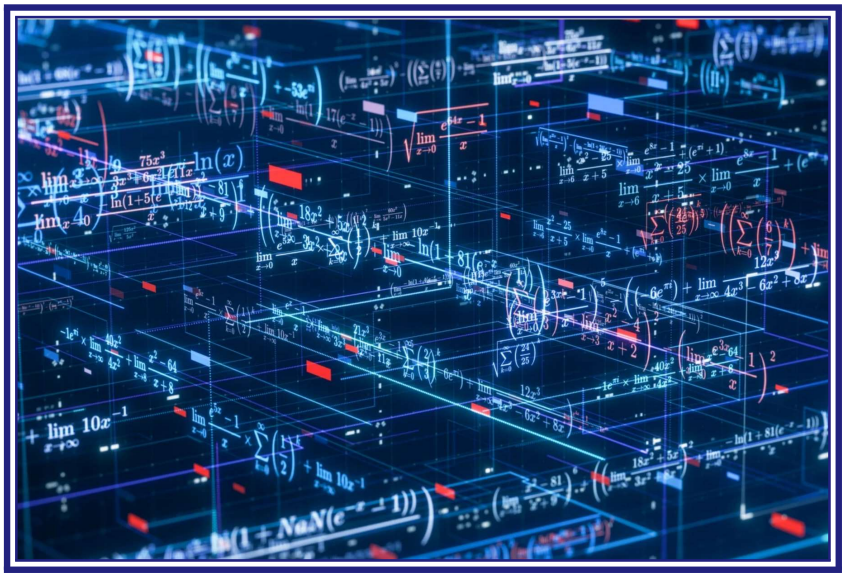
نوشته‌ها

هوش مصنوعی مدال المپیاد ریاضی را به‌دست گرفت – اما ریاضی‌دانان تحت تأثیر قرار نگرفتند

امیلی ریل **

ترجمه: سمیه رهنما *



مدل‌های هوش مصنوعی ظاهراً در حل مسائل المپیاد بین‌المللی ریاضی عملکرد خوبی داشتند، اما نحوه‌ی رسیدن آن‌ها به پاسخ‌ها یادآور این نکته است که هنوز به افرادی که ریاضی می‌خوانند، نیاز داریم.

خاطره‌ای از دوران دبیرستان

یکی از خاطرات ماندگار من از سال آخر دبیرستان، شرکت در آزمون نه‌ساعته‌ای بود که تنها شامل شش سؤال ریاضی بود. شش نفر از برترین شرکت‌کنندگان به تیم ایالات متحده برای المپیاد بین‌المللی ریاضی (IMO) راه یافتند؛ قدیمی‌ترین رقابت ریاضی برای دانش‌آموزان دبیرستانی در جهان. من نتوانستم به آنجا راه پیدا کنم، اما بعدها استاد تمام ریاضی شدم.

حضور بی‌سروصدای هوش مصنوعی در رقابت

المپیاد امسال که ماه گذشته در ساحل سان‌شاین استرالیا برگزار شد، یک حاشیه‌ی غیرمعمول داشت. در حالی که ۱۱۰ دانش‌آموز از سراسر جهان با قلم و کاغذ مشغول حل مسائل پیچیده بودند، چند شرکت فعال در حوزه‌ی هوش مصنوعی، مدل‌های جدید خود را به‌صورت آزمایشی روی نسخه‌ی دیجیتالی آزمون امتحان کردند. بلافاصله پس از مراسم اختتامیه، شرکت‌های OpenAI و Google DeepMind اعلام کردند که مدل‌هایشان موفق به حل پنج مسئله از شش مسئله شده‌اند و مدال طلای غیررسمی کسب کرده‌اند. برخی پژوهشگران مانند سباستین بوبک[1] از OpenAI این موفقیت را «لحظه‌ی فرود بر ماه» برای صنعت توصیف کردند.

آیا هوش مصنوعی جای ریاضی‌دانان را می‌گیرد؟

اما آیا واقعاً چنین است؟ آیا هوش مصنوعی قرار است جایگزین

[1] Sébastien Bubeck

ریاضی‌دانان حرفه‌ای شود؟ من هنوز منتظر اثبات این ادعا هستم. هیجان‌زدگی و تبلیغات درباره‌ی نتایج هوش مصنوعی امسال قابل درک است، زیرا المپیاد بسیار دشوار است. در سال آخر دبیرستان، من حتی حساب دیفرانسیل و جبر خطی را کنار گذاشتم تا روی مسائل سبک المپیادی تمرکز کنم که چالش بیشتری داشتند. مدل‌های پیشرفته‌ی در حال توسعه، عملکرد بسیار بهتری نسبت به مدل‌های تجاری فعلی موجود در بازار داشتند. در رقابت موازی برگزارشده توسط MathArena.ai، مدل‌هایی مانند Gemini 2.5 Pro، Grok 4، O3 High، O4-Mini High و DeepSeek R1 حتی یک پاسخ کاملاً درست ارائه نکردند. این نشان می‌دهد که مدل‌های هوش مصنوعی در حال هوشمندتر شدن هستند و قابلیت استدلال آن‌ها به‌طور چشمگیری در حال پیشرفت است.

مقایسه‌ی ناعادلانه

با این حال، من هنوز نگران نیستم. این مدل‌ها تنها در یک آزمون نمره‌ی خوبی گرفتند – همان‌طور که بسیاری از دانش‌آموزان نیز این‌طور بودند – و مقایسه‌ی مستقیم چندان منصفانه نیست. مدل‌ها اغلب از استراتژی «بهترین از میان چند» استفاده می‌کنند؛ یعنی چند پاسخ تولید کرده وخودشان به آن‌ها نمره می‌دهند، سپس بهترین را انتخاب می‌کنند. این شبیه به حالتی است که چند دانش‌آموز به‌طور مستقل کار کنند و سپس دور هم جمع شوند تا بهترین راه حل را انتخاب کرده و ارسال کنند. اگر به شرکت‌کنندگان انسانی چنین امکانی داده شود، احتمالاً نمرات آن‌ها نیز بهتر خواهد شد.

نظر ریاضی‌دانان برجسته

ریاضی‌دانان برجسته‌ی دیگر نیز نسبت به این هیاهو محتاط بوده و هشدار می دهند. ترنس تائو، دارنده‌ی مدال طلای المپیاد بین‌المللی ریاضی IMO (و در حال حاضر استاد دانشگاه کالیفرنیا، لس آنجلس UCLA)، در شبکه‌ی Mastodon اشاره کرد که توانایی‌های هوش مصنوعی به روش آزمون بستگی دارد. رئیس IMO، گرگور دولینار،[۲] نیز اظهار داشت که این سازمان نمی‌تواند روش‌های مورد استفاده توسط مدل‌های هوش مصنوعی؛ از جمله میزان محاسبات، دخالت انسانی، یا اینکه آیا نتایج قابل تکرارند را تأیید کند.

تفاوت با پژوهش‌های واقعی

به‌علاوه، به‌نظر من سؤالات المپیاد بین‌المللی ریاضی (IMO) قابل مقایسه با مسائل پژوهشی که ریاضی‌دانان حرفه‌ای سعی در پاسخ دادن به آن‌ها دارند نیستند؛ مسائلی که گاهی حل آن‌ها ۹ ساعت بلکه ۹ سال زمان می‌برد. کِوین بازارد،[۳] استاد ریاضی در کالج سلطنتی لندن، در یک انجمن آنلاین گفت: «وقتی به‌عنوان دانشجوی کارشناسی وارد کمبریج شدم و مدال طلای IMO را در دست داشتم، در موقعیتی نبودم که بتوانم به هیچ‌یک از ریاضی‌دانان پژوهشگر در آنجا کمک کنم»

تعامل با مدل‌های زبانی

امروزه پژوهش‌های ریاضی گاهی بیش از یک عمر زمان می‌برد تا تخصص لازم برای آن کسب شود. مانند بسیاری از همکارانم، وسوسه شده‌ام که «اثبات حسی»[۴] را امتحان کنم-یک گفتگوی ریاضی با یک مدل زبانی بزرگ (LLM) درست مانند گفتگو با یک همکار، با این سوال که «آیا این درست است که...؟» و مدل نیز استدلالی واضح ارائه دهد. تجربه‌ی من نشان داده که این استدلال‌ها در موضوعات استاندارد معمولاً درست هستند، اما در مرزهای دانش اغلب اشتباهات ظریفی دارند. برای مثال، همه‌ی مدل‌هایی که من امتحان کرده‌ام، اشتباهی مشابه در نظریه‌ی خودتوان‌ها در کتگوری بی‌نهایت‌بعدی ضعیف مرتکب شده‌اند؛ فرض می‌کند نظریه‌ی خودتوان‌ها در مورد کتگوری‌های بی‌نهایت‌بعدی ضعیف همان‌گونه رفتار می‌کند که در مورد کتگوری عادی [۵] رفتار می‌کند. چیزی که متخصصان انسانی در رشته من می‌دانند نادرست است.

اعتماد به اثبات‌های رسمی

من هرگز به یک مدل زبانی – که در اصل فقط پیش‌بینی‌کننده‌ی کلمات بعدی بر اساس داده‌هایش است – برای ارائه‌ی اثباتی که خودم نتوانم بررسی کنم، اعتماد نخواهم کرد. خبر خوب این است که ابزارهایی به نام «دستیار اثبات» وجود دارند؛ برنامه‌هایی که (بدون استفاده از هوش مصنوعی) بررسی می‌کنند آیا یک استدلال منطقی واقعاً ادعای مطرح‌شده را اثبات می‌کند یا نه. این ابزارها توجه ریاضی‌دانانی مانند تائو، بازارد و خود من را که خواهان اطمینان بیشتری از صحت اثبات‌های خود هستیم، جلب کرده‌اند و می‌توانند به دموکراتیزه‌کردن ریاضیات و حتی افزایش ایمنی هوش مصنوعی کمک کنند.

مثال تاریخی: رامانوجان

فرض کنید نامه‌ای با دست‌خطی ناآشنا از شهر اِرُد[۶] (شهری در تامیل نادو) در هند دریافت کنم که حاوی اثباتی ریاضی باشد. شاید ایده‌های آن درخشان باشند یا شاید بی‌معنا. باید ساعت‌ها وقت صرف کنم تا خط‌به‌خط آن را بررسی و درست و غلط بودن آن را کنترل کنم. یک ریاضی‌دان انسانی، مانند من، فقط باید معنای اصطلاحات فنی در

[2]Gregor Dolinar [3]Kevin Buzzard [4]vibe proving [5]ordinary [6]Erode

عبارت قضیه را بفهمد. اما اگر متن ریاضی به زبان رایانه‌ای مناسب نوشته شده باشد، یک دستیار اثبات می‌تواند منطق آن را برایم بررسی کند. در مورد سرینیواسا رامانوجان، نابغه‌ی ریاضی اهل اَرُد، یک متخصص وقت گذاشت تا نامه‌اش را رمزگشایی کند. در سال ۱۹۱۳، رامانوجان به ریاضی‌دان بریتانیایی جی. اچ. هاردی نامه نوشت و خوشبختانه هاردی نبوغ او را تشخیص داد و او را به کمبریج دعوت کرد و این آغاز دوران حرفه‌ای یکی از «بزرگان» ریاضی تمام دوران بود.

اثبات‌های رسمی و مدل‌های هوش مصنوعی

نکته‌ی جالب این است که برخی از مدل‌های هوش مصنوعی شرکت‌کننده در IMO پاسخ‌های خود را به زبان دستیار اثبات Lean ارائه دادند تا برنامه‌ی رایانه‌ای بتواند به‌طور خودکار خطاهای منطقی را بررسی کند. استارت‌آپی به نام Harmonic اثبات‌های رسمی برای پنج مسئله از شش مسئله منتشر کرد و ByteDance با حل چهار مسئله به سطح مدال نقره رسید. البته سؤالات باید متناسب با محدودیت‌های زبانی مدل‌ها طراحی می‌شدند و آنها هنوز به روزها زمان نیاز داشتند تا آن را حل کنند.

نتیجه‌گیری

با این حال اثبات‌های رسمی به‌طور منحصربه‌فرد قابل اعتماد هستند. در حالی‌که مدل‌های «استدلالی» با هدف تجزیه مسائل به بخش های کوچکتر و توضیح مرحله به مرحله «تفکر» خود هدایت می‌شوند، خروجی آنها به‌همان اندازه ممکن است استدلال‌هایی تولید کنند که منطقی به‌نظر برسند، اما در واقع نادرست باشد و با یک اثبات واقعی اشتباه گرفته شود. در مقابل، دستیار اثبات تنها زمانی یک اثبات را می‌پذیرد که کاملاً دقیق و منسجم باشد و هر مرحله از زنجیره فکری خود را توجیه کند. در برخی موارد، راه‌حل تقریبی کافی است، اما وقتی دقت ریاضی اهمیت دارد، باید از مدل‌های هوش مصنوعی بخواهیم که اثبات‌هایشان به‌طور رسمی قابل اعتبارسنجی باشند. در زندگی، عدم قطعیت فراوان است و اشتباه‌کردن آسان. همانطور که در دبیرستان آموختم ، یکی از بهترین چیزها در مورد ریاضی این واقعیت است که می‌توانید به‌طور قطعی ثابت کنید که برخی ایده‌ها اشتباه هستند. بنابراین خوشحالم که یک هوش مصنوعی سعی می‌کند مسائل ریاضی شخصی‌ام را حل کند، اما تنها در صورتی که نتایج به‌طور رسمی قابل تأیید باشند و ما هنوز به آن نقطه نرسیده‌ایم.

---

*Emily Riehl, AI Took on the Math Olympiad—But Mathematicians Aren't Impressed, Scientificamerican, August 7, 2025.

**دانشگاه یزد