عامل‌ها؛ ایده‌ای که هفتاد سال منتظر زمان خودش ماند

مهرناز حسینی
دانشجوی مهندسی کامپیوتر
دانشکدگان فارابی دانشگاه تهران
mehrnazhosseini4@gmail.com

۱۱ دقیقه

این روزها وقتی از عامل‌های[1] هوش مصنوعی صحبت می‌کنیم، معمولاً منظورمان سامانه‌ای است که می‌تواند هدفی را دریافت کند، درباره‌ی آن تصمیم بگیرد، از ابزارهای مختلف استفاده کند و در نهایت کاری را انجام دهد [1]. شاید این مفهوم کاملاً تازه به نظر برسد، اما ایده‌ی اصلی آن نزدیک به هفتاد سال قدمت دارد.

جالب اینجاست که تاریخچه‌ی عامل‌ها، تاریخچه‌ی اختراع یک فناوری جدید نیست؛ بلکه تاریخچه‌ی برطرف شدن یک محدودیت است. هر نسل از پژوهشگران به مشکلی برخورد، راه‌حلی برای آن پیدا کرد و در همان لحظه، محدودیت تازه‌ای آشکار شد. همین زنجیره، عامل‌های امروزی را شکل داده است.

دهه‌ی ۵۰ و ۶۰ میلادی | اولین سؤال: آیا کامپیوتر می‌تواند خودش نتیجه بگیرد؟

در دهه‌ی ۱۹۵۰، برنامه‌ها چیزی بیش از مجموعه‌ای از دستورهای از پیش نوشته‌شده نبودند. اگر شرایط تغییر می‌کرد، برنامه هم ممکن بود متوقف شود؛ چون در شرایط پیش‌بینی‌نشده راهی برای تصمیم‌گیری نداشت.

در سال ۱۹۵۶، در کارگاه معروف دارتموث[2]، اصطلاح «هوش مصنوعی» برای نخستین بار مطرح شد [2]. دو سال بعد، جان مک‌کارتی در مقاله‌ی Programs with Common Sense ایده‌ای را مطرح کرد که بعدها یکی از ریشه‌های مفهوم عامل‌ها شد [3]. او برنامه‌ای فرضی به نام Advice Taker را توصیف کرد؛ برنامه‌ای که به‌جای اینکه تمام رفتارهایش از قبل نوشته شوند، بتواند مجموعه‌ای از دانسته‌ها را دریافت کند، میان آن‌ها استدلال کند و خودش به نتیجه‌ی تازه برسد.

امروز این ایده بدیهی به نظر می‌رسد، اما در آن زمان بسیار جلوتر از امکانات موجود بود. کامپیوترها توان پردازشی محدودی داشتند و روش‌های استدلال ماشینی هنوز در ابتدای راه بودند. با این حال، یک تغییر مهم اتفاق افتاده بود: برای نخستین بار «دانش» از «برنامه» جدا شد و تصمیم‌گیری به خودِ سیستم سپرده شد.

این همان بذری بود که بعدها در تمام نسل‌های عامل‌ها رشد کرد.

دهه‌ی ۷۰ و ۸۰ | آیا کامپیوتر می‌تواند مثل یک متخصص تصمیم بگیرد؟

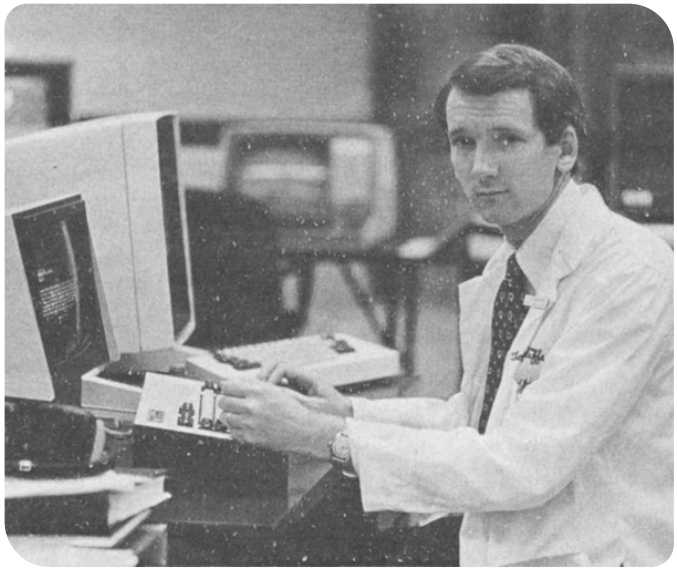
وقتی اصلِ استدلال پذیرفته شد، پرسش بعدی شکل گرفت: اگر دانش یک متخصص را به کامپیوتر منتقل کنیم، آیا می‌تواند مانند او تصمیم بگیرد؟

پاسخ آن دوران، «سیستم‌های خبره»[3] بود؛ سامانه‌هایی که دانش متخصصان را به صدها یا هزاران قانون منطقی تبدیل می‌کردند. مشهورترین نمونه، MYCIN بود که برای کمک به تشخیص عفونت‌های باکتریایی و پیشنهاد درمان توسعه یافت. در ارزیابی‌های انجام‌شده، عملکرد آن در بسیاری از سناریوها با نظر متخصصان بیماری‌های عفونی قابل مقایسه بود؛ هرچند به دلایل فنی، حقوقی و اخلاقی هرگز به‌طور گسترده وارد محیط درمان نشد [4].

نمونه‌ی موفق‌تر، XCON بود که از سال ۱۹۸۰ در شرکت Digital Equipment Corporation برای پیکربندی کامپیوترهای VAX به کار گرفته شد و با خودکار کردن فرایندی پیچیده، صرفه‌جویی اقتصادی قابل‌توجهی ایجاد کرد [5]. این موفقیت نشان داد که تصمیم‌گیری ماشینی می‌تواند از آزمایشگاه خارج شود و در صنعت ارزش واقعی خلق کند.

اما محدودیت این نسل نیز به‌تدریج آشکار شد. هر قانون باید به‌صورت دستی نوشته، بررسی و به‌روزرسانی می‌شد. هرچه مسئله بزرگ‌تر می‌شد، تعداد قوانین نیز افزایش پیدا می‌کرد و نگهداری آن‌ها دشوارتر می‌شد. سیستم‌های خبره ثابت کردند که ماشین می‌تواند مانند یک متخصص تصمیم بگیرد؛ اما نه در دنیایی که دانش آن هر روز تغییر می‌کند.

بنابراین، مسئله دیگر «باهوش‌تر کردن یک سیستم» نبود؛ بلکه پیدا کردن راهی برای تقسیم دانش و تصمیم‌گیری میان چند عامل مستقل بود.



دهه‌ی ۹۰ | اگر یک عامل کافی نیست، چند عامل چطور؟

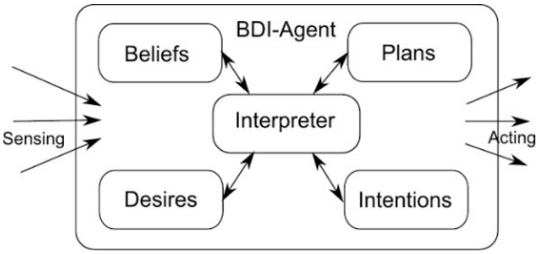
در دهه‌ی ۱۹۹۰ نگاه پژوهشگران به عامل تغییر کرد. به‌جای ساختن

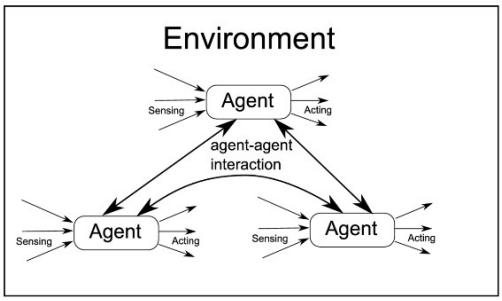
۱. Agent
۲. Dartmouth
۳. Expert Systems

یـک سیسـتم بـزرگ کـه همه‌چیـز را بدانـد، ایـده‌ی جدیـد ایـن بـود کـه هـر عامـل مسـئول بخشـی از مسـئله باشـد و بـرای رسـیدن بـه یـک هـدف مشـترک بـا عامل‌هـای دیگـر همـکاری کنـد [6].

در همیـن دوره، مـدل BDI[۴] چارچوبـی بـرای تصمیم‌گیـری عامل‌هـا ارائـه داد. بـر اسـاس ایـن مـدل، تصمیم‌گیـری عامـل بـر پایـه‌ی باورهـا، خواسـته‌ها و قصدهـای آن انجـام می‌شـود [7]. هم‌زمـان، پژوهشـگران زبان‌هـا و اسـتانداردهایی طراحـی کردنـد تـا عامل‌هـا بتواننـد اطلاعـات، درخواسـت‌ها و نتایـج را به‌صـورت سـاختاریافته بـا یکدیگـر ردوبـدل کننـد.

ایـن تغییـر، مفهوم عامـل را از یک «برنامه‌ی تصمیم‌گیـر» به عضوی از یک «سـامانه‌ی همکاری‌کننـده» تبدیـل کـرد. امـا یـک مشـکل اساسـی همچنان باقـی بـود. ایـن عامل‌هـا می‌توانسـتند بـا یکدیگـر پیـام ردوبـدل کننـد، امـا هنـوز درک عمیقـی از زبـان طبیعـی نداشـتند و تعامـل آن‌ها با دنیـای واقعی محـدود بـه قوانینـی بود کـه از قبـل برایشـان تعریف شـده بود.





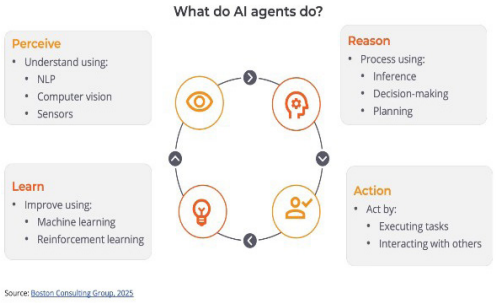
از ۲۰۲۰ تا امروز | وقتی عامل‌ها توانستند عمل کنند

ظهـور مدل‌هـای زبانـی بـزرگ[۵]، ایـن محدودیـت را تـا حـد زیـادی برطـرف کـرد. ایـن مدل‌هـا می‌توانسـتند زبـان طبیعـی را درک کننـد، اطلاعـات را خلاصـه کننـد و دربـاره‌ی مسـائل اسـتدلال کننـد؛ امـا در ابتـدا هنـوز فقـط تولیدکننـده‌ی متـن بودنـد و توانایـی انجـام عمـل نداشـتند.

نقطه‌ی عطف، معرفی روش‌هایی بود که «استدلال» و «عمل» را در یک چرخه قرار دادند [8]. از اینجا به بعد، عامل دیگر فقط پاسخ تولید نمی‌کرد؛ می‌توانست هنگام حل مسئله از ابزارهای بیرونی استفاده کند، نتیجه را ارزیابی کند، در صورت نیاز مسیرش را تغییر دهد و دوباره تصمیم بگیرد. هم‌زمان، استانداردها و زیرساخت‌هایی نیز شکل گرفتند تا ارتباط میان مدل‌های زبانی، ابزارها و منابع داده ساختاریافته‌تر و قابل‌اعتمادتر شود.

البتـه ایـن نسـل هـم کامل نیسـت. عامل‌هـای امـروزی ممکن اسـت دچار خطـا شـوند، اطلاعـات نادرسـت تولیـد کننـد یـا در اسـتفاده از ابزارهـا اشـتباه کننـد. بـه همین دلیـل، پژوهش دربـاره‌ی روش‌هـای افزایش دقـت، قابلیت اطمینـان و نظـارت بـر آن‌هـا همچنان ادامـه دارد.

شـاید ظاهـر عامل‌هـای امـروزی بـا آنچـه پژوهشـگران دهـه‌ی ۱۹۵۰ تصور می‌کردنـد کامـلاً متفـاوت باشـد؛ امـا ایـده‌ی اصلـی تغییـر چندانـی نکـرده اسـت. هنـوز هـم یـک عامـل موفق بایـد بتوانـد محیـط را درک کنـد، درباره‌ی آن تصمیـم بگیـرد و بـرای رسـیدن بـه هدفش عمل کنـد. آنچه در ایـن هفتاد سـال تغییـر کـرده، نـه خودِ ایـده، بلکـه ابزارهایی اسـت که آن رویـای قدیمی را هـر بـار یـک قـدم بـه واقعیـت نزدیک‌تـر کرده‌اند.



منابع

[1] Russell, S. J., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

[2] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

[3] McCarthy, J. (1959). Programs with common sense. In Proceedings of the Teddington Conference on the Mechanization of Thought Processes (pp. 75–91). Her Majesty's Stationery Office.

[4] Buchanan, B. G., & Shortliffe, E. H. (Eds.). (1984). Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley.

[5] Bachant, J., & McDermott, J. (1984). R1 revisited: Four years in the trenches. AI Magazine, 5(3), 21–32. https://doi.org/10.1609/aimag.v5i3.445

[6] Wooldridge, M. J., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. The Knowledge Engineering Review, 10(2), 115–152. https://doi.org/10.1017/S0269888900008122

[7] Rao, A. S., & Georgeff, M. P. (1995). BDI agents: From theory to practice. In Proceedings of the First International Conference on Multiagent Systems (pp. 312–319). AAAI Press.

[8] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. International Conference on Learning Representations. https://arxiv.org/abs/2210.03629

۴. Belief–desire–intention

۵. LLMs (Large Language Models)