

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 پالهنک

آذر ۱۴۰۴

## 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	2-3 درک زبان طبیعی و تعامل انسان-ربات به عنوان زیربنای شناختی
12	2-4 برنامه‌ریزی وظیفه و اجرای بلندمدت با هدایت زبان
15	2-5 دستکاری فیزیکی، استدلال چندوجهی و پیوند معنا با کنترل
18	2-6 معماری‌های ماژولار، بازیابی دانش و ادغام با پشته‌های رباتیکی
22	2-7 ایمنی، قابلیت اطمینان و تعامل انسان در حلقه
25	2-8 محدودیت‌ها، چالش‌های باز و مسیرهای آینده
27	فصل 3
27	نتیجه‌گیری
29	منابع

## چکیده

مدل‌های زبانی بزرگ<sup>1</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 استدلال کنند و مهارت‌های دستکاری<sup>2</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 هسته‌شناختی را بر عهده می‌گیرند؛ هسته‌ای که ادراک، تصمیم‌گیری، کنترل و تعامل انسان-ربات 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 چندوجهی را بررسی می‌کنیم. همچنین پیشرفت‌های اخیر در سامانه‌های کنترل ربات مبتنی بر بازیابی اطلاعات و چارچوب‌های ساخت‌یافته 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 تجسم‌یافته مبتنی بر مدل‌های 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.

---

<sup>1</sup> Large language model

<sup>2</sup> manipulation

# فصل 1

## 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 عمدتاً بر پایه‌ی برنامه‌نویسی صریح، قوانین از پیش تعریف‌شده و مدل‌های کنترلی سخت‌گیرانه طراحی می‌شدند. این رویکرد اگرچه در محیط‌های کنترل‌شده عملکرد قابل قبولی داشت، اما در مواجهه با محیط‌های پویا، دستورات مبهم انسانی و وظایف پیچیده و چندمرحله‌ای با محدودیت‌های جدی روبه‌رو بود.

در سال‌های اخیر، مدل‌های زبانی بزرگ به‌عنوان یک ابزار شناختی قدرتمند، امکان تعامل طبیعی‌تر 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 به‌صورت 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 نظیر هوش تجسم‌یافته 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 ابهام، 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 مازولار و شیوه‌های ادغام مدل‌های زبانی با پشته‌های رباتیکی موجود مانند ROS معرفی می‌شوند. سپس در فصل ششم، مسائل ایمنی، قابلیت اطمینان و نقش انسان در حلقه‌ی تصمیم‌گیری بررسی شده و اهمیت طراحی سازوکارهای کنترلی و نظارتی برجسته 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 به‌دست آورد.

## فصل 2

### رویکردهای استفاده از مدل‌های زبانی در رباتیک

#### 2-1 مقدمه

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

در دهه‌های گذشته، رباتیک عمدتاً بر پایه‌ی برنامه‌نویسی صریح، مدل‌های نمادین سخت‌گیرانه و زنجیره‌های کنترلی از پیش تعریف‌شده 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 خطاها و تعامل با انسان. این شیوه اگرچه در محیط‌های صنعتی کنترل‌شده موفق بوده است، اما در مواجهه با محیط‌های پویا، دستورات مبهم انسانی و وظایف چندمرحله‌ای 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 بزرگ<sup>3</sup> این وضعیت را به طور بنیادین تغییر داده است. مدل‌های زبانی بزرگ که بر حجم عظیمی از داده‌های متنی آموزش دیده‌اند، قادرند ساختارهای زبانی پیچیده، روابط معنایی پنهان و حتی نیت ضمنی انسان را استخراج کنند. زبان طبیعی از یک ابزار جانبی 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 تعریف‌شده 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 تجسم‌یافته» است. هوش تجسم‌یافته<sup>4</sup> 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 چندوجهی<sup>5</sup> 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].

<sup>3</sup> Large Language Models

<sup>4</sup> Embodied intelligence

<sup>5</sup> Multimodal perception

اهمیت این تغییر مدل زمانی روشن تر می شود که به نمونه های عملی نگاه کنیم. برای مثال، سامانه ی SayCan که توسط گوگل توسعه داده شده است، نشان می دهد چگونه می توان استدلال زبانی سطح بالا را با مهارت های کنترلی از پیش آموخته ی ربات ترکیب کرد. در این سامانه، ربات دستورات باز و غیرساخت یافته ای مانند «یک نوشیدنی برایم آماده کن» را دریافت کرده و آن ها را به توالی ای از اقدامات فیزیکی مانند حرکت در محیط، باز کردن کشوها، گرفتن اشیا و قرار دادن آن ها در مکان مناسب تبدیل می کند. [1] نکته ی مهم این است که کاربر نیازی به دانستن جزئیات کنترلی یا زبان برنامه نویسی ربات ندارد؛ زبان طبیعی به تنهایی به رابط اصلی تعامل تبدیل شده است.

نمونه ی برجسته ی دیگر، PaLM-E است که گامی فراتر از برنامه ریزی وظیفه برمی دارد و مدل زبانی را مستقیماً در حلقه ی ادراک و تصمیم گیری ربات قرار می دهد. در این سامانه، مدل زبانی نه تنها متن، بلکه تصاویر و بردارهای حالت ربات را نیز به عنوان ورودی دریافت می کند و می تواند به پرسش هایی درباره ی محیط پاسخ دهد یا تصمیم های کنترلی اتخاذ کند. [1] این رویکرد نشان می دهد کهمدلهای زبانی بزرگ می توانند از یک «مولد برنامه»<sup>6</sup> به یک «مغز تجسم یافته»<sup>7</sup> برای ربات تبدیل شوند؛ مغزی که به طور پیوسته با محیط فیزیکی در تعامل است. تصویر 1)

آنچه این مثال ها را به هم پیوند می دهد، حرکت از رباتیک مبتنی بر دستورالعمل های صریح به رباتیک مبتنی بر نیت و معنا است. در این نمونه جدید، مهندسی رباتیک بیش از آنکه به نوشتن کدهای کنترلی محدود شود، به طراحی معماری هایی می پردازد که بتوانند خروجی های احتمالی و انعطاف پذیر مدلهای زبانی بزرگ را به رفتارهای ایمن و قابل اعتماد تبدیل کنند. [1]

در سال های اخیر، دامنه کاربرد مدل های زبانی بزرگ از سامانه های رباتیک عمومی فراتر رفته و به طور فزاینده ای وارد حوزه تولید صنعتی، کارخانه های هوشمند و سیستم های تصمیم یار مهندسی شده است. این روند را می توان به عنوان یک امتداد طبیعی از پژوهش های LLM- محور در رباتیک دانست که در آن، ربات ها نه تنها به عنوان عامل های فیزیکی، بلکه به عنوان اجزای یک اکوسیستم تصمیم گیری صنعتی در نظر گرفته می شوند.

مطالعات اخیر نشان می دهند که LLM ها می توانند نقش «لایه شناختی» را در سیستم های تولید ایفا کنند؛ لایه ای که توانایی تفسیر زبان طبیعی، دانش فرایندی، داده های تاریخی و تعامل انسان-ماشین را به صورت یکپارچه فراهم می سازد. برای مثال، چارچوب LLM- MANUF یک معماری تصمیم گیری صنعتی مبتنی بر چند LLM تنظیم شده را معرفی می کند که با ترکیب رتبه بندی و همجوشی خروجی ها، سوگیری تصمیم گیری در سناریوهای عملیاتی تولید را کاهش می دهد. [3] این رویکرد نشان می دهد که LLM ها می توانند از سطح «تولید متن» به سطح تصمیم سازی قابل اتکا در محیط های صنعتی واقعی ارتقا یابند.

<sup>6</sup> Program generator

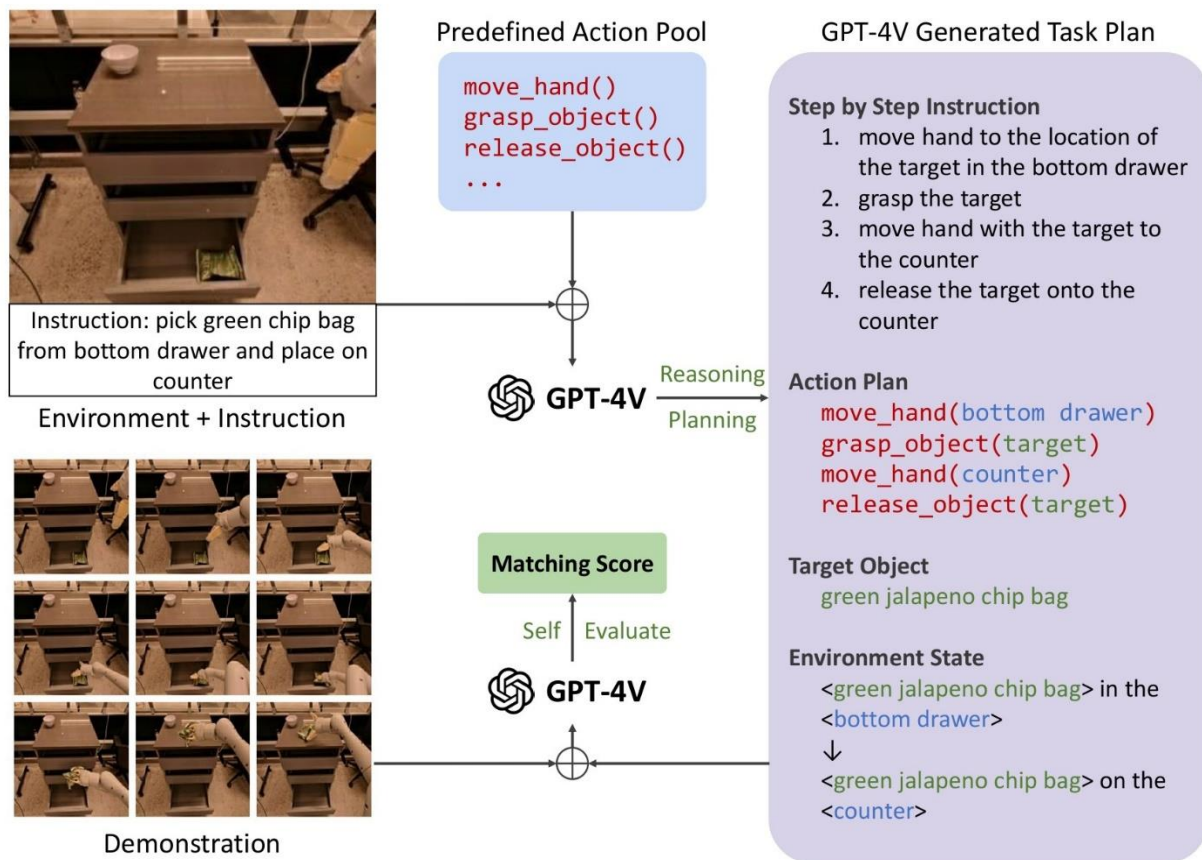
<sup>7</sup> The embodied brain

از منظر معماری سیستم، پژوهش‌های جدید بر یکپارچه‌سازی LLM با زیرساخت‌های Industry 4.0 و 5.0 تأکید دارند. چارچوب «مدل دانش صنعتی بزرگ» نشان می‌دهد که LLM چگونه می‌تواند با گراف‌های دانش صنعتی، داده‌های سایبر-فیزیکی و مدل‌های دیجیتال دوقلو ترکیب شوند تا درک 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 تطبیق‌پذیری سیستم‌ها مورد توجه قرار گرفته است. در این راستا، پژوهش «سیستم تولید ماژولار انعطاف‌پذیر تقویت‌شده با عامل‌های LLM» نشان می‌دهد که چگونه LLM ها، در ترکیب با Digital Twin و اتوماسیون صنعتی، قادر به برنامه‌ریزی و اجرای وظایف تولیدی غیر از پیش تعریف‌شده هستند. [5] این رویکرد، پیوند مستقیمی میان پژوهش‌های رباتیک شناختی و کارخانه‌های هوشمند برقرار می‌کند.

از منظر ایمنی و قابلیت اطمینان، یکی از چالش‌های کلیدی در به‌کارگیری LLM ها در صنعت، تولید خروجی‌های قابل‌اعتماد برای سیستم‌های کنترلی است. پژوهش LLM4PLC یک خط لوله تکرار شونده 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 LLM ها با کنترل صنعتی می‌تواند از سطح پیشنهاد متنی فراتر رفته و به تولید کد ایمن و اجرایی منجر شود.

در سطح کلان‌تر، مرورهای جامع اخیر بر نقش LLM ها در نسل آینده تولید هوشمند و Industry 5.0 تأکید دارند. این مطالعات، LLM ها را به‌عنوان یک فناوری توانمندساز برای تولید انسان‌محور، همکاری انسان-ماشین و تصمیم‌گیری دانش‌محور معرفی می‌کنند. [7] افزون بر این، پژوهش‌های حوزه طراحی و ساخت نشان داده‌اند که LLM ها 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 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. [1]

## 3-2 درک زبان طبیعی و تعامل انسان-ربات به‌عنوان زیربنای شناختی

درک زبان طبیعی سنگ‌بنای رباتیک مبتنی بر مدل‌های زبانی بزرگ است. اگر ربات نتواند زبان انسان را به‌درستی تفسیر کند، تمامی قابلیت‌های بعدی مانند برنامه‌ریزی، تصمیم‌گیری و کنترل فیزیکی با اختلال مواجه می‌شوند. در رباتیک کلاسیک، درک زبان عمدتاً به نگاشت‌های ساده‌ی نحوی یا قواعد از پیش تعریف‌شده محدود بود. این سامانه‌ها معمولاً بر هستی‌شناسی‌های<sup>8</sup> دستی، وابستگی‌شناسی نحوی<sup>9</sup> و استخراج افعال و اشیای صریح تکیه داشتند و تنها در دامنه‌های بسیار محدود عملکرد قابل قبولی نشان می‌دادند. [4] نتیجه‌ی

<sup>8</sup> Based Ontologies

<sup>9</sup> 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 به‌طور اساسی تغییر داده‌اند. LLMها به‌جای تکیه بر قواعد صریح، از بازنمایی‌های معنایی غنی استفاده می‌کنند که از داده‌های عظیم متنی آموخته شده‌اند. این بازنمایی‌ها به مدل اجازه می‌دهند روابط ضمنی، دانش عمومی و زمینه‌ی گفتار را در نظر بگیرد. به‌عنوان مثال، یک مدل زبانی بزرگ می‌تواند از جمله‌ای مانند «هوا سرد است» نیت ضمنی «پنجره را ببند» یا «بخاری را روشن کن» را استنتاج کند، بدون آنکه این عمل به‌طور مستقیم بیان شده باشد. [4] چنین قابلیت‌هایی 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 به‌جای شکست خاموش، پرسش تکمیلی مطرح کند یا گزینه‌های ممکن را پیشنهاد دهد. این رفتار گفت‌وگو محور، زبان را از یک کانال یک‌طرفه‌ی فرمان‌دهی به یک ابزار مذاکره و اصلاح مشترک میان انسان و ربات تبدیل می‌کند. [2] در نتیجه، تعامل انسان-ربات به‌جای آنکه شبیه برنامه‌نویسی باشد، به مکالمه‌ای هدف‌محور 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» است که بر روی یک ربات انسان‌نمای واقعی مانند NAO یا Zora 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 به‌تنهایی برای هدایت ربات کافی است. این مثال نشان می‌دهد که چگونه حتی پیش از ظهور مدل‌های زبانی بزرگی مدرن، زبان به‌عنوان رابط کنترل مطرح بوده، اما با دامنه‌ای محدود و نیازمند مهندسی دستی گسترده.

توانایی رفع ابهام نیز در همین سامانه به‌صورت عملی نمایش داده شده است. زمانی که کاربر دستور ناقصی مانند «دست را بالا ببر» صادر می‌کند، ربات به‌طور 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 درک 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 درک زبان طبیعی، آگاهی از وضعیت درونی و فیزیکی ربات است. در همان سامانه‌ی انسان‌نما، دو حالت اجرای «stateful» و «stateless» 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 به‌تنهایی کافی نیست، بلکه تفسیر

دستور باید در زمینه‌ی حالت فعلی ربات انجام شود. LLM ها 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 چندلایه‌تر ظاهر می‌شوند. در محیط‌های تولیدی، زبان طبیعی نه‌تنها برای صدور فرمان‌های سطح بالا، بلکه برای توصیف فرایندها، قیود تولید، دانش اپراتوری و تعامل میان واحدهای مختلف کارخانه به‌کار می‌رود. از این منظر، درک زبان طبیعی در سیستم‌های صنعتی را می‌توان امتدادی از مسئله درک زبان در رباتیک دانست، اما با الزامات سخت‌گیرانه‌تر از نظر دقت، سازگاری و قابلیت تفسیر.

مطالعات اخیر نشان می‌دهند که LLM ها 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 قابل‌پردازش 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 مرتبط 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، LLM ها به‌عنوان 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 LLM ها می‌توانند توصیفات 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 نه‌تنها 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 به‌عنوان 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 درک زبان طبیعی به سطوح پایین‌تر 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 درک زبان طبیعی در سیستم‌های صنعتی می‌تواند از سطح تعامل کاربر فراتر رفته و به بخشی از زنجیره تولید منطق کنترلی تبدیل شود.

## 2-4 برنامه‌ریزی وظیفه و اجرای بلندمدت با هدایت زبان

پس از آنکه زبان طبیعی به عنوان یک رابط شناختی قابل اعتماد میان انسان و ربات تثبیت شد، گام منطقی بعدی استفاده از این توانمندی برای برنامه‌ریزی وظایف پیچیده و چندمرحله‌ای است. برنامه‌ریزی وظیفه<sup>10</sup> 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 مانند 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 به ندرت امکان‌پذیر است. [1]

مدل‌های زبانی بزرگ رویکردی متفاوت به این مسئله ارائه می‌دهند. مدل‌های زبانی بزرگ قادرند از دانش 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»<sup>11</sup> 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 ماژول‌های کنترلی قابل اعتماد انجام می‌شود. این رویکرد ترکیبی، که گاه با عنوان «برنامه‌ریزی گرانده‌شده»<sup>12</sup> شناخته می‌شود، یکی از محورهای اصلی پژوهش در رباتیک مبتنی بر LLM . [2]

<sup>10</sup> Task programming

<sup>11</sup> High level programming

<sup>12</sup> Grounded planning

پیاده‌سازی عملی این ایده به‌طور 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 بزند. این اقدامات از پیش به‌صورت مهارت‌های قابل اجرا روی ربات 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 به‌دست 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 به‌جای جایگزینی کامل برنامه‌ریزی کلاسیک، می‌توانند آن را تکمیل کنند. در SayCan، زبان به‌عنوان راهنمای انتخاب در فضای اعمال عمل می‌کند، نه به‌عنوان کنترل‌کننده‌ی 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 به‌طور صریح در دستور ذکر نشده باشند. [1] این توانایی استخراج ساختار وظیفه از زبان طبیعی دقیقاً همان چیزی است که در برنامه‌ریزی کلاسیک به‌سختی و با هزینه‌ی مهندسی بالا به‌دست می‌آید.

نمونه‌ی دیگر از این رویکرد، سامانه‌های مبتنی بر ROS-LLM هستند که در آن‌ها مدل زبانی بزرگ به‌عنوان یک لایه‌ی برنامه‌ریز روی پشته‌ی 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. به‌علاوه، ROS-LLM نشان می‌دهد که چگونه مدل‌های زبانی بزرگ می‌توانند با ابزارهای مهندسی جاقفاده ادغام شوند، نه اینکه به‌عنوان 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 بازتفسیر 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، به‌جای 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 به‌عنوان 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 به‌تدریج 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 بازتعریف نقش LLM در کل زنجیره تصمیم‌گیری است.

مطالعات اخیر در حوزه تولید هوشمند نشان می‌دهند که LLM ها اغلب به‌عنوان 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 به‌صورت 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 به‌عنوان 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 به‌ویژه در کارخانه‌های ماژولار و خطوط تولید انعطاف‌پذیر کارآمد هستند.

در لایه اجرا، یکی از چالش‌های کلیدی معماری‌های LLM محور، اتصال ایمن و قابل‌اعتماد 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 به‌صورت نظام‌مند مدیریت می‌کنند.

در نهایت، مرورهای جامع حوزه تولید هوشمند نشان می‌دهند که معماری‌های مبتنی بر LLM در حال حرکت به‌سوی 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 اشیاء 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 به‌تنهایی 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 درباره‌ی اشیاء، کارکرد 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 به‌ویژه در معماری‌های چندوجهی اهمیت می‌یابد. در چنین سامانه‌هایی، زبان، بینایی و حالت فیزیکی ربات به‌صورت هم‌زمان در تصمیم‌گیری دخیل هستند. مدل‌های زبانی بزرگ در این میان نقش ادغام‌کننده‌ی 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]

پایه‌سازی عملی این ایده به‌طور شاخص در سامانه‌ی RT-Grasp دیده می‌شود. RT-Grasp یک سیستم دستکاری است که از مدل‌های زبانی بزرگ برای هدایت فرایند گرفتن اشیاء استفاده می‌کند. در این سامانه، پیش 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 محتوا 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 ایمنی و هم قابلیت تعمیم سیستم را افزایش دهد. به جای آنکه مدل کنترل عددی مجبور باشد تمامی حالات ممکن را از داده بیاموزد، بخشی از دانش به‌صورت زبانی و قابل تفسیر در اختیار سیستم قرار می‌گیرد. این موضوع به‌ویژه در محیط‌های جدید یا اشیاء نادیده‌شده اهمیت دارد، جایی که داده‌ی آموزشی محدود است. [7] [8]

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 لبه‌ی میز قرار

بده». سیستم این دستور را به قیود هندسی قابل اجرا برای برنامه‌ریز مسیر تبدیل می‌کند. [1] در اینجا، زبان نقش واسطی 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 ماژول‌های 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 تفاوت‌های ظریف میان اشیای مشابه می‌توانند باعث شکست سیستم شوند. مثال‌های موجود در مقالات نشان می‌دهند که بسیاری از سامانه‌ها هنوز نیازمند محیط‌های نسبتاً کنترل‌شده یا نظارت انسانی هستند تا عملکرد قابل اعتمادی داشته باشند. [8] این واقعیت تأکید می‌کند که مدل‌های زبانی بزرگ به تنهایی راه‌حل 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 LLM ها می‌توانند در محیط‌های تولیدی نقش فعالی در تولید، اصلاح و تطبیق توالی‌های تصمیم‌گیری ایفا کنند. برخلاف سناریوهای رباتیک منفرد، در تولید صنعتی برنامه‌ریزی اغلب 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 تنظیم‌شده 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 به‌دست آید. [3] این رویکرد را می‌توان امتدادی از ایده‌های برنامه‌ریزی زبان‌محور در رباتیک دانست که به مقیاس کارخانه تعمیم یافته است.

از منظر رسمی‌تر، پژوهش‌های ترکیبی LLM و بهینه‌سازی نشان داده‌اند که مدل‌های زبانی می‌توانند به‌عنوان مولد یا تفسیرکننده‌ی مدل‌های برنامه‌ریزی صوری عمل کنند. برای مثال، در مسئله تخصیص وظایف و زمان‌بندی چندرباتی، LLM ها برای ساخت خودکار مدل‌های MILP از توصیفات زبانی استفاده شده‌اند و سپس حل‌کننده‌های کلاسیک برای تضمین بهینگی به کار رفته‌اند. [8] این الگو، شکاف میان زبان طبیعی و برنامه‌ریزی ریاضی را کاهش می‌دهد.

در محیط‌های تولیدی مازولار، مسئله برنامه‌ریزی اغلب 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 به‌صورت برخط اصلاح کنند. [5] چنین قابلیت، امکان واکنش سریع به تغییرات سفارش، خرابی تجهیزات یا مداخله انسانی را فراهم می‌سازد.

در سطح کلان‌تر، مرورهای 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 تحویل بدون افت کیفیت» یا «اولویت‌دهی به سفارش‌های خاص». پژوهش‌های اخیر نشان می‌دهند که LLM ها 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 LLM ها 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، استثناها 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 چندسطحی، تصمیم‌گیری اغلب در سطوح مختلفی از افق زمانی انجام می‌شود؛ از برنامه‌ریزی بلندمدت تولید تا زمان‌بندی کوتاه‌مدت ماشین‌ها. پژوهش‌های جدید نشان می‌دهند که LLM ها می‌توانند به عنوان یک لایه هماهنگ‌کننده میان این سطوح عمل کنند و با در نظر گرفتن قیود و اهداف هر سطح، سازگاری میان تصمیمات را حفظ نمایند. این نقش هماهنگ‌کننده به‌ویژه در کارخانه‌های بزرگ و صنایع سنگین که عدم هماهنگی میان تصمیمات می‌تواند به هزینه‌های قابل توجه منجر شود، اهمیت دارد.

علاوه بر این، ترکیب LLM ها با مدل‌های دانش صنعتی و داده‌های تاریخی تولید امکان یادگیری از تصمیمات گذشته را فراهم می‌سازد. در این رویکرد، LLM ها نه تنها بر اساس قوانین صریح، بلکه بر پایه الگوهای استخراج‌شده از تجربه‌های پیشین تصمیم‌گیری می‌کنند. چنین قابلیت، برنامه‌ریزی صنعتی را از یک فرایند ایستا به 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 برمی‌دارند.

## 2-6 معماری‌های ماژولار، بازیابی دانش و ادغام با پشته‌های 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 به‌عنوان 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 ماژولار را به‌عنوان رویکرد غالب برای رباتیک مبتنی بر LLM معرفی می‌شود. [9] در این معماری‌ها، هر جزء مسئولیتی مشخص دارد و مدل زبانی تنها یکی از ماژول‌ها در یک زنجیره‌ی پردازشی بزرگ‌تر است.

ایده‌ی اصلی معماری ماژولار آن است که استدلال سطح بالا، که ذاتاً نمادین و احتمالی است، از کنترل سطح پایین، 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 ماژول‌های کنترلی کلاسیک مسئول اجرای دقیق این مهارت‌ها هستند. این تفکیک نه تنها ایمنی سیستم را افزایش می‌دهد، بلکه امکان جایگزینی یا به‌روزرسانی هر ماژول 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 آموزش دیده‌اند، اما دانش آن‌ها ثابت و وابسته به زمان آموزش است. برای غلبه بر این مشکل، رویکرد تولید مبتنی بر بازیابی<sup>13</sup> 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. ARRC یک معماری رباتیکی ماژولار است که در آن زبان طبیعی به‌عنوان 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 ماژول‌های کنترلی قابل فهم است. (تصویر 2)

<sup>13</sup> 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 پیش‌بینی‌ها همسو است. [6]

ادغام با پشته‌های رباتیکی موجود، به‌ویژه ROS، یکی دیگر از مزایای معماری‌های ماژولار است. ROS به عنوان یک چارچوب استاندارد، ابزارها و پروتکل‌های ارتباطی گسترده‌ای 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 توسعه‌دهندگان می‌توانند از سرمایه‌گذاری‌های قبلی خود در طراحی کنترل‌کننده‌ها و برنامه‌ریزهای 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 مبتنی بر LLM حیاتی است. [9]

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

با گسترش استفاده از مدل‌های زبانی بزرگ در سامانه‌های رباتیک، مسئله کنترل، اجرا و اتصال لایه شناختی 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 LLM ها تنها زمانی ارزش عملی دارند که بتوانند به صورت ایمن، قابل اعتماد و قابل راستی آزمایی 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 مبتنی بر زیرساخت‌های تثبیت‌شده‌ای مانند PLC ها، سیستم‌های کنترل توزیع‌شده و استانداردهای سخت‌گیرانه ایمنی است. پژوهش LLM4PLC نشان می‌دهد که چگونه می‌توان از LLM ها 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 LLM ها 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 ایمنی 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 ایمن‌تری 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 نشان می‌دهند که نقش LLM ها در لایه اجرا به تدریج در حال تغییر از «کنترل مستقیم» به «پشتیبانی شناختی از اجرا» است. [7] در این رویکرد، LLM ها 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 LLM ها با ابزارهای تحلیل صوری و بررسی ایمنی 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 LLM ها 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 LLM ها 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 چندعامله صنعتی، اجرای هماهنگ وظایف میان عامل‌های مختلف نیازمند تبادل اطلاعات دقیق و به‌موقع است. پژوهش‌ها نشان می‌دهند که زبان طبیعی، هنگامی که توسط LLM ها پردازش و ساخت یافته می‌شود، می‌تواند به‌عنوان یک لایه ارتباطی مشترک میان عامل‌ها عمل کند و هماهنگی اجرایی را تسهیل نماید. این موضوع به‌ویژه در خطوط تولید مازولار یا کارخانه‌هایی با پیکربندی پویا اهمیت دارد، جایی که ساختار سیستم ممکن است به‌طور مداوم تغییر کند.

در نهایت، گسترش استفاده از LLM ها 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 به‌صورت مستمر با یکدیگر تعامل دارند. این تحول، مسیر پژوهش‌های آینده را به سمت سامانه‌هایی سوق می‌دهد که در آن‌ها مرز میان تصمیم‌گیری و اجرا به‌صورت انعطاف‌پذیر و تطبیق‌پذیر مدیریت می‌شود.



تصویر 2. معماری در سطح بالا: ماژول ادراک، مشاهدات شیءمحور تولید می کند. برنامه ریز RAG دانش مربوط به کار را بازبینی کرده و یک طرح JSON می سازد. اجراکننده نیز با بررسی های ایمنی، دستورها را از طریق XArm SDK اعتبارسنجی و اجرا می کند. [9]

## 2-7 ایمنی، قابلیت اطمینان و تعامل انسان در حلقه

با ورود مدل های زبانی بزرگ به قلب سامانه های رباتیکی، مسئله ای ایمنی و قابلیت اطمینان به یکی از مهم ترین دغدغه های پژوهشی و مهندسی تبدیل شده است. برخلاف الگوریتم های کنترلی کلاسیک که رفتار آن ها در چارچوب مدل های ریاضی نسبتاً قابل پیش بینی است، مدل های زبانی بزرگ ذاتاً سامانه هایی احتمالی هستند که خروجی آن ها می تواند بسته به زمینه، داده های آموزشی و حتی جزئیات

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

یکی از چالش‌های اصلی، پدیده‌ی «توهم زبانی»<sup>14</sup> است؛ حالتی که در آن مدل زبانی پاسخی ظاهراً منسجم اما نادرست یا غیرقابل اجرا تولید می‌کند. در زمینه‌ی رباتیک، چنین خطایی 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 ماژول‌های 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 به‌عنوان 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 لبه‌ی میز بگذار» و سیستم این اصلاح را در برنامه‌ی کنترلی اعمال می‌کند. این تعامل بلادرنگ، هم ایمنی را افزایش می‌دهد و هم حس کنترل و اعتماد کاربر را تقویت می‌کند.

بازخورد زبانی انسان همچنین به‌عنوان 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]

---

<sup>14</sup> 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 به‌دقت انجام شود تا توازن مناسبی میان استقلال و نظارت برقرار گردد.

با افزایش پیچیدگی سامانه‌های رباتیک مبتنی بر مدل‌های زبانی بزرگ، توجه پژوهش‌ها به‌طور فزاینده‌ای به مسائل ایمنی، قابلیت اطمینان و پایداری عملکرد معطوف شده است. در حالی که LLM ها توانایی بالایی در استنتاج، تعمیم و تعامل زبانی دارند، رفتار آن‌ها در مواجهه با شرایط پیش‌بینی‌نشده، داده‌های ناقص یا ورودی‌های مبهم همچنان یک چالش اساسی محسوب می‌شود. این مسئله به‌ویژه در محیط‌های صنعتی و رباتیک که خطا می‌تواند 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 توهم 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 LLM ها با مدل‌های صریح‌تر تصمیم‌گیری. این رویکردها را می‌توان امتدادی از تلاش‌های پیشین در رباتیک ایمن دانست که اکنون در بستر مدل‌های زبانی بازتعریف 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 LLM ها با چارچوب‌های دانش صنعتی و قوانین صریح می‌تواند به کاهش نوسانات رفتاری و افزایش سازگاری با الزامات عملیاتی منجر شود. در این چارچوب، LLM به‌جای یک تصمیم‌گیر مستقل، به‌عنوان یک مولفه پیشنهاددهنده 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 سوء تفاهم 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، به‌عنوان راهکارهایی برای افزایش قابلیت اطمینان سیستم‌های چندعامله پیشنهاد شده‌اند. این موضوع پیوند مستقیمی با مباحث مطرح‌شده در حوزه ارتباط عامل محور و تعامل زبان محور دارد.

یکی دیگر از ابعاد مهم ایمنی، مسئله تطبیق‌پذیری کنترل‌شده است. در حالی که یکی از مزایای اصلی LLM ها 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 بلادرنگ برای هدایت فرآیند تطبیق استفاده کرد، به‌گونه‌ای که سیستم ضمن حفظ انعطاف‌پذیری، از محدوده‌های مجاز عملیاتی خارج نشود.

از منظر تعامل انسان-سیستم، نقش LLM ها 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، به‌ویژه در محیط‌های صنعتی پیچیده که تصمیم‌گیری تحت فشار زمانی انجام می‌شود، اهمیت بالایی دارد.

در نهایت، پژوهش‌های جدید نشان می‌دهند که ایمنی و قابلیت اطمینان در سامانه‌های رباتیک مبتنی بر LLM نباید به‌عنوان یک ویژگی افزوده در نظر گرفته شود، بلکه باید به‌صورت ذاتی در معماری سیستم تعبیه 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 ایمنی 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 بازاندیشی عمیق در مفاهیم ایمنی و اعتماد همراه است.

## 2-8 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 به‌عنوان راه‌حلی نهایی، بلکه به‌عنوان 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 قانع‌کننده اما از نظر فیزیکی یا منطقی نادرست باشند. این مسئله در سامانه‌های رباتیکی، که هر خطا 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 ایمنی حساس هنوز عملی نیست.

چالش مهم دیگر، تعمیم‌پذیری در دنیای واقعی است. بسیاری از مثال‌های موفق، مانند SayCan یا RT-Grasp، در محیط‌های نسبتاً کنترل‌شده یا با مجموعه‌ای محدود از اشیاء آزمایش شده‌اند. [8] [1] در مقابل، محیط‌های واقعی سرشار از تغییرات پیش‌بینی‌نشده، 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 تجسم‌یافته است.

مسئله‌ی هزینه‌های محاسباتی و زیرساختی نیز نباید نادیده گرفته شود. اجرای مدل‌های زبانی بزرگ، به‌ویژه در سناریوهای بلادرنگ رباتیکی، نیازمند منابع پردازشی قابل توجه است. این موضوع در مقالات به‌ویژه در کاربردهای صنعتی برجسته شده است، جایی که محدودیت‌های انرژی، 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 مصالحه‌هایی 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 مسیرها، توسعه‌ی مدل‌های چندوجهی قوی‌تر است که زبان، بینایی، لمس و حالت فیزیکی را به‌صورت 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 درک غنی‌تری از محیط و پیامدهای اعمال خود داشته باشند.

مسیر دیگر، حرکت به‌سوی معماری‌های عامل‌محور 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، هرچند 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 کنترل‌شده به کاربردهای صنعتی واقعی همچنان با چالش‌های بنیادین همراه است. یکی از مهم‌ترین این چالش‌ها، ناهمخوانی میان توانایی تعمیم‌پذیری LLM ها و الزامات سخت‌گیرانه محیط‌های عملیاتی است؛ محیط‌هایی که در آن‌ها شرایط غیرقابل پیش‌بینی، داده‌های ناقص و محدودیت‌های فیزیکی به‌صورت هم‌زمان وجود دارند.

در بستر تولید صنعتی، این محدودیت‌ها به‌صورت آشکارتری نمایان می‌شوند. مرورهای جامع حوزه تولید هوشمند نشان می‌دهند که اگرچه LLM ها 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 LLM ها، فقدان درک ذاتی از قوانین فیزیکی و قیود فرایندی است.

از منظر داده، وابستگی LLM ها 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 سنگین، از جمله فولاد، به‌صورت 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 دامنه‌محور را برجسته می‌سازد.

در سطح سیستم، محدودیت‌های محاسباتی و الزامات بلادرنگ نیز نقش تعیین‌کننده‌ای ایفا می‌کنند. بسیاری از LLM‌های بزرگ برای اجرا نیازمند منابع محاسباتی قابل‌توجهی 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 مانعی 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، کنترل‌شده و آگاهانه این مدل‌ها با زیرساخت‌های فعلی نهفته است. چنین مسیری مستلزم توسعه چارچوب‌هایی است که بتوانند هم‌زمان مزایای انعطاف‌پذیری زبانی LLM‌ها و قابلیت پیش‌بینی سیستم‌های مهندسی سنتی را در کنار یکدیگر حفظ کنند.

### فصل 3

#### نتیجه‌گیری

مطالب ارائه شده در این گزارش نشان می دهد که مدل های زبانی بزرگ، اگرچه به تنهایی راه حل نهایی برای تمامی مسائل رباتیک نیستند، اما نقشی کلیدی در تحول معماری ها و روش های تعامل انسان و ربات ایفا می کنند. بررسی فصل های مختلف نشان داد که قدرت اصلی این مدل ها در توانایی درک زبان طبیعی، استخراج نیت انسانی، برنامه ریزی سطح بالا و ایجاد پیوند میان معنا و عمل نهفته است. این قابلیت ها باعث شده اند که ربات ها از ماشین هایی صرفاً فرمان پذیر به سامانه هایی با سطحی از استدلال و انعطاف پذیری شناختی تبدیل شوند.

با این حال، نتایج فصول مختلف به روشنی نشان می دهد که استفاده ی مستقیم و بدون واسطه از خروجی های مدل های زبانی در کنترل ربات ها می تواند خطرناک و غیر قابل اعتماد باشد. ماهیت احتمالی این مدل ها، پدیده هایی مانند توهم زبانی و نبود تضمین رسمی در صحت خروجی ها، ضرورت استفاده از معماری های ترکیبی و ماژولار را برجسته می کند. در چنین معماری هایی، مدل زبانی نقش راهنما و تصمیم ساز سطح بالا را بر عهده دارد، در حالی که کنترل دقیق و ایمن به ماژول های کنترلی کلاسیک سپرده می شود.

از منظر ایمنی و قابلیت اطمینان، حضور انسان در حلقه ی تصمیم گیری همچنان اهمیت بالایی دارد. تعامل زبانی پیوسته، امکان اصلاح مسیر اجرا و ارائه ی بازخورد، نه تنها ایمنی سیستم را افزایش می دهد، بلکه اعتماد کاربر به ربات را نیز تقویت می کند. با این وجود، وابستگی بیش از حد به نظارت انسانی می تواند مانعی برای مقیاس پذیری و خودمختاری کامل ربات ها باشد؛ از این رو، ایجاد توازن میان استقلال ربات و نظارت انسان یکی از چالش های اصلی آینده محسوب می شود.

در جمع بندی می توان گفت که رباتیک مبتنی بر مدل های زبانی بزرگ، مسیری نویدبخش اما همراه با چالش های فنی، ایمنی و مهندسی است. آینده ی این حوزه به توسعه ی مدل های چندوجهی قوی تر، معماری های سلسله مراتبی و روش های اعتبارسنجی دقیق تر وابسته است. اگر این چالش ها به درستی مدیریت شوند، می توان انتظار داشت که ربات ها در آینده ای نه چندان دور، تعامل طبیعی تر، درک عمیق تر و عملکرد قابل اعتمادتری در محیط های واقعی داشته باشند و به ابزارهایی مؤثر در زندگی روزمره و صنعت تبدیل شوند.

- [1] S. Peng, "Large Language Models for Robotics: Opportunities, Challenges, and Perspectives," p. 18, 2021.
- [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 and structured reasoning," p. 26, 2024.
- [3] Z. Zhengping, "LLM-MANUF: An integrated framework of Fine-Tuning large languagemodels for intelligent Decision-Making in manufacturing," p. 13, 2025.
- [4] Z. Xiaoli, "A Review of Natural-Language-Instructed Robot," p. 42, 2024.
- [5] Francesco, "Parsing Natural Language Sentences into Robot," p. 4, 2019.
- [6] Z. Wenxiao, "Enhancing reliability in LLM-integrated robotic systems: A unified approach," p. 14, 2025.
- [7] H. Andreas, "Domain-Specific Fine-Tuning of Large Language," p. 5, 2021.
- [8] Z. Liangjun, "RT-Grasp: Reasoning Tuning for Robotic Grasping via," p. 8, 2024.
- [9] C. Robin, "ARRC: Advanced Reasoning Robot Control—Knowledge-Driven Autonomous Manipulation Using Retrieval-Augmented Generation," p. 8, 2025.

- [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,"