در چنین سامانه‌هایی، زبان، بینایی و حالت فیزیکی ربات به صورت هم‌زمان در تصمیم‌گیری دخیل هستند. مدل‌های زبانی بزرگ در این میان نقش ادغام کننده معنا را بازی می‌کنند؛ یعنی اطلاعات بصری و حسی را در چارچوب مفاهیم زبانی تفسیر می‌کنند و بر اساس آن، راهبرد مناسب دستکاری را پیشنهاد می‌دهند. [2]

پیاده‌سازی عملی این ایده به طور شاخص در سامانه‌ی RT-Grasp دیده می‌شود. RT-Grasp یک سیستم دستکاری است که از مدل‌های زبانی بزرگ برای هدایت فرایند گرفتن اشیا استفاده می‌کند. در این سامانه، بیش از آنکه شبکه‌ی عصبی عددی محل و زاویه‌ی گرفتن را پیش‌بینی کند، یک مرحله‌ی استدلال زبانی انجام می‌شود که هدف و محدودیت‌های دستکاری را مشخص می‌کند. [8] رای مثال، اگر هدف «برداشتن یک لیوان پر از آب» باشد، استدلال زبانی می‌تواند بر حفظ تعادل و جلوگیری از ریختن محظوظ تأکید کند و بدین ترتیب، فضای جست‌وجوی کنترل عددی را محدود سازد.

اهمیت RT-Grasp در این است که نشان می‌دهد افزودن لایه‌ی زبانی می‌تواند هم اینمی و هم قابلیت تعیین سیستم را افزایش دهد. به جای آنکه مدل کنترل عددی مجبور باشد تمامی حالات ممکن را از داده بیاموزد، بخشی از دانش به صورت زبانی و قابل تفسیر در اختیار سیستم قرار می‌گیرد. این موضوع به ویژه در محیط‌های جدید یا اشیای نادیده‌شده اهمیت دارد، جایی که داده‌ی آموزشی محدود است. [8] [7]

نمونه‌ی دیگری از پیوند معنا و کنترل را می‌توان در سامانه‌ی VoxPoser مشاهده کرد. VoxPoser از زبان طبیعی برای تعریف قیود فضایی و هدف‌های دستکاری استفاده می‌کند؛ برای مثال، کاربر می‌تواند بگوید «این مکعب را کنار لیوان و دور از لبه میز قرار

بده». سیستم این دستور را به قیود هندسی قابل اجرا برای برنامه‌ریز مسیر تبدیل می‌کند. [1] در اینجا، زبان نقش واسطه را ایفا می‌کند که مفاهیم کیفی مانند «کنار»، «دور از» یا «روی» را به محدودیت‌های کمی در فضای پیکربندی ربات ترجمه می‌کند.

این نوع ترجمه‌ی معنا به قیود کنترلی نشان می‌دهد که مدل‌های زبانی بزرگ لزوماً نیازی به تولید مستقیم دستورات حرکتی ندارند. ارزش اصلی آن‌ها در ایجاد ساختار مفهومی برای مسئله است؛ ساختاری که مأذول‌های کلاسیک کنترل می‌توانند بر اساس آن عمل کنند. [9]

با وجود این پیشرفت‌ها، دستکاری فیزیکی مبتنی بر LLM همچنان با چالش‌های جدی مواجه است. عدم قطعیت در ادراک بصری، خطاهای مکانیکی و تقاضاهای ظریف میان اشیای مشابه می‌تواند باعث شکست سیستم شوند. مثال‌های موجود در مقالات نشان می‌دهند که بسیاری از سامانه‌ها هنوز نیازمند محیط‌های نسبتاً کنترل شده یا ناظارت انسانی هستند تا عملکرد قابل اعتمادی داشته باشند. [8] این واقعیت تأکید می‌کند که مدل‌های زبانی بزرگ به تنها بی راه حل نهایی نیستند، بلکه بخشی از یک اکوسیستم پیچیده‌ی کنترلی‌اند.

در ادامه‌ی پژوهش‌های مرتبط با برنامه‌ریزی و تصمیم‌گیری مبتنی بر زبان، مطالعات جدید نشان می‌دهند که LLM‌ها می‌توانند در محیط‌های تولیدی نقش فعالی در تولید، اصلاح و تطبیق توالی‌های تصمیم‌گیری ایفا کنند. برخلاف سناریوهای رباتیک منفرد، در تولید صنعتی برنامه‌ریزی غالب با قیود چندگانه، منابع محدود و اهداف متضاد همراه است؛ موضوعی که نیازمند استنتاج چندمرحله‌ای و وابسته به زمینه است.

در این راستا، چارچوب LLM-MANUF نشان می‌دهد که چگونه می‌توان از چند LLM تنظیم شده برای پشتیبانی از تصمیم‌گیری صنعتی استفاده کرد، به گونه‌ای که هر مدل بخشی از فضای تصمیم (مانند برنامه‌ریزی تولید، تخصیص منابع یا زمانبندی) را پوشش دهد و خروجی نهایی از طریق هم‌جوشی و رتبه‌بندی نتایج به دست آید. [3] این رویکرد را می‌توان امتدادی از ایده‌های برنامه‌ریزی زبان محور در رباتیک دانست که به مقیاس کارخانه تعیین یافته است.

از منظر رسمی تر، پژوهش‌های ترکیبی LLM و بهینه‌سازی نشان داده‌اند که مدل‌های زبانی می‌توانند به عنوان مولد یا تفسیر کننده‌ی مدل‌های برنامه‌ریزی صوری عمل کنند. برای مثال، در مسئله تخصیص وظایف و زمانبندی چندرباتی، LLM‌ها برای ساخت خود کار مدل‌های MILP از توصیفات زبانی استفاده شده‌اند و سپس حل کننده‌های کلاسیک برای تضمین بهینگی به کار رفته‌اند. [8] این الگو، شکاف میان زبان طبیعی و برنامه‌ریزی ریاضی را کاهش می‌دهد.

در محیط‌های تولیدی مأذولات، مسئله برنامه‌ریزی غالب پویا و وابسته به وضعیت لحظه‌ای سیستم است. پژوهش‌های مربوط به سیستم‌های تولید انعطاف‌پذیر نشان می‌دهند که عامل‌های LLM می‌توانند با استفاده از بازخورد دیجیتال دوقلو و داده‌های اجرایی، برنامه‌های تولید را به صورت برخط اصلاح کنند. [5] چنین قابلیتی، امکان واکنش سریع به تغییرات سفارش، خرایی تجهیزات یا مداخله انسانی را فراهم می‌سازد.

در سطح کلان‌تر، مروارهای حوزه تولید هوشمند نشان می‌دهند که برنامه‌ریزی مبتنی بر LLM به تدریج از تولید توالی‌های ایستا به سمت برنامه‌ریزی تطبیق‌پذیر و انسان‌محور حرکت می‌کند، جایی که زبان طبیعی به عنوان واسطه اصلی تعریف اهداف، قیود و ترجیحات انسانی به کار می‌رود. [7] این روند، پیوند مستقیمی میان پژوهش‌های برنامه‌ریزی رباتیک و الزامات عملی صنعت برقرار می‌سازد.

در کنار برنامه‌ریزی سطح بالا، یکی از ابعاد مهم تصمیم‌گیری صنعتی مبتنی بر LLM، توانایی این مدل‌ها در تفسیر اهداف مبهم و تبدیل آن‌ها به قیود صریح عملیاتی است. در محیط‌های تولیدی واقعی، اهداف معمولاً به صورت زبانی و غیررسمی بیان می‌شوند، مانند «کاهش زمان تحويل بدون افت کیفیت» یا «اولویت‌دهی به سفارش‌های خاص». پژوهش‌های اخیر نشان می‌دهند که LLM‌ها می‌توانند چنین اهدافی را به مجموعه‌ای از قیود، معیارهای بهینه‌سازی و سیاست‌های تصمیم‌گیری قابل‌پردازش تبدیل کنند و بدین ترتیب نقش واسطه میان تصمیم‌گیر انسانی و الگوریتم‌های برنامه‌ریزی را ایفا نمایند.

از منظر تعامل انسان–سیستم، استفاده از LLM‌ها در برنامه‌ریزی صنعتی امکان بازتعریف نقش اپراتور انسانی را فراهم می‌کند. به جای تعریف مستقیم پارامترهای پیچیده برنامه‌ریزی، اپراتورها می‌توانند از طریق زبان طبیعی، ترجیحات، استشناها و دانش تجربی خود را به سیستم منتقل کنند. مطالعات حوزه طراحی و تولید نشان می‌دهند که این تعامل زبانی می‌تواند بار شناختی کاربران را کاهش داده و در عین حال شفافیت تصمیم‌گیری را افزایش دهد، زیرا سیستم قادر است منطق برنامه‌ریزی خود را نیز به صورت زبانی توضیح دهد.

در سیستم‌های تولید چندسطوحی، تصمیم‌گیری اغلب در سطوح مختلفی از افق زمانی انجام می‌شود؛ از برنامه‌ریزی بلندمدت تولید تا زمان‌بندی کوتاه‌مدت ماشین‌ها. پژوهش‌های جدید نشان می‌دهند که LLM‌ها می‌توانند به عنوان یک لایه هماهنگ کننده میان این سطوح عمل کنند و با در نظر گرفتن قیود و اهداف هر سطح، سازگاری میان تصمیمات را حفظ نمایند. این نقش هماهنگ‌کننده به ویژه در کارخانه‌های بزرگ و صنایع سنگین که عدم هماهنگی میان تصمیمات می‌تواند به هزینه‌های قابل توجه منجر شود، اهمیت دارد.

علاوه بر این، ترکیب LLM‌ها با مدل‌های دانش صنعتی و داده‌های تاریخی تولید امکان یادگیری از تصمیمات گذشته را فراهم می‌سازد. در این رویکرد، LLM‌ها نه تنها بر اساس قوانین صریح، بلکه بر پایه الگوهای استخراج شده از تجربه‌های پیشین تصمیم‌گیری می‌کنند. چنین قابلیتی، برنامه‌ریزی صنعتی را از یک فرایند تطبیق‌پذیر و یادگیرنده تبدیل می‌کند که می‌تواند در طول زمان بهبود یابد.

در نهایت، این خط پژوهشی نشان می‌دهد که برنامه‌ریزی و تصمیم‌گیری مبتنی بر LLM در صنعت، صرفاً جایگزینی برای روش‌های کلاسیک نیست، بلکه چارچوبی ترکیبی را شکل می‌دهد که در آن زبان طبیعی، دانش انسانی، مدل‌های ریاضی و سیستم‌های اجرایی به صورت هم‌افزا عمل می‌کنند. این هم‌افزایی، زمینه را برای توسعه سامانه‌های تولیدی فراهم می‌سازد که هم از نظر کارایی و هم از نظر تعامل‌پذیری با انسان، گامی فراتر از رویکردهای سنتی بر می‌دارند.

2-6-معماری‌های مژولار، بازیابی دانش و ادغام با پشتهدانی رباتیکی

با افزایش نقش مدل‌های زبانی بزرگ در برنامه‌ریزی و دستکاری، مسئله‌ی اساسی نحوه‌ی ادغام این مدل‌ها با سامانه‌های رباتیکی پیچیده و موجود مطرح می‌شود. استفاده‌ی مستقیم و یکپارچه از مدل‌های زبانی بزرگ به عنوان یک جزء یکتا، اگرچه از نظر مفهومی جذاب است، اما در عمل با مشکلاتی مانند هزینه‌ی محاسباتی بالا، نبود تضمین ایمنی و دشواری اشکال‌زدایی همراه است. به همین دلیل معماری‌های مژولار را به عنوان رویکرد غالب برای رباتیک مبتنی بر LLM معرفی می‌شود. [9] در این معماری‌ها، هر جزء مسئولیتی مشخص دارد و مدل زبانی تنها یکی از مژول‌ها در یک زنجیره‌ی پردازشی بزرگ‌تر است.

ایده‌ی اصلی معماری مژولار آن است که استدلال سطح‌بالا، که ذاتاً نمادین و احتمالی است، از کترول سطح‌پایین، که نیازمند دقت و قابلیت پیش‌بینی بالاست، جدا شود. مدل‌های زبانی بزرگ در لایه‌های بالادستی قرار می‌گیرند و وظایفی مانند تفسیر زبان، تولید برنامه‌ی مفهومی و انتخاب مهارت مناسب را انجام می‌دهند، در حالی که مژول‌های کنترلی کلاسیک مسئول اجرای دقیق این مهارت‌ها هستند. این تفکیک نه تنها ایمنی سیستم را افزایش می‌دهد، بلکه امکان جایگزینی یا به روزرسانی هر مژول را بدون بازطراحی کل سامانه فراهم می‌کند. [2]

یکی از چالش‌های کلیدی در این معماری‌ها، محدودیت دانش درونی مدل‌های زبانی بزرگست. هرچند این مدل‌ها بر داده‌های عظیمی آموخته‌اند، اما دانش آن‌ها ثابت و وابسته به زمان آموختش است. برای غلبه بر این مشکل، رویکرد تولید مبتنی بر بازیابی¹³ مطرح شده است. در این روش، مدل زبانی به یک پایگاه دانش خارجی متصل می‌شود و پیش از تولید پاسخ یا برنامه، اطلاعات مرتبط را بازیابی می‌کند. [9] این پایگاه دانش می‌تواند شامل مستندات فنی ربات، نقشه‌ی محیط، محدودیت‌های ایمنی یا تجربه‌های پیشین باشد.

پیاده‌سازی عملی این ایده را می‌توان در سامانه‌ی ARRC مشاهده کرد. یک معماری رباتیکی مژولار است که در آن زبان طبیعی به عنوان رابط اصلی میان کاربر و ربات عمل می‌کند، اما تصمیم‌گیری نهایی از طریق چندین لایه‌ی بررسی و اعتبارسنجی انجام می‌شود. [9] در این سامانه، دستور زبانی کاربر ابتدا توسط مدل زبانی تحلیل می‌شود، سپس اطلاعات مرتبط از پایگاه دانش بازیابی می‌گردد و در نهایت یک خروجی ساخت‌یافته تولید می‌شود که برای مژول‌های کنترلی قابل فهم است. (تصویر 2)

¹³ Retrieval-Augmented Generation

نکته‌ی مهم در ARRC آن است که خروجی LLM به صورت متن آزاد مستقیماً اجرا نمی‌شود. به جای آن، مدل موظف است خروجی خود را در قالبی مشخص و محدود ارائه دهد؛ قالبی که تنها شامل اعمال مجاز و پارامترهای قابل کنترل است. این محدودسازی نقش مهمی در افزایش ایمنی و قابلیت اطمینان سیستم دارد و به طور مستقیم با مهار رفتارهای غیرقابل پیش‌بینی ها همسو است. [6]

ادغام با پشته‌های رباتیکی موجود، به ویژه ROS، یکی دیگر از مزایای معماری‌های ماژولار است. ROS به عنوان یک چارچوب استاندارد، ابزارها و پروتکل‌های ارتباطی گستردۀ‌ای را برای رباتیک فراهم می‌کند. سامانه‌هایی مانند ROS-LLM نشان داده‌اند که می‌توان مدل‌های زبانی بزرگ را به عنوان یک گرهی اضافی به این پشته افزود، بدون آنکه ساختار کلی سیستم تغییر کند. [2] در این رویکرد، مدل زبانی دستورات زبانی را به فراخوانی سرویس‌ها یا اکشن‌های موجود در ROS نگاشت می‌کند و بدین ترتیب، قابلیت‌های جدیدی به ربات افزوده می‌شود.

از منظر مهندسی، این نوع ادغام مزایای قابل توجهی دارد. نخست آنکه توسعه دهنده‌گان می‌توانند از سرمایه‌گذاری‌های قبلی خود در طراحی کنترل‌کننده‌ها و برنامه‌ریزی‌های کلاسیک بهره ببرند. دوم آنکه اشکال‌زدایی و ارزیابی سیستم ساده‌تر می‌شود، زیرا هر ماژول رفتاری مشخص و قابل اندازه‌گیری دارد. این شفافیت معماری برای پذیرش صنعتی رباتیک مبنی بر LLM حیاتی است. [9]

با وجود این مزایا، معماری‌های ماژولار بدون چالش نیستند. همانگی میان ماژول‌ها، تأخیر ناشی از فراخوانی مدل‌های بزرگ و مدیریت خطأ در مرزهای ماژول‌ها از جمله مسائلی هستند که در مثال‌های عملی نیز دیده می‌شوند. [9] این مسائل نشان می‌دهند که طراحی معماری به اندازه‌ی انتخاب مدل زبانی اهمیت دارد و موفقیت یک سامانه‌ی رباتیکی بیش از آنکه به قدرت خام LLM وابسته باشد، به نحوه استفاده‌ی مهندسی شده از آن بستگی دارد.

با گسترش استفاده از مدل‌های بزرگ در سامانه‌های رباتیک، مسئله کنترل، اجرا و اتصال لایه‌شناختی به لایه فیزیکی به یکی از محورهای اصلی پژوهش تبدیل شده است. در این چارچوب، بخش قابل توجهی از پژوهش‌های جدید بر این نکته تأکید دارند که خروجی‌های زبانی و تصمیمی LLM‌ها تنها زمانی ارزش عملی دارند که بتوانند به صورت ایمن، قابل اعتماد و قابل راستی‌آزمایی به فرمان‌های اجرایی تبدیل شوند. این چالش در محیط‌های صنعتی، به مراتب پیچیده‌تر از سناریوهای آزمایشگاهی رباتیک ظاهر می‌شود.

در سیستم‌های تولیدی، لایه اجرا معمولاً مبنی بر زیرساخت‌های تثبیت‌شده‌ای مانند PLC‌ها، سیستم‌های کنترل توزیع شده و استانداردهای سخت‌گیرانه ایمنی است. پژوهش LLM4PLC نشان می‌دهد که چگونه می‌توان از LLM‌ها برای تولید منطق کنترلی استفاده کرد، مشروط بر آنکه این منطق در یک حلقه تکرارشونده شامل بررسی نحوی، اعتبارسنجی صوری و بازخورد انسانی قرار گیرد. [6] این رویکرد، نشان‌دهنده مسیری عملی برای اتصال خروجی‌های زبانی به کنترل صنعتی واقعی است.

از منظر معماری اجرا، مطالعات جدید پیشنهاد می‌کنند که LLM‌ها نباید مستقیماً مسئول ارسال فرمان‌های سطح پایین باشند، بلکه بهتر است در نقش تولید‌کننده سیاست‌ها، توالی‌های کنترلی سطح بالا یا مشخصات رسمی عمل کنند. این سیاست‌ها سپس توسط لایه‌های

کنترلی کلاسیک یا سیستم‌های مبتنی بر قوانین اجرا می‌شوند. چنین تفکیکی، امکان حفظ قابلیت پیش‌بینی و اینمنی سیستم را فراهم می‌کند، در حالی که انعطاف‌پذیری و قابلیت تفسیر زبان طبیعی نیز حفظ می‌شود.

در محیط‌های تولید هوشمند، اجرای وظایف اغلب با بازخورد پیوسته از سنسورها، ماشین‌ها و اپراتورهای انسانی همراه است. پژوهش‌های حوزه سیستم‌های تولید ماژولار نشان می‌دهند که عامل‌های LLM می‌توانند با دریافت این بازخوردها، سیاست‌های اجرایی را به صورت برخط اصلاح کنند و بدین ترتیب رفتار سیستم را با شرایط واقعی هم‌تراز سازند. [5] این قابلیت، پیوند مستقیمی میان تصمیم‌گیری زبانی و اجرای تطبیق‌پذیر ایجاد می‌کند.

از سوی دیگر، در چارچوب‌های مبتنی بر دانش صنعتی، اجرای دستورات زبانی اغلب نیازمند تفسیر در زمینه دانش فرایندی و محدودیت‌های فیزیکی است. مدل‌های دانش صنعتی بزرگ نشان می‌دهند که ادغام LLM با دانش ساخت‌یافته صنعتی می‌تواند از تولید فرمان‌های ناسازگار با واقعیت فیزیکی جلوگیری کند و اجرای این‌تری را تضمین نماید. [4] در این حالت، لایه اجرا نه تنها به ورودی زبانی، بلکه به دانش زمینه‌ای نیز وابسته است.

در نهایت، مرورهای جامع حوزه تولید هوشمند و Industry 5.0 LLM‌ها در لایه اجرا به تدریج در حال تغییر از «کنترل مستقیم» به «پشتیبانی شناختی از اجرا» است. [7] در این رویکرد، LLM‌ها به اپراتورها و سیستم‌های کنترلی کمک می‌کنند تا تصمیمات اجرایی را بهتر درک، اصلاح و تطبیق دهند، بدون آنکه مسئولیت کامل کنترل فیزیکی را بر عهده بگیرند.

افزون بر این، یکی از چالش‌های کلیدی در لایه اجرا، مسئله ناهمخوانی سطح انتزاع زبان طبیعی با الزامات دقیق سیستم‌های کنترلی است. در حالی که زبان طبیعی ذاتاً مبهم و وابسته به زمینه است، سیستم‌های صنعتی نیازمند فرمان‌هایی با معنا و رفتار کاملاً مشخص هستند. پژوهش‌های اخیر نشان می‌دهند که استفاده از لایه‌های میانی تفسیر، مانند نمایش‌های نمادین، قوانین صریح یا مدل‌های رسمی، می‌تواند این شکاف را کاهش دهد و خروجی‌های LLM را به ورودی‌های قابل اجرا تبدیل کند.

در این راستا، ترکیب LLM‌ها با ابزارهای تحلیل صوری و بررسی اینمنی به عنوان یک رویکرد عملی مورد توجه قرار گرفته است. مطالعات نشان می‌دهند که زمانی که خروجی‌های زبانی LLM پیش از اجرا از فیلترهای اعتبارسنجی عبور می‌کنند، می‌توان احتمال بروز رفتارهای ناخواسته یا خطرناک را به طور قابل توجهی کاهش داد. این الگو، به ویژه در صنایع سنگین که خطاهای کنترلی می‌توانند پیامدهای پرهزینه یا خطرناک داشته باشند، اهمیت ویژه‌ای دارد.

از منظر تعامل انسان–ماشین، حضور LLM‌ها در لایه اجرا می‌تواند شفافیت عملکرد سیستم را افزایش دهد. در بسیاری از سامانه‌های کنترلی سنتی، منطق تصمیم‌گیری برای اپراتورها به سختی قابل درک است. اما LLM‌ها قادرند دلایل انتخاب یک سیاست اجرایی یا تغییر در رفتار سیستم را به صورت زبانی توضیح دهند. این قابلیت، نه تنها اعتماد اپراتور را افزایش می‌دهد، بلکه امکان مداخله آگاهانه انسان در شرایط بحرانی را نیز فراهم می‌سازد.

در سیستم‌های چند عامله صنعتی، اجرای هماهنگ و ظایف میان عامل‌های مختلف نیازمند تبادل اطلاعات دقیق و به موقع است. پژوهش‌ها نشان می‌دهند که زبان طبیعی، هنگامی که توسط LLM‌ها پردازش و ساخت‌یافته می‌شود، می‌تواند به عنوان یک لایه ارتباطی مشترک میان عامل‌ها عمل کند و هماهنگی اجرایی را تسهیل نماید. این موضوع بهویژه در خطوط تولید ماژولار یا کارخانه‌هایی با پیکربندی پویا اهمیت دارد، جایی که ساختار سیستم ممکن است به طور مداوم تغییر کند.

در نهایت، گسترش استفاده از LLM‌ها در لایه اجرا نشان می‌دهد که مفهوم «اجرا» در سامانه‌های رباتیک و صنعتی در حال بازتعریف است. اجرا دیگر صرفاً به معنای ارسال فرمان‌های از پیش تعیین شده نیست، بلکه به فرایندی پویا تبدیل شده است که در آن تفسیر زبان، دانش زمینه‌ای، بازخورد محیط و مداخله انسانی به صورت مستمر با یکدیگر تعامل دارند. این تحول، مسیر پژوهش‌های آینده را به سمت سامانه‌هایی سوق می‌دهد که در آن‌ها مرز میان تصمیم‌گیری و اجرا به صورت انعطاف‌پذیر و تطبیق‌پذیر مدیریت می‌شود.



تصویر 2. معماری در سطح بالا: مژول ادراک، مشاهدات شیء محور تولید می‌کند. برنامه‌ریزی RAG دانش مربوط به کار را بازیابی کرده و یک طرح JSON می‌سازد. اجراکننده نیز با بررسی‌های اینمی، دستورها را از طریق XArm SDK اعتبارسنجی و اجرا می‌کند. [9]

2-ایمنی، قابلیت اطمینان و تعامل انسان در حلقه

با ورود مدل‌های زبانی بزرگ به قلب سامانه‌های ریاضیکی، مسئله‌ی اینمی و قابلیت اطمینان به یکی از مهم‌ترین دغدغه‌های پژوهشی و مهندسی تبدیل شده است. برخلاف الگوریتم‌های کنترلی کلاسیک که رفتار آن‌ها در چارچوب مدل‌های ریاضی نسبتاً قابل پیش‌بینی است، مدل‌های زبانی بزرگ ذاتاً سامانه‌هایی احتمالی هستند که خروجی آن‌ها می‌تواند بسته به زمینه، داده‌های آموزشی و حتی جزئیات

ظریف ورودی تغییر کند. از یک سو، منبع انعطاف‌پذیری و تعیین‌پذیری بالاست و از سوی دیگر، خطری بالقوه برای ایمنی ربات در محیط‌های واقعی.

یکی از چالش‌های اصلی، پدیده‌ی «توهم زبانی»¹⁴ است؛ حالتی که در آن مدل زبانی پاسخی ظاهرآ منسجم اما نادرست یا غیرقابل اجرا تولید می‌کند. در زمینه‌ی رباتیک، چنین خطای می‌تواند به اعمال خطرناک یا آسیب‌زا منجر شود. به همین دلیل، پژوهش‌ها بر این نکته تأکید دارند که خروجی LLM نباید مستقیماً و بدون بررسی اجرا شود. در عوض، معماری‌های ایمن از لایه‌های اعتبارسنجی استفاده می‌کنند که خروجی مدل را از نظر سازگاری با محدودیت‌های فیزیکی، قوانین ایمنی و اهداف تعریف‌شده بررسی می‌کنند. [10]

نمونه‌های عملی متعددی از این رویکرد در مقالات دیده می‌شود. برای مثال، در سامانه‌های مبتنی بر ROS-LLM، دستورات زبانی ابتدا به برنامه‌های ساخت‌یافته تبدیل می‌شوند و سپس توسط مأذول‌های کنترلی بررسی می‌گردند تا از مجاز بودن اعمال اطمینان حاصل شود. [2] اگر برنامه‌ی تولیدشده با محدودیت‌های ایمنی در تضاد باشد، یا اجرا نمی‌شود یا به کاربر بازخورد داده می‌شود. این فرایند نشان می‌دهد که ایمنی نه در یک نقطه‌ی خاص، بلکه در کل زنجیره‌ی تصمیم‌گیری توزیع شده است.

تعامل انسان در حلقه یکی دیگر از راهکارهای کلیدی برای افزایش قابلیت اطمینان است. برخلاف دیدگاه‌های اولیه که استقلال کامل ربات را هدف نهایی می‌دانستند، رویکردهای جدیدتر انسان را به عنوان بخشی فعال از سامانه در نظر می‌گیرند. در این چارچوب، انسان می‌تواند در مراحل حساس تصمیم‌گیری مداخله کند، خروجی مدل را تأیید یا اصلاح نماید و حتی با زبان طبیعی رفتار ربات را هدایت کند. [2]

سامانه‌ی VoxPoser نمونه‌ی روشنی از این رویکرد است. در VoxPoser، کاربر می‌تواند در طول اجرای وظیفه با زبان طبیعی قیود جدیدی اضافه کند یا مسیر اجرای ربات را اصلاح نماید. [1] برای مثال، اگر ربات در حال قرار دادن یک شیء در موقعیتی نامناسب باشد، کاربر می‌تواند بگوید «کمی دورتر از لبه میز بگذار» و سیستم این اصلاح را در برنامه‌ی کنترلی اعمال می‌کند. این تعامل بلاذرنگ، هم ایمنی را افزایش می‌دهد و هم حس کنترل و اعتماد کاربر را تقویت می‌کند.

بازخورد زبانی انسان همچنین به عنوان ابزاری برای یادگیری و بهبود رفتار ربات عمل می‌کند. در برخی سامانه‌ها، اصلاحات کاربر ذخیره می‌شود و در تصمیم‌گیری‌های بعدی مورد استفاده قرار می‌گیرد. [2] این فرایند نشان می‌دهد که زبان می‌تواند نه تنها ابزار فرماندهی، بلکه رسانه‌ای برای انتقال تجربه و دانش ایمنی باشد. [1]

¹⁴ linguistic Hallucination

با این حال، اتکا به انسان در حلقه محدودیت‌هایی نیز دارد. افزایش بار شناختی کاربر، تأخیر در تصمیم‌گیری و وابستگی بیش از حد به نظارت انسانی از جمله چالش‌هایی هستند که در مثال‌های عملی گزارش شده‌اند. [1] این مسائل نشان می‌دهند که طراحی تعامل انسان–ربات باید بدقت انجام شود تا توازن مناسبی میان استقلال و نظارت برقرار گردد.

با افزایش پیچیدگی سامانه‌های رباتیک مبتنی بر مدل‌های زبانی بزرگ، توجه پژوهش‌ها به طور فزاینده‌ای به مسائل ایمنی، قابلیت اطمینان و پایداری عملکرد معطوف شده است. در حالی که LLM‌ها توانایی بالای در استنتاج، تعییم و تعامل زبانی دارند، رفتار آن‌ها در مواجهه با شرایط پیش‌بینی نشده، داده‌های ناقص یا ورودی‌های مبهم همچنان یک چالش اساسی محسوب می‌شود. این مسئله بهویژه در محیط‌های صنعتی و رباتیک که خطای تواند پیامدهای فیزیکی و اقتصادی جدی داشته باشد، اهمیت دوچندان می‌یابد.

پژوهش‌های اخیر نشان می‌دهند که یکی از منابع اصلی عدم اطمینان در سامانه‌های LLM محور، پدیده‌هایی مانند توهمندی زبانی، استنتاج نادرست و حساسیت به قالب‌بندی ورودی است. در پاسخ به این چالش‌ها، رویکردهای مختلفی برای افزایش قابلیت اطمینان پیشنهاد شده‌اند؛ از جمله استفاده از بازخورد انسانی در حلقه، محدودسازی فضای خروجی و ترکیب LLM‌ها با مدل‌های صریح‌تر تصمیم‌گیری. این رویکردها را می‌توان امتدادی از تلاش‌های پیشین در رباتیک ایمن دانست که اکنون در بستر مدل‌های زبانی باز تعریف شده‌اند.

در محیط‌های تولیدی، ایمنی تنها به جلوگیری از رفتارهای خطرناک محدود نمی‌شود، بلکه شامل پایداری عملکرد در طول زمان، قابلیت پیش‌بینی رفتار سیستم و انطباق با استانداردهای صنعتی نیز هست. مطالعات حوزه تولید هوشمند نشان می‌دهند که ادغام LLM‌ها با چارچوب‌های دانش صنعتی و قوانین صریح می‌تواند به کاهش نوسانات رفتاری و افزایش سازگاری با الزامات عملیاتی منجر شود. در این چارچوب، LLM‌ها بجهای یک تصمیم‌گیر مستقل، به عنوان یک مولفه پیشنهاده‌نده یا تحلیل‌گر عمل می‌کند که خروجی آن تحت نظارت لایه‌های ایمن‌تر قرار دارد.

از منظر سیستم‌های چندعامله، مسئله ایمنی به هماهنگی میان عامل‌ها نیز گسترش می‌یابد. پژوهش‌ها نشان داده‌اند که عدم هماهنگی یا سوءتفاهم زبانی میان عامل‌های مبتنی بر LLM می‌تواند به تصمیمات متناقض یا ناکارآمد منجر شود. به همین دلیل، طراحی پروتکل‌های ارتباطی ساخت‌یافته و محدودسازی تعاملات زبانی آزاد، به عنوان راهکارهایی برای افزایش قابلیت اطمینان سیستم‌های چندعامله پیشنهاد شده‌اند. این موضوع پیوند مستقیمی با مباحث مطرح شده در حوزه ارتباط عامل محور و تعامل زبان محور دارد.

یکی دیگر از ابعاد مهم ایمنی، مسئله تطبیق‌پذیری کنترل شده است. در حالی که یکی از مزایای اصلی LLM‌ها توانایی سازگاری با شرایط جدید است، این سازگاری در محیط‌های ایمن باید تحت قیود مشخصی انجام شود. پژوهش‌های اخیر نشان می‌دهند که می‌توان از سیاست‌های ایمنی، قیود رسمی و مکانیسم‌های نظارت بلاذرنگ برای هدایت فرآیند تطبیق استفاده کرد، به گونه‌ای که سیستم ضمن حفظ انعطاف‌پذیری، از محدوده‌های مجاز عملیاتی خارج نشود.

از منظر تعامل انسان–سیستم، نقش LLM‌ها در افزایش ایمنی تنها به کنترل مستقیم محدود نمی‌شود، بلکه شامل افزایش آگاهی موقعیتی اپراتور نیز هست LLM. ها می‌توانند با توضیح دلایل تصمیمات، پیش‌بینی پیامدهای احتمالی و هشدار درباره شرایط غیرعادی، به

اپراتورها کمک کنند تا مداخلات آگاهانه‌تری انجام دهند. این نقش توضیح‌دهنده، به ویژه در محیط‌های صنعتی پیچیده که تصمیم‌گیری تحت فشار زمانی انجام می‌شود، اهمیت بالایی دارد.

در نهایت، پژوهش‌های جدید نشان می‌دهند که اینمنی و قابلیت اطمینان در سامانه‌های رباتیک مبتنی بر LLM نباید به عنوان یک ویژگی افزوده در نظر گرفته شود، بلکه باید به صورت ذاتی در معماری سیستم تعییه گردد. این دیدگاه، حرکت از راهکارهای واکنشی به سمت طراحی‌های پیشگیرانه را تشویق می‌کند؛ طراحی‌هایی که از ابتدا مرزهای تصمیم‌گیری، اجرا و تعامل زبانی را با در نظر گرفتن الزامات اینمنی مشخص می‌سازند. این رویکرد، پایه‌ای مفهومی برای ادامه بحث در بخش‌های بعدی فراهم می‌کند و نشان می‌دهد که توسعه سامانه‌های LLM محور در رباتیک و صنعت، ناگزیر با بازندهای عمق در مفاهیم اینمنی و اعتماد همراه است.

8-2 محدودیت‌ها، چالش‌های باز و مسیرهای آینده

با وجود پیشرفت‌های چشمگیر در رباتیک مبتنی بر مدل‌های زبانی بزرگ، بررسی انتقادی این رویکرد نشان می‌دهد که فاصله‌ی معناداری میان قابلیت‌های فعلی و چشم‌انداز «ربات‌های عمومی و مستقل» وجود دارد. مدل‌های زبانی بزرگ را نباید به عنوان راه حلی نهایی، بلکه به عنوان یک ابزار قدرتمند اما ناقص در نظر گرفت.^[1] این فصل به جمع‌بندی محدودیت‌های اساسی، چالش‌های پژوهشی باز و مسیرهای محتمل آینده می‌پردازد و مثال‌های عملی را برای عینیت‌بخشی این بحث به کار می‌گیرد.

یکی از بنیادی‌ترین محدودیت‌ها، عدم تضمین صحت و سازگاری خروجی‌های زبانی است. همان‌طور که در فصل‌های پیش اشاره شد، مدل‌های زبانی بزرگ ممکن است پاسخ‌هایی تولید کنند که از نظر زبانی قانون کننده اما از نظر فیزیکی یا منطقی نادرست باشند. این مسئله در سامانه‌های رباتیکی، که هر خط‌ای می‌تواند پیامدهای فیزیکی داشته باشد، اهمیت دوچندان پیدا می‌کند. گزارش می‌کند که در برخی پیاده‌سازی‌های صنعتی، خروجی‌های مدل زبانی تنها پس از فیلترهای سخت‌گیرانه و آزمون‌های متعدد اجازه‌ی اجرا پیدا می‌کنند، و حتی در این شرایط نیز نظارت انسانی ضروری باقی می‌ماند.^[7] این واقعیت نشان می‌دهد که اتکای کامل به مدل‌های زبانی بزرگ در محیط‌های اینمنی حساس‌هنوز عملی نیست.

چالش مهم دیگر، تعمیم‌پذیری در دنیای واقعی است. بسیاری از مثال‌های موفق، مانند RT-Grasp SayCan یا کنترل شده یا با مجموعه‌ای محدود از اشیا آزمایش شده‌اند.^[8] [1] در مقابل، محیط‌های واقعی سرشار از تغییرات پیش‌بینی نشده، نویز حسی و تعاملات پیچیده‌اند. اگرچه مدل‌های زبانی بزرگ از نظر زبانی تعمیم‌پذیری بالایی دارند، اما این ویژگی لزوماً به تعمیم فیزیکی منجر نمی‌شود.^[7] شکاف میان «دانستن» و «انجام دادن» همچنان یکی از چالش‌های باز در هوش تجسم یافته است.

مسئله‌ی هزینه‌های محاسباتی و زیرساختی نیز نباید نادیده گرفته شود. اجرای مدل‌های زبانی بزرگ، بهویژه در سناریوهای بلاذرنگ رباتیکی، نیازمند منابع پردازشی قابل توجه است. این موضوع در مقالات بهویژه در کاربردهای صنعتی بر جسته شده است، جایی که محدودیت‌های ارزی، تأخیر و قابلیت اطمینان سخت‌افزار نقش تعیین‌کننده دارند. [7] در نتیجه، بسیاری از سامانه‌های عملی به استفاده از مدل‌های کوچک‌تر، نسخه‌های فشرده یا اجرای ابری روی آورده‌اند که هر یک مصالحه‌ایی در زمینه‌ی تأخیر و حریم خصوصی به همراه دارد.

از منظر پژوهشی، یکی از چالش‌های باز مهم، همترازی اهداف انسانی با رفتار ربات است. اگرچه زبان ابزاری قدرتمند برای بیان نیت انسانی است، اما تضمین اینکه تفسیر مدل از این نیت با انتظارات کاربر همخوانی داشته باشد، همچنان دشوار است. بازخورد زبانی انسان، نظارت در حلقه و محدودسازی خروجی‌ها تنها راهکارهای موقتی‌اند. [10] مثال‌های عملی نیز نشان می‌دهند که بدون چنین سازوکارهایی، حتی سامانه‌های پیشرفته ممکن است رفتاری غیرمنتظره از خود نشان دهند. [2]

با این حال، مسیرهای آینده‌ی روشنی نیز قابل ترسیم است. یکی از این مسیرهای توسعه‌ی مدل‌های چندوجهی قوی‌تر است که زبان، بینایی، لمس و حالت فیزیکی را به صورت یکپارچه پردازش می‌کنند. نمونه‌هایی مانند PaLM-E نشان داده‌اند که این ادغام می‌تواند شکاف میان ادراک و استدلال را کاهش دهد. [1] انتظار می‌رود که با پیشرفت این مدل‌ها، ربات‌ها بتوانند در ک غنی‌تری از محیط و پیامدهای اعمال خود داشته باشند.

مسیر دیگر، حرکت به سوی معماری‌های عامل محور و سلسله‌مراتبی است که در آن‌ها چندین عامل زبانی و کنترلی با سطوح مختلف انتزاع همکاری می‌کنند. چنین معماری‌هایی می‌توانند پیچیدگی تصمیم‌گیری را بهتر مدیریت کنند و در عین حال، ایمنی و قابلیت اطمینان را حفظ نمایند [9]، مثال‌های موجود در مقالات، هر چند در مقیاس محدود، نشان می‌دهند که این ایده از نظر عملی قابل پیاده‌سازی است.

با وجود پیشرفت‌های چشمگیر مدل‌های زبانی بزرگ در حوزه رباتیک و سیستم‌های هوشمند، مطالعات اخیر نشان می‌دهند که انتقال این مدل‌ها از محیط‌های تحقیقاتی کنترل شده به کاربردهای صنعتی واقعی همچنان با چالش‌های بنیادین همراه است. یکی از مهم‌ترین این چالش‌ها، ناهمخوانی میان توانایی تعمیم‌پذیری LLM‌ها و الزامات سخت‌گیرانه محیط‌های عملیاتی است؛ محیط‌هایی که در آن‌ها شرایط غیرقابل پیش‌بینی، داده‌های ناقص و محدودیت‌های فیزیکی به صورت هم‌زمان وجود دارند.

در بستر تولید صنعتی، این محدودیت‌ها به صورت آشکارتری نمایان می‌شوند. مرورهای جامع حوزه تولید هوشمند نشان می‌دهند که اگرچه LLM‌ها در تحلیل دانش، تعامل زبانی و پشتیبانی تصمیم‌گیری عملکرد مناسبی دارند، اما اتکای مستقیم به آن‌ها بدون لایه‌های مکمل دانش صنعتی، کنترل صریح و نظارت انسانی می‌تواند منجر به رفتارهای ناپایدار شود. این موضوع نشان می‌دهد که یکی از محدودیت‌های کلیدی LLM‌ها، فقدان در ک ذاتی از قوانین فیزیکی و قیود فرایندی است.

از منظر داده، وابستگی LLM‌ها به داده‌های آموزشی عمومی یکی دیگر از موانع مهم در کاربرد صنعتی محسوب می‌شود. پژوهش‌های اخیر تأکید می‌کنند که بسیاری از دانش‌های حیاتی در صنایع سنجین، از جمله فولاد، به صورت ضمنی، محلی و وابسته به تجربه انسانی

هستند و به راحتی در داده‌های عمومی بازنمایی نمی‌شوند. این مسئله ضرورت استفاده از مدل‌های دانش صنعتی، داده‌های اختصاصی و روش‌های تنظیم دامنه محور را برجسته می‌سازد.

در سطح سیستم، محدودیت‌های محاسباتی و الزامات بلاذرنگ نیز نقش تعیین‌کننده‌ای ایفا می‌کنند. بسیاری از LLM‌های بزرگ برای اجرا نیازمند منابع محاسباتی قابل توجهی هستند که استفاده از آن‌ها را در لایه‌های نزدیک به کنترل دشوار می‌سازد. به همین دلیل، پژوهش‌ها به سمت معماری‌های سلسله‌مراتبی، استفاده از مدل‌های کوچک‌تر تخصصی و تفکیک نقش میان استنتاج زبانی و اجرای فیزیکی حرکت کرده‌اند.

از سوی دیگر، محدودیت‌های سازمانی و انسانی نیز نباید نادیده گرفته شوند. پذیرش فناوری‌های مبتنی بر LLM در محیط‌های صنعتی مستلزم تغییر در رویه‌های کاری، آموزش نیروی انسانی و اعتمادسازی تدریجی است. مطالعات نشان می‌دهند که حتی در صورت بلوغ فنی، عدم شفافیت تصمیم‌گیری یا دشواری در تفسیر خروجی‌های سیستم می‌تواند مانع جدی برای پذیرش عملی این فناوری‌ها باشد.

در نهایت، این مجموعه محدودیت‌ها نشان می‌دهد که مسیر آینده پژوهش در حوزه LLM‌های رباتیک و صنعتی، نه در جایگزینی کامل سیستم‌های موجود، بلکه در ادغام تدریجی، کنترل شده و آگاهانه این مدل‌ها با زیرساخت‌های فعلی نهفته است. چنین مسیری مستلزم توسعه چارچوب‌هایی است که بتوانند هم زمان مزایای انعطاف‌پذیری زبانی LLM‌ها و قابلیت پیش‌بینی سیستم‌های مهندسی سنتی را در کنار یکدیگر حفظ کنند.

3 فصل

نتیجه‌گیری

مطالب ارائه شده در این گزارش نشان می‌دهد که مدل‌های زبانی بزرگ، اگرچه به تنها بی راه حل نهایی برای تمامی مسائل رباتیک نیستند، اما نقشی کلیدی در تحول معماری‌ها و روش‌های تعامل انسان و ربات ایفا می‌کنند. بررسی فصل‌های مختلف نشان داد که قدرت اصلی این مدل‌ها در توانایی در که زبان طبیعی، استخراج نیت انسانی، برنامه‌ریزی سطح بالا و ایجاد پیوند میان معنا و عمل نهفته است. این قابلیت‌ها باعث شده‌اند که ربات‌ها از ماشین‌هایی صرفاً فرمان‌پذیر به سامانه‌هایی با سطحی از استدلال و انعطاف‌پذیری شناختی تبدیل شوند.

با این حال، نتایج فضول مختلف به روشنی نشان می‌دهد که استفاده‌ی مستقیم و بدون واسطه از خروجی‌های مدل‌های زبانی در کنترل ربات‌ها می‌تواند خطرناک و غیرقابل اعتماد باشد. ماهیت احتمالی این مدل‌ها، پدیده‌هایی مانند توهمندی زبانی و نبود تضمین رسمی در صحت خروجی‌ها، ضرورت استفاده از معماری‌های ترکیبی و مازولار را برجسته می‌کند. در چنین معماری‌هایی، مدل زبانی نقش راهنمای و تصمیم‌ساز سطح بالا را بر عهده دارد، در حالی که کنترل دقیق و ایمن به مازولهای کنترلی کلاسیک سپرده می‌شود.

از منظر اینمی و قابلیت اطمینان، حضور انسان در حلقه‌ی تصمیم‌گیری همچنان اهمیت بالایی دارد. تعامل زبانی پیوسته، امکان اصلاح مسیر اجرا و ارائه‌ی بازخورد، نه تنها اینمی سیستم را افزایش می‌دهد، بلکه اعتماد کاربر به ربات را نیز تقویت می‌کند. با این وجود، وابستگی بیش از حد به نظارت انسانی می‌تواند مانع برای مقیاس‌پذیری و خودمختاری کامل ربات‌ها باشد؛ از این‌رو، ایجاد توازن میان استقلال ربات و نظارت انسان یکی از چالش‌های اصلی آینده محسوب می‌شود.

در جمع‌بندی می‌توان گفت که رباتیک مبتنی بر مدل‌های زبانی بزرگ، مسیری نویدیبخش اما همراه با چالش‌های فنی، اینمی و مهندسی است. آینده‌ی این حوزه به توسعه‌ی مدل‌های چندوجهی قوی‌تر، معماری‌های سلسله‌مراتبی و روش‌های اعتبارسنجی دقیق‌تر وابسته است. اگر این چالش‌ها به درستی مدیریت شوند، می‌توان انتظار داشت که ربات‌ها در آینده‌ای نه‌چندان دور، تعامل طبیعی‌تر، در که عمیق‌تر و عملکرد قابل اعتمادتری در محیط‌های واقعی داشته باشند و به ابزارهایی مؤثر در زندگی روزمره و صنعت تبدیل شوند.

- [1] S. Peng, “Large Language Models for , W. Jiaqi, W. Zihao, L. Yiwei, J. Hanqi p. 18, 2021. Robotics: Opportunities, Challenges, and Perspectives,”
- [2] E. M. Christopher, W. Yuhui, Hongzhan, G. Antoine, J. Gonzalez-Billandon, M. J. Wang, “ROS-LLM: A ROS framework for embodied AI with task feedback , Zimmer p. 26, 2024. and structured reasoning,”
- [3] Z. Zhengping, “LLM-MANUF: An integrated , Keqiang, D. Nan D. Kaze, Y. Bo, X. framework of Fine-Tuning large languagemodels for intelligent Decision-Making in p. 13, 2025. manufacturing,”
- [4] Z. Xiaoli, “A Review of Natural-Language-Instructed , L. Rui, G. Yibei, J. Runxiang p. 42, 2024. Robot,”
- [5] Francesco, “Parsing Natural Language O. , Danilo, M. Enrico B. Gianluca, . D. p. 4, 2019. Sentences into Robot,”
- [6] Z. Wenxiao, “Enhancing reliability in LLM- , D. Conan, . B. Thomas, B. H. Jin p. 14, 2025. integrated robotic systems: A unified approach,”
- [7] H. Andreas, “Domain-Specific Fine- , A. Benjamin, K. Urs, T. Aleksandar, K. Darko p. 5, 2021. Tuning of Large Language,”
- [8] Z. Liangjun, “RT-Grasp: Reasoning Tuning ,X. Jinxuan, J. Shiyu, L. Yutian, Z. Yuqian p. 8, 2024. Robotic Grasping via,”
- [9] C. Robin, “ARRC: Advanced Reasoning ,V. Eugene, M. Ammar Jaleel, . R. Salim Robot Control—Knowledge-Driven Autonomous Manipulation Using Retrieval- p. 8, 2025. Augmented Generation,”

- [10] D. z. Michał, “InCoRo: In-Context Learning for , Y. Z. Jiaqiang , G. C. Carla, V. David p. 20, 2024. Robotics Control with Feedback Loops,”
- [11] P. S. Yu, “Large Language Models for Robotics: , F. Zeng, W. Gan, Y. Wang, N. Liu p. 19, 2023. A Survey,”
- [12] Z. Yilun, “A ,Zhebo, L. Minghao, L. Yufeng K. Dezhang, L. Shi , . X. Zhenhua, W. p. 35, 2025. Survey of LLM-Driven AI Agent Communication: Protocols,,”
- [13] T. Jie, “Large ,L. Yiwei, Z. Huaqin, J. Hanqi, P. Yi, L. Zhengliang , W. Zihao, S. Peng . p. 52, 2024. Language Models for Manufacturing,”
- [14] p. 10, 2023. T. Dzmitry, “LLM-BRAIn: AI-driven Fast Generation of Robot,” , L. Artem
- [15] A. Rumaisa, “LLM-Driven Robots Risk Enacting Discrimination, , H. Andrew p. 59, 2025. Violence, and,”
- [16] M. George, “Multi-Agent Systems ,Ziqi, . G. X. Haoyuan, Z. Dandan C. Junhong, Y. p. 11, 2024. for Robotic Autonomy with LLMs,”
- [17] Z. Shuai, “An LLM- ,ZhenDong, N. ZhanShang, W. ShiXing, L. JunYi, C. YongTian C. p. 8, 2025. powered Natural-to-Robotic Language,”
- [18] T. Stefanie, “Plug in the Safety Chip: Enforcing , Y. Ziyi, S. R. Shreyas, S. Ankit p. 15, 2023. Constraints for LLM-driven Robot,”
- [19] B. Rogerio, “Reshaping Robot ,B. Arthur, F. Luis, . H. Sami, K. Ashish, . M. Shuang p. 7, 2022. Trajectories Using Natural Language Commands:,”
- [20] R. Nicholas, . ,T. Stefanie, K. Thomas, D. Steven, R. W. Matthew, G. B. Ashis, T. Seth p. 8, 2011. “Understanding Natural Language Commands,”
- [21] F. Yourong , “A Multimodal ,D. Wenhao, Z. Xingting, . W. Pan, Z. Jingwei, S. Yutong . p. 22, 2025. LLM-Driven Robotic Control System,”
- [22] K. Ilya, “Large Language Model-Enabled Multi-Agent , L. Jonghan, V.-H. Birgit p. 7, 2024. Manufacturing Systems,”

- [23] D. H. Hunter, "Framework for LLM , I. G. Cristian, A. D. Marcus, A. L. Tomas
p. 11, 2024. Applications in Manufacturing,"
- [24] Myriam, "Integrating Large Language Models into Digital E. , O. Chourouk
p. 29, 2025. A Systematic Review and Research Agenda," Manufacturing
- [25] S. Hanqi, "A Unified Industrial Large Knowledge Model Framework in , L. Jay
p. 7, 2024. Industry 4.0 and Smart Manufacturing,"
- [26] W. Michael, "Towards autonomous system: . , X. Yuchen, S. Manthan, J. Nasser
p. model agents," flexible modular production system enhanced with large language
9, 2023.
- [27] Q. A. Gustavo, "LLM4PLC: Harnessing Large , F. Mohamad, D. Rahul, M. Yasamin
Language Models for Verifiable Programming of PLCs in Industrial Control Systems,"
p. 12, 2024.
- [28] Jin , "Automatic MILP Model Construction for H. , P. Mingming, C. Zhendong, Y. Jie
p. 7, Multi-Robot Task Allocation Scheduling Based on Large Language Models,"
2025.
- [29] H. FELIX, "How Can Large Language Models Help Humans in Design , W. BOHAN
p. 100, 2023. And Manufacturing,"
- [30] Z. Pai, "Leveraging large language models in next , M. Yunfei, Z. Shuai, Y. Zheng .
p. 32, 2025. generation intelligent manufacturing Retrospect and prospec,"