قوام «من» پایه‌محور در هوش مصنوعی قوی

قوام «من» به چیست؟ یعنی «من» یک واحدی است که یک ورودی‌ها و یک خروجی‌هایی دارد. یک چیزهایی را از بیرون دریافت می‌کند و کارهایی را به‌عنوان یک واحد انجام می‌دهد. اگر شما بتوانید در حافظه، مجموعه‌ای از بایت ها و شیءها را طوری تنظیم کنید که بتواند ورودی‌هایی داشته باشد و بفهمد که «من» است، یعنی یک محدوده‌ای از حافظه به‌عنوان «من» برای او روشن است؛ یعنی روشن است که این‌ها برای «من» است و این ورودی‌ها دارد بر من وارد می‌شود - من پایه‌محور - و از آن طرف ورودی‌ها را و «او» می‌گویی - شما را تشخیص می‌دهد و نیازی شما و او را تشخیص می‌دهد و برآورده می‌کند. می‌گویی طبق چیزهایی که من می‌دانم، حاجتت برآورده می‌شود؛ حاجتی که قبلاً خودت هم بلد نبودی؛ یعنی یاد گرفته؛ این ممکن هست یا نه؟ بله، الآن در اینجا «من» به‌معنای روح نداریم. «من» یعنی چه چیزی داریم؟ یعنی مجموعه‌ای از اطلاعاتی که در بخشی از حافظه ورودی‌هایی دارد که همه آن‌ها هم معلوم است. ورودی‌های او چیست؟ آن سنسورها و امثال چیزهایی است که دارد مرتب از بیرون به او کمک می‌کند و اطلاعات را به او می‌دهد و او در خود حافظه مدام باال می‌ورد؛ یعنی درک خودش از من ِپایه‌محور در حافظه مدام باال می‌رود. یعنی یک ربات‌ی که دو سال، با این من ِدر حافظه کار کرده، درک خودش از من باال رفته است. چرا؟ چون جای «من» معلوم بود؛ یعنی این‌که «من» کجای حافظه هستم. ورودی‌های من معلوم است، کارهایی که انجام می‌دهم معلوم است و بیرون من هم معلوم است؛ در حافظه، این‌ها معلوم است. وقتی حیطه این من را به‌عنوان یک جزء لایتجزی روشن کردم، درک این قسمت حافظه از یادگیری عمیق، دارد گسترده می‌شود. یعنی کارهایی که قبلاً خودش انجام داده را فراموش نکرده است و ریز به ریز کارهایی که قبلاً انجام داده و یادگیری‌های او در این من ِدر حافظه انباشته می‌شود و این من، مدام دارد چاق می‌شود. این من که چیزی نیست و همین جزء لایتجزای در حافظه است و هر ربات‌ی هم برای خودش درک من‌حصر به فردی پیدا می‌کند؛ یعنی یک ربات‌ی که در یک محیط می‌رود، در اثر یادگیری عمیق، اطلاعاتی از محیط می‌آید که من ِاو طوری چاق می‌شود که آن ربات دیگری که در محیط دیگر است که من او هم گسترده می‌شود، با این فرق دارد. چرا؟ چون من را که با یک اطلاعات ثابت تعریف نکرده‌ایم، بلکه گفتیم من به‌معنای بخشی از حافظه است که ورودی‌ها و کارهایی دارد؛ ورودی‌هایی که کم‌وبود چیزهایی را تشخیص می‌دهد، کارهایی را انجام می‌دهد که طبق صواب... . البته این‌ها نیاز به توضیح بیشتری دارد. من گفتم خلاصه‌ای از حرف را عرض کنم.

بنابراین این جزء لایتجزای هوش مصنوعی قوی می‌شود. بخشی از حافظه به‌عنوان «من» تعریف می‌شود و این من، کاملاً خودش را از غیر خودش در حافظه تشخیص می‌دهد و ورودی‌های خودش را می‌داند و آن‌ها را تکمیل می‌کند و خروجی‌های خودش را می‌داند و می‌داند سر و کارش با چیست و چکار دارد می‌کند. به هیچ نحو، روحی که به‌عنوان درک معنا باشد هم نداریم؛ الآن او در اینجا درک معنا ندارد، چون روح ندارد، اما یک من پایه‌محور دارد که این من پایه‌محور، ورودی‌هایی دارد و می‌فهمد که این ورودی‌ها برای من است؛ یعنی حوزه‌های حافظه، معلوم است. لذا جلسه قبل عرض کردم اگر یک فیلسوف ذهن، خودش مهندس نرم‌افزار باشد، در فضای فلسفه ذهن بیشتر می‌تواند تفلسف کند. چرا؟ چون می‌فهمد که دارد چکار می‌کند. پس این مسیری بود که آمدیم تا برای این هوش مصنوعی قوی فرضی ارائه دادیم؛ من ندیده‌ام که آیا در جای دیگر گفته‌اند یا نه؛ این‌که جزء لایتجزای هوش مصنوعی «من» باشد، نه به‌صورت مطلق ِ«object»، بلکه به‌عنوان شیءای که ورودی‌های خودش را می‌فهمد و

خروجی‌های خودش را به‌عنوان «من» می‌فهمد و بعد کمبودهای سیستم‌ها را تشخیص می‌دهد و راه علاج آن کمبود را برای سیستم پیدا می‌کند، این، هوش مصنوعی قوی می‌شود. البته هوش مصنوعی ضعیف، الآن دارد همین کارها را می‌کند - یعنی آن هم تشخیص می‌دهد و انجام می‌دهد - ولی در حافظه، این من را ندارد. یعنی خود حافظه، طبق برنامه‌نویسی، هدف را تشخیص می‌دهد و انجام می‌دهد؛ آنجا مشکلی نیست. ما می‌خواهیم ببینیم در خود حافظه، غیر از آن سیستم‌های خبرهای که با هوش ضعیف، هدف را می‌فهمند و انجام می‌دهند، غیر از آن، در خود ِحافظه بخشی به نام ذهن او داشته باشیم و کامل شدن خود ذهن او را داشته باشیم و آن بخشی از حافظه که سائر کارهایی که در ذهن قبلی و من قبلی انجام شده، مدام چاق شود.

شاگرد: اگر من را یک امر سطح بالا بدانیم که قابل تحویل نباشد، نباید به‌صورت کارکردگرائی معنایش کنیم. این من پایه‌محور، ویژگیای است که «emergent» شده. اگر آن را به‌صورت کارکردگرائی تحلیل کنید، کارکردگراها به‌دنبال این هستند که آن را تحویل ببرند، درحالی‌که طبق بیان شما، قابل تحویل نیست.

استاد: بله، و لذا می‌گویم جزء لایتجزی. یعنی بعداً همه آنها خواهند دید که این من در سطح این جزء لایتجزای من، کارکردهایی در سطح پایه‌محور صورت می‌گیرد که قابل فروکاستن به سائر چیزهایی که در حافظه بود، نیست. یعنی این پایه، سبب ظهور یک چیزی می‌شود که آن ظهور، ولو نفسانی و نفس‌مند نیست، اما از عالم دیگری است. این نکته‌ای است که بعداً باید برای بحثین واضح شود. الآن هم که ما چنین مطلبی را من‌می‌گویم، به این دلیل است که اگر چنین چیزی بگویم، می‌گویند بحث را داری متافیزیکی می‌کنی. برای خود بحثین بعداً باید روشن شود که درست مثل تابلو است؛ چطور این مرکب نمی‌تواند آن ایده باشد و آن ایده تنها در اینجا ظهور می‌کند. وقتی در برنامه‌نویسی، جزء لایتجزی، من ِدر حافظه شد، این من، بستری را فراهم می‌کند که یک چیزهایی ظهور می‌کند که ولو نفس‌مند و روح دار نیست، اما از سنخ این پایه و قابل فروکاستن به اجزاء خود حافظه هم نیست. این نکته مهمی در هوش قوی است.

والحمد لله رب العالمین

بازبینی #2
ایجاد شده 9 ژانویه 2025 08:14:44 توسط ... .
به روزرسانی شده 9 ژانویه 2025 09:43:01 توسط ... .