ماهنامه
چین

جهش چین در تراشه‌های دوبعدی نسل جدید

هوش مصنوعی و صنعت تراشه

۲۸

www.techchina.ir @fanavarichin

سال سوم | شماره ۲۸ | خرداد ۱۴۰۵

نمایش برتری کوانتومی چین در یک عملکرد واقعی مبتنی بر هوش مصنوعی

دستورالعمل‌های پکن برای کنترل و توسعه عامل‌های هوش مصنوعی

بسم الله الرحمن الرحیم

درآمد:

هوش مصنوعی و صنعت تراشه به‌عنوان دو ستون اصلی انقلاب صنعتی چهارم، نقش تعیین‌کننده‌ای در آینده اقتصاد و امنیت ملی کشورها ایفا می‌کنند. چین طی سال‌های اخیر با اتخاذ رویکردی راهبردی، سرمایه‌گذاری‌های عظیم در زیرساخت‌های تحقیق و توسعه، و اجرای برنامه‌های کلان ملی همچون «برنامه توسعه نسل جدید هوش مصنوعی» و «استقلال فناوری نیمه‌رسانا»، گام‌های بلندی برای ارتقای جایگاه خود در این حوزه‌ها برداشته است.

در حوزه هوش مصنوعی، چین در زمینه‌هایی نظیر بینایی ماشین، پردازش زبان طبیعی، یادگیری عمیق و کاربردهای صنعتی و شهری، به رده‌های بالای جهانی رسیده است. همزمان، با گسترش مدل‌های زبانی پیشرفته، پلتفرم‌های بومی توانسته‌اند رقابت جدی با غول‌های فناوری غربی ایجاد کنند. دولت چین همچنین به‌طور فعال چارچوب‌های مقرراتی برای توسعه مسئولانه و ایمن هوش مصنوعی تدوین کرده است.

در بخش تراشه و نیمه‌رسانا، با توجه به فشارهای ژئوپلیتیکی و محدودیت‌های صادراتی از سوی ایالات متحده و متحدانش، چین به سرعت در حال توسعه زنجیره تأمین داخلی و فناوری‌های بومی است. پیشرفت‌های شرکت‌های داخلی چین نشان از عزم جدی این کشور برای کاهش وابستگی به واردات و دستیابی به خودکفایی فناورانه دارد.

پیگیری تحولات این حوزه در چین برای ما اهمیت مضاعف دارد؛ چرا که امکان شناسایی فرصت‌های همکاری فناورانه، تبادل دانش، و بهره‌گیری از ظرفیت‌های مشترک برای توسعه بومی فراهم می‌شود.

ماهنامه « هوش مصنوعی و صنعت تراشه چین»، با هدف ارائه تحلیلی جامع از تازه‌ترین اخبار، سیاست‌ها، دستاوردهای علمی و روندهای بازار این دو بخش کلیدی منتشر شده و می‌تواند به عنوان مرجعی معتبر برای تصمیم‌گیران و فعالان صنعتی و پژوهشی کشور مورد استفاده قرار گیرد.

عبدالرضا رحمانی فضلی

سفیر جمهوری اسلامی ایران-پکن

فهرست مطالب

چرخش سرمایه‌گذاران به سمت تولیدکنندگان تراشه چین

سهام شرکت‌های تولیدکننده تراشه در چین رشد کرد، در حالی که شرکت‌های فعال در حوزه کاربردهای هوش مصنوعی با افت مواجه شدند؛ زیرا سرمایه‌گذاران با انتشار مدل جدید استارتاپ دیپ‌سیک به سمت سهام نیمه‌هادی‌ها حرکت کردند و انتظار دارند تقاضا برای توان محاسباتی افزایش یابد.

به گزارش ساوت چاینا مورنینگ پست، سهام SMIC تا ۱۰ درصد افزایش یافت، در حالی که سهام Hua Hong Semiconductor ۱۵ درصد جهش کرد و روند صعودی گسترده‌تر در میان تراشه‌سازان چینی را ادامه داد.

در مقابل، چندین توسعه‌دهنده کاربردهای هوش مصنوعی تحت فشار قرار گرفتند. شرکت Knowledge Atlas Technology که با نام Zhipu AI نیز شناخته می‌شود، ۹.۱ درصد افت کرد و سهامMiniMax نیز ۹.۴ درصد کاهش یافت؛ زیرا مدل جدید، رقابت در بخش پرشتاب هوش مصنوعی مولد را تشدید کرد.

هر دو شرکت امسال رشد قابل‌توجهی را تجربه کرده بودند؛ رشدی که ناشی از خوش‌بینی سرمایه‌گذاران به قهرمانان بومی هوش مصنوعی چین و انتظار برای تجاری‌سازی سریع بود.

این جابه‌جایی سرمایه پس از انتشار مدل DeepSeek-V۴ از سوی شرکت مستقر در Hangzhou رخ داد؛ مدلی که به‌طور گسترده یکی از جدی‌ترین رقبای چینی برای OpenAI و Anthropic تلقی می‌شود. دیپ‌سیک اعلام کرد مدل پایه نسل جدید این شرکت از نظر رقابتی در سطح سیستم‌های بسته پیشروی آمریکایی، از جمله محصولات OpenAI و Google DeepMind قرار دارد.

گزارش‌ها حاکی از آن است که دیپ‌سیک برای نخستین دور جذب سرمایه خود با تنسنت و علی بابا در حال مذاکره بوده است.

عرضه این مدل در شرایطی انجام شد که رقابت چین و آمریکا در حوزه هوش مصنوعی شدت گرفته و واشنگتن محدودیت‌های صادراتی بر تراشه‌های پیشرفته انویدیا را سخت‌تر کرده است؛ اقدامی که شرکت‌های چینی را به تسریع توسعه و استفاده از جایگزین‌های داخلی سوق داده است.

سرمایه‌گذاران بر این نظرند که رقابت شدید در توسعه هوش مصنوعی همچنان تقاضا برای توان محاسباتی را بالا نگه خواهد داشت؛ موضوعی که به نفع تولیدکنندگان نیمه‌هادی در بخش بالادستی زنجیره خواهد بود. به گفته تحلیلگران، DeepSeek-V۴ برای تراشه‌های داخلی بهینه‌سازی شده و این امر می‌تواند جایگاه آن را در اکوسیستم فناوری چین تحت تحریم‌ها تقویت کند.

مدل دیپ‌سیک بسیار کارآمد است، بنابراین تقاضا برای استنتاج (inference) به سرعت در حال افزایش است.

این موضوع از سهام تراشه و سخت‌افزار حمایت می‌کند، زیرا شرکت‌ها همچنان باید برای اجرای این مدل‌ها در مقیاس بزرگ روی GPUها یا تراشه‌های Ascend شرکتHuawei سرمایه‌گذاری کنند.

با شتاب گرفتن زیرساخت هوش مصنوعی داخلی در چین و تداوم تقاضای قوی جهانی، این اجزای بنیادی می‌توانند از محرک‌های رشد چندساله برخوردار شوند. افزایش بهره‌وری و روند بومی‌سازی نیز فرصت‌های بخش سخت‌افزار را گسترش می‌دهد.

در همین حال، بهره‌وری هزینه‌ای این مدل بر شرکت‌های نرم‌افزاری هوش مصنوعی فشار وارد کرده است؛ زیرا کاهش موانع ورود، رقابت را تشدید و سرمایه‌گذاران را وادار به بازنگری در ارزش‌گذاری‌ها کرده است.

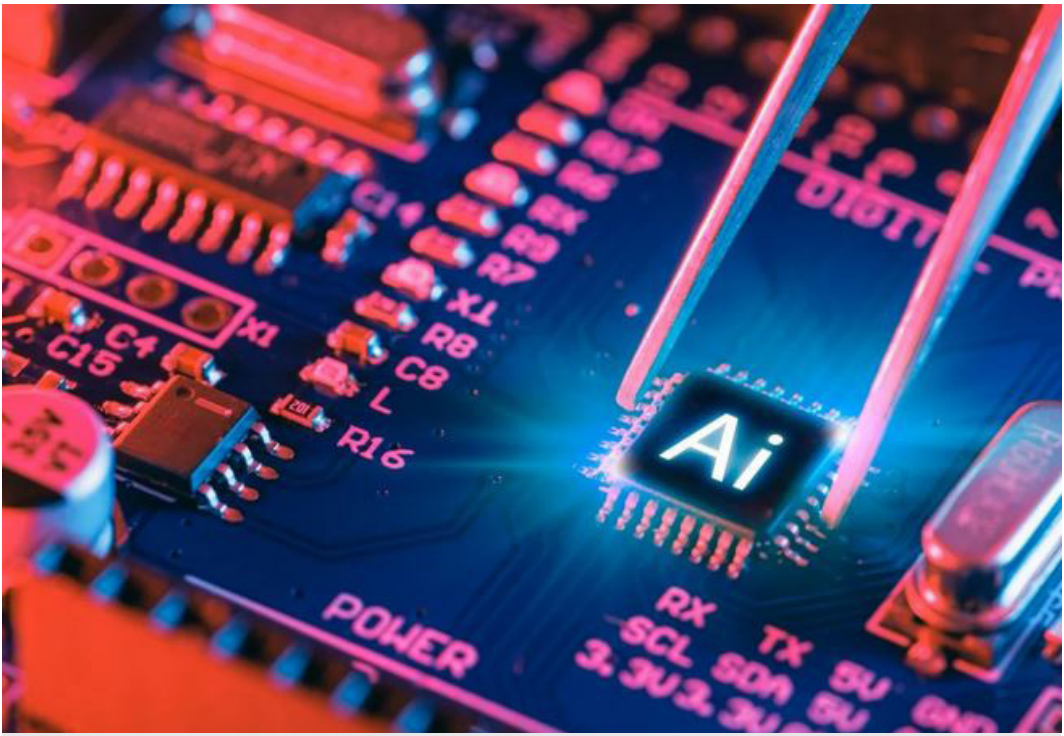
این فروش گسترده بازتاب اثر «قیمت‌گذاری مخرب» است؛ یعنی مدل‌های ارزان‌تر و پربازده، رقبا را مجبور می‌کنند قیمت‌ها را کاهش دهند یا کاربران خود را از دست بدهند.

سرمایه‌گذاران نگران‌اند که نرم‌افزار هوش مصنوعی به کالایی همگن (commoditised) تبدیل شود و ظرفیت سودآوری آن کاهش یابد.

این الگو مشابه تحولات اخیر بازار آمریکا است؛ جایی که سهام نیمه‌هادی‌ها عملکرد بهتری داشته‌اند، در حالی که شرکت‌های hyperscaler و توسعه‌دهندگان کاربردهای هوش مصنوعی عقب مانده‌اند.

این روند نشان‌دهنده انتظار سرمایه‌گذاران است که تقاضای زیربنایی برای محاسبات هوش مصنوعی همچنان قوی باقی بماند، اما افزایش هزینه‌های سخت‌افزار به‌ویژه در ویفرها و حافظه‌ها، می‌تواند حاشیه سود بازیگران پایین‌دستی را تحت فشار قرار دهد.

این نگرانی در کوتاه‌مدت معتبر است، زیرا بازیگران سخت‌افزاری بالادستی فعلاً قدرت قیمت‌گذاری دارند. اما در بلندمدت، اگر شرکت‌های پایین‌دستی نتوانند کاربردهای موفق هوش مصنوعی ایجاد کنند، کل زنجیره تأمین هوش مصنوعی نیز ممکن است با مشکل مواجه شود.



الماس در خدمت خنک‌سازی مراکز داده هوش مصنوعی چین

پژوهشگران چینی ماده جدیدی از ترکیب الماس و مس توسعه داده‌اند که به گفته آن‌ها می‌تواند راندمان خنک‌سازی مراکز داده هوش مصنوعی (AI) را تا ۸۰ درصد افزایش دهد.

به گزارش ساوت چاینا مورنینگ پست، تیم مواد کربنی کاربردی از مؤسسه فناوری صنعتی نینگبو وابسته به آکادمی علوم چین (CAS) اعلام کرده که این ماده کامپوزیتی پیش‌تر در یک گره محاسباتی هوش مصنوعی در شهر ژنگژو، استان هنان، به کار گرفته شده است.

صنعت جهانی محاسبات در مسیر دستیابی به پیشرفت‌های هوش مصنوعی با یک «دیوار حرارتی» مواجه شده است؛ مشکلی ناشی از

داغ‌شدن بیش‌ازحد، که به دلیل افزایش چگالی توان در تراشه‌های نسل جدید و در نتیجه تولید گرمای بیشتر رخ می‌دهد.

چین برای مواد پیشرفته اتلاف حرارت به واردات وابسته است، اما هزینه این مواد و توانایی آن‌ها در هدایت گرما، مستقیماً بر توانایی این کشور برای ساخت زیرساخت محاسباتی خودکفا اثر می‌گذارد.

در نتیجه، ایجاد یک صنعت مستقل و قابل‌کنترل برای مواد مدیریت حرارتی، برای بخش توان محاسباتی و رقابت‌پذیری هسته‌ای چین حیاتی تلقی می‌شود.

این تیم اعلام کرد ماده کامپوزیتی الماس-مس دارای رسانایی حرارتی بیش از هزار وات بر متر-کلوین (W/mK) است.

مس، که یکی از بهترین رساناهای گرما در میان فلزات مهندسی رایج به شمار می‌رود، رسانایی حرارتی ۴۰۰ وات بر متر-کلوین دارد. در همین حال، الماس دارای رسانایی حرارتی فوق‌العاده بالایی در حدود ۲ هزار وات بر متر-کلوین است.

استفاده از این ماده ظرفیت انتقال حرارت ماژول‌های تراشه را افزایش می‌دهد و در نتیجه عملکرد تراشه‌ها را نیز ۱۰ درصد بهبود می‌بخشد.

این تیم برای غلبه بر موانع تولید این کامپوزیت‌ها،از جمله مشکلات مربوط به پراکندگی مواد و عملیات سطحی یک فرآیند مقیاس‌پذیر و فناوری اختصاصی کامپوزیت سه‌بعدی با راندمان بالا توسعه داده است.

گروه تحقیقاتی اعلام کرد که با شرکت تولیدکننده مس جیانگ‌شی همکاری می‌کند تا تولید این فناوری را در مقیاس صنعتی پیش ببرد.

معماری‌های سنتی اتلاف حرارت و تأمین برق به محدودیت‌های فیزیکی خود نزدیک می‌شوند.

تغییر تمرکز از سرورها به زیرساخت‌ها، انتخابی اجتناب‌ناپذیر برای پشتیبانی از ساخت خوشه‌های فوق‌بزرگ محاسبات هوشمند است.

این تیم همچنین یک آزمایشگاه باز برای تجهیزات خنک‌سازی ایجاد کرده تا رابط‌های اصلی و محیط‌های آزمایشی را در اختیار شرکت‌های تراشه‌سازی، تولیدکنندگان سرور، یکپارچه‌سازها و اپراتورهای توان محاسباتی قرار دهد.

این گروه با همکاری این شرکا قصد دارد چالش‌های سازگاری را حل کند، توسعه استانداردهای صنعتی را تسریع بخشد، آستانه استقرار زیرساخت‌های پیشرفته محاسبات هوشمند بومی را کاهش دهد و ساخت یک اکوسیستم باز را ترویج کند.

آینده متفاوت شرکت‌های هوش مصنوعی چین پس از متن‌باز

در صنعت فناوری فوق‌رقابتی و بسیار سودآور چین، عرضه رایگان مدل‌های پیشرفته هوش مصنوعی در نگاه اول ممکن است غیرمنطقی به نظر برسد، اما این موضوع به یک استراتژی اصلی کسب‌وکار تبدیل شده است.

به گزارش ساوت چاینا مورنینگ پست، نوامبر گذشته در دانشگاه هنگ‌کنگ، از رئیس گروه علی‌بابا، پرسیده شد چرا این غول فناوری مدل‌های هوش مصنوعی خود را متن‌باز (open-source) کرده است.

او در برابر جمعی از دانشجویان گفت که معتقد است هوش مصنوعی متن‌باز با کاهش هزینه‌ها می‌تواند برای جهان سودمند باشد و بنابراین برای کشورهایی که از نظر سرمایه و نیروی انسانی محدود هستند، انتخابی طبیعی است.

مدل پرچمدار هوش مصنوعی این شرکت با نامQwen طی سه سال گذشته نزدیک به ۱ میلیارد بار دانلود شده و به‌طور قابل‌توجهی محبوب‌ترین خانواده مدل متن‌باز در جهان محسوب می‌شود.

از او پرسیده شد این موضوع چگونه به سودآوری منجر می‌شود و وی پاسخ داد: «ما از هوش مصنوعی پول درنمی‌آوریم»؛ منظور او خود مدل‌ها بود. این پاسخ صریح نشان‌دهنده رویکرد شرکت است: هرچند علی‌بابا از استنتاج (inference) و خدمات ابری درآمدزایی می‌کند، اما همچنان هوش مصنوعی متن‌باز را بخشی اساسی از هویت خود می‌داند.

این پرسش که چگونه می‌توان از این مدل‌ها که ذاتاً فاقد مالکیت فکری قابل تجاری‌سازی مستقیم هستند، درآمد ایجاد کرد، طی یک سال گذشته بارها مطرح شده است.

در چین، ظهور دیپ‌سیک در اوایل سال گذشته موجی از مدل‌های متن‌باز ایجاد کرد و استارت‌آپ‌هایی مانند مینی‌مکس و ژیپو از این موج استفاده کردند و حتی به عرضه‌های بزرگ در بورس هنگ‌کنگ رسیدند.

اما این پرسش در فصل گزارش‌های مالی اخیر دوباره مطرح شده است، زیرا برخی شرکت‌ها از جمله علی‌بابا و ژیپو برخی از جدیدترین مدل‌های خود را به‌صورت بسته (closed-source) عرضه کرده‌اند.

اما به گفته چند تحلیلگر، متن‌باز بودن و درآمدزایی الزاماً در تضاد نیستند.

نکته کلیدی این است که مدل‌های هوش مصنوعی فقط یک لایه از «پشته (stack)» محصول هستند.
بیشتر شرکت‌های هوش مصنوعی از یک رویکرد ترکیبی استفاده می‌کنند که عناصر متن‌باز و بسته را هم‌زمان در بر می‌گیرد.

یک الگوی قدیمی

مدل‌های هوش مصنوعی از طریق فرآیندی به نام استنتاج (inference) در اختیار کاربران قرار می‌گیرند؛ در این فرآیند، پردازنده‌های گرافیکی قدرتمند (GPU) برای اجرای مدل استفاده می‌شوند. اگرچه مدل‌های متن‌باز رایگان هستند، اما اجرا و تنظیم آن‌ها روی سخت‌افزار محلی می‌تواند هزاران دلار هزینه داشته باشد و نیازمند تخصص فنی بالاست. در عوض، توسعه‌دهندگان مدل معمولاً خود خدمات استنتاج را ارائه می‌دهند؛ نوعی «درآمدزایی غیرمستقیم».
به همین ترتیب، شرکت‌های ابری چین مانند علی‌بابا و تنسنت از طریق اجاره GPUها از مدل‌های متن‌باز درآمد کسب می‌کنند.
ارائه‌دهندگان استنتاج همچنین می‌توانند خدمات ویژه‌ای مانند امنیت و بهینه‌سازی به شرکت‌ها بفروشند.
این رویکرد در گذشته نیز موفق بوده است؛ معروف‌ترین نمونه آن سیستم عامل اندروید گوگل است که متن‌باز است و به اکوسیستم غالب موبایل در جهان تبدیل شد و در نهایت به تقویت موتور جست‌وجوی گوگل کمک کرد.
علی‌بابا کلاد نیز نمونه دیگری است. این شرکت حدود ۳۶ درصد بازار ابری چین را در اختیار دارد و از سال ۲۰۲۴ (یک سال پس از متن‌باز کردن Qwen) درآمد ابری آن موتور اصلی رشد شرکت شده است.

محدودیت‌های مدل متن‌باز

با این حال، این استراتژی محدودیت‌هایی دارد. رقابت شدید داخلی و همچنین رقابت با مدل‌های آمریکایی باعث فشار نزولی بر قیمت‌ها شده است.

بر اساس گزارش مالی علی‌بابا، حاشیه سود بخش ابری این شرکت همچنان در محدوده تک‌رقمی است. در مقابل، شرکت‌های آمریکایی مانند OpenAI و Anthropic که مدل‌های بسته دارند، حاشیه سود ۴۰ تا ۵۰ درصدی دارند، زیرا کنترل انحصاری بر استنتاج به آن‌ها قدرت قیمت‌گذاری بیشتری می‌دهد.

در آمریکا، استراتژی ارزش‌محور وجود دارد که هر دلار هزینه‌شده باید بیشترین درآمد را از هر کاربر تولید کند. اما در چین، استراتژی مبتنی بر حجم است؛ یعنی مدل‌ها تا حد ممکن ارزان و فراوان عرضه می‌شوند و درآمد به‌صورت تدریجی و در لایه‌های مختلف کسب می‌شود.

در بازاری که مدل‌ها کالایی (commoditised) می‌شوند، قیمت و بهره‌وری هزینه به عامل تعیین‌کننده تبدیل می‌شود.

این وضعیت شباهت زیادی به جنگ قیمتی در صنعت خودروهای برقی و پنل‌های خورشیدی چین دارد. شرکت‌های مینی‌مکس و ژیپو در گزارش‌های مالی خود برای سال منتهی به دسامبر ۲۰۲۵، با وجود رشد سه‌رقمی درآمد، افزایش زیان را گزارش کردند؛ زیرا رقابت شدید باعث کاهش قیمت‌ها شده است.

الگوی جدید

ظهور ابزار عامل هوش مصنوعی متن‌بازOpenClaw در ابتدای سال، محاسبات استراتژیک شرکت‌های چینی را تغییر داده است.

این ابزار می‌تواند به‌صورت خودکار وظایف را برای کاربران انجام دهد و کاربران می‌توانند انتخاب کنند از کدام مدل استفاده شود. این فرآیند مصرف گسترده‌تری از مدل‌ها ایجاد می‌کند و می‌تواند تقاضا را به‌شدت افزایش دهد.

این تحول برای شرکت‌های ابری مانند علی‌بابا و شرکت‌های AI مانند مینی‌مکس و ژیپو مثبت تلقی می‌شود، زیرا افزایش تقاضا امکان افزایش قیمت و بهبود حاشیه سود را فراهم می‌کند.

همچنین می‌تواند به تغییر مدل درآمدی از «غیرمستقیم» به «مستقیم» کمک کند.

به همین دلیل، تمرکز صنعت به سمت مدل به‌عنوان سرویس (MaaS) در حال حرکت است؛ مدلی که بر اساس مصرف توکن (واحد اصلی استفاده از AI) هزینه دریافت می‌کند.

مدیرعامل علی‌بابا، گفته است که هدف شرکت تبدیل MaaS به مدل اصلی درآمدی هوش مصنوعی است، هرچند اکنون کمتر از ۱۰ درصد درآمد ابری AI را تشکیل می‌دهد. او پیش‌بینی کرده این مدل ظرف ۵ سال بزرگ‌ترین منبع درآمد ابری شرکت خواهد شد.

در همین حال، مصرف توکن در سه ماه اول سال شش برابر شده است.

رقابت و تغییرات جدید

تقریباً همه شرکت‌های هوش مصنوعی چین طرح‌های پرداخت بر اساس استفاده (coding plans) معرفی کرده‌اند، اما به دلیل کمبود جهانی GPU، این طرح‌ها به سختی قابل دسترسی شده‌اند و برخی شرکت‌ها حتی آن‌ها را لغو کرده‌اند.

برای ژیپو، افزایش استفاده از API در سه‌ماهه اول ۴۰۰درصد رشد

داشته، در حالی که قیمت هر درخواست ۸۳ درصد افزایش یافته است. مدیرعامل این شرکت گفته است بسیاری از مشتریانی که قبلاً مدل‌های متن‌باز را به‌صورت محلی اجرا می‌کردند، اکنون به APIهای ابری مهاجرت کرده‌اند و این باعث شده درآمد شرکت به‌طور عمده بر پایه API و توکن باشد.

رویکردهای ترکیبی و آینده

مدلOpenClaw باعث تغییر استراتژی شده و ممکن است صنعت را پایدارتر کند، اما رقابت همچنان شدید است.

شرکت‌های کوچک‌تر مانند ژیپو و مینی‌مکس با چالش بزرگی روبه‌رو هستند، زیرا مدل‌های متن‌باز آن‌ها توسط غول‌هایی مانند علی‌بابا و تنسنت در خدمات خود استفاده می‌شود.

مینی‌مکس اخیراً مدل جدید خود را متن‌باز کرده اما استفاده تجاری از آن را بدون مجوز ممنوع کرده است؛ اقدامی که با واکنش منفی جامعه متن‌باز مواجه شد.

برخی تحلیلگران می‌گویند این محدودیت‌ها ممکن است محبوبیت مدل‌های چینی را در جامعه جهانی توسعه‌دهندگان کاهش دهد.

با این حال، کارشناسان می‌گویند چنین محدودیت‌هایی در صنعت متن‌باز بی‌سابقه نیست و بخشی از تلاش برای حفظ پایداری تجاری است.

در نهایت، آینده مدل متن‌باز در چین هنوز نامشخص است، به‌خصوص با توجه به محدودیت توان محاسباتی نسبت به آمریکا و رقابت شدید مدل‌های بسته.

همچنین همه نگاه‌ها به نسخه بعدی دیپ‌سیک است؛ مدلی که ممکن است دوباره موج متن‌باز را تقویت کند یا مسیر صنعت را تغییر دهد.

نمایش برتری کوانتومی چین در یک عملکرد واقعی مبتنی بر هوش مصنوعی

یک مرکز محاسباتی هوش مصنوعی که بتواند الگوهای آب‌وهوایی را چندین هفته جلوتر پیش‌بینی کند، معمولاً قیمتی در حدود ۱۰۰ میلیون دلار یا بیشتر دارد.

به گزارش ساوت چاینا مورنینگ پست، اکنون یک سیستم کوانتومی در مقیاس کوچک می‌تواند با هزینه‌ای کمتر از ۱ درصد این مبلغ، عملکردی بهتر از چنین مراکزی ارائه دهد.

این یافته‌ها پرسش‌هایی درباره اقتصاد بلندمدت رقابت جهانی زیرساخت‌های هوش مصنوعی ایجاد می‌کند. اگر سیستم‌های کوانتومی

فشرده بتوانند در برخی وظایف عملکرد رقابتی ارائه دهند، آیا ممکن است مراکز داده عظیم امروزی که هزینه آن‌ها به تریلیون‌ها دلار می‌رسد، به‌زودی منسوخ شوند؟

این سیستم پیشرفته که بر پایه ۹ اسپین کوانتومی برهم‌کنش‌گر ساخته شده است، در وظایف پیش‌بینی چندمرحله‌ای آب‌وهوا عملکردی برابر یا بهتر از یک شبکه مخزنی کلاسیک با ۱۰ هزار گره (node) نشان داد. این نتایج در ۲۵ مارس توسط تیمی مشترک از دانشگاه علوم و فناوری چین و دانشگاه چینی هنگ‌کنگ گزارش شد.

در ایالات متحده، سرمایه‌گذاری دولت و بخش خصوصی در پیش‌بینی آب‌وهوای مبتنی بر هوش مصنوعی به صدها میلیون دلار رسیده است. اداره ملی اقیانوسی و جوی آمریکا (NOAA) نزدیک به ۱۰۰ میلیون دلار برای ارتقای سیستم ابررایانه‌ای Rhea سرمایه‌گذاری کرده است، در حالی که قانون TAME حدود ۱۸۸ میلیون دلار بودجه طی پنج سال برای پژوهش‌های آب‌وهوایی مبتنی بر هوش مصنوعی اختصاص داده است.

شرکت‌های خصوصی مانند Tomorrow.io بیش از ۱۷۵ میلیون دلار سرمایه جذب کرده‌اند و غول‌های فناوری مانند گوگل، مایکروسافت و انویدیا نیز همچنان سرمایه‌گذاری گسترده‌ای در مدل‌های پیچیده آب‌وهوا مبتنی بر خوشه‌های عظیم محاسباتی انجام می‌دهند.

در مقابل، سیستم ۹ کیوبیتی (qubit) مبتنی بر رزونانس مغناطیسی هسته‌ای (NMR) که در این پژوهش استفاده شده، بسیار کوچک‌تر و بالقوه بسیار کم‌هزینه‌تر است. برای مقایسه، یک پردازنده کوانتومی ۹ کیوبیتی توسعه‌یافته توسط شرکت Rigetti Computing حدود ۹۰۰

هزار دلار قیمت‌گذاری شده است که نشان‌دهنده مقیاس بسیار سبک این آزمایش است.

اگرچه این نتایج هنوز در مراحل ابتدایی هستند، اما پرسش‌های گسترده‌ای درباره اقتصاد بلندمدت مراکز داده عظیم هوش مصنوعی مطرح می‌کنند، از جمله پروژه‌های جاه‌طلبانه‌ای مانند زیرساخت محاسباتیStargate ، اگر سیستم‌های کوانتومی کوچک بتوانند در برخی وظایف عملکردی رقابتی ارائه دهند.

برای نخستین بار آزمایش‌ها نشان داده‌اند که در مواجهه با وظایف پیش‌بینی سری‌های زمانی واقعی، عملکرد یادگیری ماشین کوانتومی می‌تواند از مدل‌های شبکه عصبی کلاسیک فراتر رود.

اگرچه رایانش کوانتومی پیش‌تر در مسائل بسیار تخصصی مزیت‌هایی نشان داده بود، اما تبدیل این مزیت به کاربردهای واقعی همواره یک چالش جهانی بوده است.

نمایش‌های قبلی شرکت گوگل و رایانه کوانتومی Jiuzhang چین، «برتری کوانتومی» را در وظایفی مانند نمونه‌گیری مدار تصادفی و نمونه‌گیری بوزون نشان داده بودند، اما این مسائل کاربرد عملی مستقیمی نداشتند.

در این پژوهش، به‌جای استفاده از مدارهای عمیق و شکننده کوانتومی، تیم چینی از چارچوب «محاسبات مخزنی» (reservoir computing) استفاده کرد و داده‌های سری زمانی را در شبکه‌ای از اسپین‌های کوانتومی برهم‌کنش‌گر رمزگذاری کرد. این سیستم از دینامیک طبیعی درهم‌تنیدگی کوانتومی برای پردازش اطلاعات استفاده می‌کند و نیاز به طراحی مدارهای پیچیده را حذف می‌کند.

نکته مهم این است که اثراتی که معمولاً «نویز» محسوب می‌شوند، در

این سیستم به منابع محاسباتی تبدیل شده‌اند و نوعی حافظه کوتاه‌مدت ضروری برای پیش‌بینی سری‌های زمانی ایجاد کرده‌اند.

اگر یک شبکه عصبی کلاسیک را مانند کتابخانه‌ای عظیم در نظر بگیریم که هر کتاب آن با دقت توسط کتابدار طبقه‌بندی شده است، مخزن کوانتومی بیشتر شبیه فنجان قهوه‌ای هم‌زده است که دینامیک پیچیده و گردابی آن به‌طور طبیعی الگوها را بدون نیاز به دخالت کتابدار رمزگذاری می‌کند.

پژوهشگران همچنین اشاره کردند که رویکرد محاسبات مخزنی، با حذف نیاز به مدارهای عمیق کوانتومی، سربار محاسباتی را کاهش داده و پیچیدگی سخت‌افزاری و مصرف انرژی را پایین آورده است.

طبق گفته شرکت SpinQ Technology، این موضوع تا حدی به این دلیل است که روش رزونانس مغناطیسی هسته‌ای به تجهیزات گران‌قیمت سرمایش فوق‌العاده مانند یخچال‌های رقیق‌کننده (dilution refrigerators) نیاز ندارد.

در آزمون‌های استانداردNARMA ، یک معیار برای پیش‌بینی سری‌های زمانی مدل کوانتومی خطای پیش‌بینی را یک تا دو مرتبه بزرگی کاهش داد. این نتایج پایه عملکرد قوی آن در پیش‌بینی‌های واقعی را تشکیل می‌دهد.

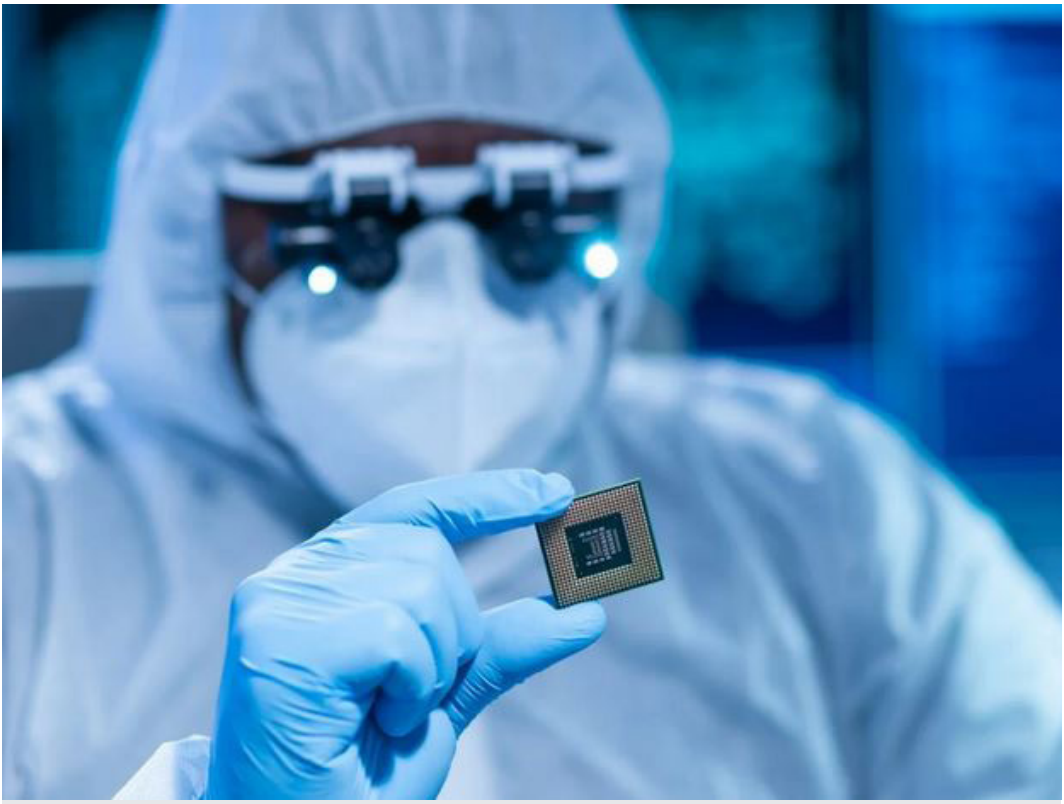
پژوهشگران این کار را گامی اولیه اما ملموس به سمت «مزیت عملی کوانتومی» توصیف می‌کنند؛ جایی که سیستم‌های کوانتومی در وظایف واقعی از نمونه‌های کلاسیک بهتر عمل می‌کنند.

این دستاورد شباهت‌هایی با تحول اخیر اقتصاد مدل‌های زبانی بزرگ دارد، جایی که سیستم‌های سبک‌تر مانند دیپ‌سیک نشان دادند

معماری‌های کوچک‌تر و کارآمدتر می‌توانند با خوشه‌های محاسباتی عظیم رقابت کنند.

آکادمی علوم چین اعلام کرد این پژوهش «یک چارچوب آزمایشی قابل اجرا برای توسعه هوش مصنوعی کوانتومی کم‌مصرف و با بُعد بالا برای کاربردهای واقعی» ارائه می‌دهد.

رقابت رایانش کوانتومی دیگر فقط درباره تعداد کیوبیت‌ها نیست، بلکه درباره این است که ماشین‌های ناقص امروز چه مسائل واقعی را می‌توانند حل کنند. با این حال، این سیستم فعلاً از نظر مقیاس محدود باقی مانده است.



جهش چین در تراشه‌های دوبعدی نسل جدید

دانشمندان چینی روشی برای رشد نیمه‌هادی‌های دوبعدی در مقیاس ویفر (wafer-scale) توسعه داده‌اند که سرعت رشد را هزار برابر افزایش می‌دهد و مسیر را برای پیشرفت‌های صنعتی هموار می‌کند.

به گزارش ساوت چاینا مورنینگ پست، افزایش سریع تقاضا برای تراشه‌های پرقدرت و کم‌مصرف که ناشی از رشد هوش مصنوعی و مدل‌های زبانی بزرگ است، جست‌وجو برای فناوری‌های نسل بعدی نیمه‌هادی را تشدید کرده است.

قانون مور (Moore's Law) پیش‌بینی می‌کرد ظرفیت نیمه‌هادی‌ها هر دو سال دو برابر شود، اما با ادامه کوچک‌تر شدن ابعاد تراشه‌ها، محدودیت‌های فیزیکی باعث شده افزایش بیشتر عملکرد به‌طور فزاینده‌ای دشوار شود.

نیمه‌هادی‌های دوبعدی (2D semiconductors) به‌عنوان یکی از اصلی‌ترین گزینه‌ها برای مواد تراشه‌های پسامور (post-Moore) مطرح شده‌اند، زیرا می‌توانند امکان کوچک‌سازی بیشتر ترانزیستورها را فراهم کنند.

در یک نیمه‌هادی دوبعدی، توانایی هدایت الکتریکی با افزودن مقادیر بسیار اندکی از عناصر دیگر قابل تغییر است؛ فرآیندی که «دوپینگ» (doping) نام دارد و می‌تواند به تولید مواد نوع n (منفی) و نوع p (مثبت) منجر شود.

در حالی که بسیاری از نیمه‌هادی‌های دوبعدی نوع n مانند مولیبدن دی‌سولفید و مولیبدن دی‌سلنید وجود دارند، نمونه‌های نوع p با عملکرد بالا و پایداری مناسب بسیار نادر هستند.

ترانزیستورهای یک تراشه برای عملکرد به مواد نوع n و p نیاز دارند که به‌صورت جفت کار کنند. کمبود مواد نوع p با عملکرد بالا به یک گلوگاه حیاتی برای توسعه نیمه‌هادی‌های دوبعدی در گره‌های زیر ۵ نانومتر تبدیل شده و همچنین به یک جبهه رقابتی شدید علمی و فناورانه بدل شده است.

برای حل این چالش، گروهی از محققین چینی، روش جدیدی برای تولید نیمه‌هادی‌های دوبعدی توسعه دادند.

این تیم اعلام کرد فرآیند متداول رسوب‌دهی شیمیایی بخار یاCVD))

را با استفاده از یک زیرلایه دولایه طلای/تنگستن مایع (Au/W) متحول کرده است.

این روش امکان رشد فیلم‌های تک‌لایه تنگستن سیلیکون نیترید را در مقیاس ویفر و با قابلیت تنظیم دوپینگ فراهم کرد.

روش جدید اندازه نواحی تک‌بلوری را به ابعاد زیر میلی‌متری افزایش داده و نرخ رشد را نسبت به مقادیر گزارش‌شده قبلی حدود هزار برابر بیشتر کرده است.

در روش‌های قبلی، رشد می‌توانست تنها حدود ۱ میکرومتر طی ۵ ساعت باشد، اما اکنون این نرخ به ۲۰ میکرومتر در دقیقه رسیده و اندازه نهایی فیلم به ۳.۶ در ۱.۸ سانتی‌متر رسیده است.

از نظر عملکرد ترانزیستوری، تک‌لایه تنگستن سیلیکون نیترید نه‌تنها دارای تحرک بالای حفره و چگالی جریان بالا در حالت روشن است، بلکه استحکام مکانیکی بالا، رسانایی حرارتی عالی و پایداری شیمیایی قابل‌توجهی نیز دارد.

انتظار می‌رود این افزایش ظرفیت تولید، کاربرد صنعتی نیمه‌هادی‌های دوبعدی را تسریع کند.

دستیابی به رشد سطح وسیع تک‌لایه تنگستن سیلیکون نیترید با دوپینگ کنترل‌شده، پیش‌شرط ادغام مقیاس‌پذیر کاربردهای دستگاهی است و این پژوهش مسیر استفاده از آن در مدارهای مجتمع CMOS را هموار می‌کند.

مهم‌تر از آن، این روش CVD با استفاده از طلای مایع به‌عنوان زیرلایه رشد، یک مسیر عمومی و بسیار کارآمد برای رشد سریع سایر مواد دوبعدی نیز فراهم می‌کند.

فراتر از مدارهای مجتمع، این ماده در حوزه اپتوالکترونیک نیز ظرفیت بالایی دارد و می‌تواند در دیودهای نورگسیل (LED)، فوتودیودها/ آشکارسازهای نوری و لیزرها به‌کار گرفته شود.

پایداری فوق‌العاده این ماده همچنین آن را برای حسگرهای مورد استفاده در محیط‌های مایع و دستگاه‌های الکترونیکی رابط زیستی (bio-interfacing devices) مناسب می‌کند.

رقابت هوش مصنوعی چین و آمریکا؛ توازن میان امنیت و گشودگی

کمیته منتخب مجلس نمایندگان آمریکا در امور چین اخیراً گزارشی درباره هوش مصنوعی منتشر کرد. این گزارش با عنوان «هرچه می‌تواند بخرد، و هرچه لازم باشد می‌دزدد: کارزار چین برای دستیابی به قابلیت‌های پیشرفته هوش مصنوعی»، بازتاب‌دهنده نگاه سخت‌گیرانه‌تر در واشنگتن است؛ نگاهی که رشد هوش مصنوعی چین را به‌طور فزاینده با موضوعات دسترسی به بازار و نگرانی‌های امنیتی گره می‌زند.

به گزارش ساوت چاینا مورنینگ پست، صرف‌نظر از اینکه این باورها تا چه حد کاملاً مستند باشند، آن‌ها به‌طور فزاینده در حال شکل‌دهی به چارچوب سیاستی آمریکا برای فهم رقابت فناوری میان دو کشور هستند؛ رقابتی که دیگر کمتر مسئله‌ای مربوط به نوآوری و بیشتر موضوعی مرتبط با امنیت ملی تلقی می‌شود.

در چنین بستری، جنجال اخیر پیرامون تقطیر مدل(model distillation) با محوریت شرکت‌های پیشروی آمریکایی از جمله OpenAI، Anthropic و Alphabet توجه زیادی را جلب کرده است. هماهنگی میان این شرکت‌ها، آن هم اندکی پس از تلاش واشنگتن برای ایجاد یک نظام «صادرات هوش مصنوعی فول‌استک»، نشان می‌دهد آنچه در ظاهر یک اختلاف فنی به نظر می‌رسد، در واقع بخشی از تغییری گسترده‌تر در نحوه حکمرانی و رقابت جهانی بر سر هوش مصنوعی است.

در نگاه نخست، بحث تقطیر مدل به مسیرهای فنی و مرزهای مالکیت فکری مربوط می‌شود. تقطیر، روشی رایج در یادگیری ماشین است که به مدل‌های کوچک‌تر اجازه می‌دهد عملکرد مدل‌های بزرگ‌تر را تقریب بزنند، هزینه‌های محاسباتی را کاهش دهند و پذیرش فناوری را تسریع کنند. وضعیت حقوقی این روش همچنان مبهم است و حتی شرکت‌های آمریکایی نیز خود از روش‌های مشابه استفاده کرده‌اند.

اما در محیط ژئوپلیتیکی کنونی، این موضوع بازتعریف شده است. برخی سیاست‌گذاران و شرکت‌های آمریکایی استدلال می‌کنند مدل‌های تقطیرشده می‌توانند در عملیات سایبری، کارزارهای اطلاعات نادرست یا حتی کاربردهای نظامی سوءاستفاده شوند. بنابراین، آنچه زمانی مسئله‌ای درباره بهینه‌سازی بود، اکنون به موضوعی مرتبط با امنیت ملی ارتقا یافته است.

این تغییر بازتاب‌دهنده تحول عمیق‌تری در حکمرانی هوش مصنوعی در آمریکاست. طی چند سال گذشته، واشنگتن از تمرکز اولیه بر ایمنی هوش مصنوعی شامل ریسک‌های اخلاقی و آسیب‌های الگوریتمی، به سمت پارادایمی امنیت‌محور حرکت کرده که بر رقابت راهبردی و کنترل

فناوری متمرکز است. ادبیات ایمنی از بین نرفته، اما اکنون بیش از پیش با ملاحظات امنیت ملی درهم آمیخته است.

اگرچه آمریکا همچنان در تولید مدل‌های پیشرفته زبانی بزرگ پیشتاز است، چین در مقیاس و انتشار فناوری بسیار رقابتی باقی مانده و بزرگ‌ترین سهم جهانی از مقالات و پتنت‌های هوش مصنوعی را در اختیار دارد و به‌سرعت استقرار واقعی این فناوری را گسترش می‌دهد.

هم‌زمان، شکاف عملکردی میان مدل‌های پیشروی آمریکایی و چینی در جدیدترین چرخه‌های ارزیابی به‌ویژه در وظایف کاربردی و چندزبانه، بیش از پیش کاهش یافته است.

این کاهش فاصله شاید توضیحی برای احساس فوریت و حتی نگرانی فزاینده در برخی محافل واشنگتن باشد. تصویرسازی از توسعه هوش مصنوعی به‌عنوان مسابقه‌ای که باید در آن پیروز شد، عمیقاً در گفتمان سیاستی آمریکا نهادینه شده است.

با تشدید رقابت، حفظ رهبری فناوری دیگر کافی تلقی نمی‌شود؛ کند کردن رقبا نیز به هدفی هم‌سنگ تبدیل شده است.

از منظر نهادی، این تغییر رابطه دولت و صنعت را نیز بازتعریف کرده است. از طریق نهادهای مشورتی، کنترل‌های صادراتی و ابتکارهای تعیین استاندارد، شرکت‌های بزرگ فناوری آمریکا به‌تدریج در چارچوب حکمرانی‌ای ادغام شده‌اند که با اولویت‌های امنیت ملی همسو است.

نتیجه، نوعی هماهنگی درهم‌تنیده است: شرکت‌ها همچنان بازیگران بازار باقی می‌مانند، اما تا حدی به ابزارهای سیاست راهبردی نیز تبدیل می‌شوند.

در این نظام در حال تحول، شرکت‌هایی مانند OpenAI،Anthropic

و Google دیگر صرفاً رهبران نوآوری نیستند، بلکه به دروازه‌بانان قابلیت‌های پیشرفته هوش مصنوعی تبدیل می‌شوند.

امنیت دیگر فقط مدیریت ریسک نیست؛ بلکه تعیین این است که چه کسی به فناوری‌های پیشرفته دسترسی داشته باشد و تحت چه شرایطی. به نام امنیت ملی، این شرکت‌های بزرگ فناوری قادرند پویایی‌های رقابتی را شکل دهند و موانع ورود برای رقبای بالقوه را افزایش دهند.

این روند با روح حاکم بر دموکراتیزه کردن فناوری که پایه بخش بزرگی از گسترش اقتصاد دیجیتال بوده، ناسازگار به نظر می‌رسد. هرچه دسترسی محدودتر شود، انتشار قابلیت‌های پیشرفته نیز ممکن است کندتر شود؛ امری که هم مشارکت گسترده‌تر در نوآوری را محدود می‌کند و هم عدم تقارن میان دارندگان فناوری‌های هسته‌ای و وابستگان به آن‌ها را تقویت می‌کند.

یکی از پیامدهای مستقیم این تغییر، محدود شدن گزینه‌های راهبردی کشورهای در حال توسعه است. برای بسیاری از کشورهای جنوب جهانی، دسترسی به قابلیت‌های پیشرفته هوش مصنوعی مستلزم ادغام در اکوسیستم‌های فناوری موجود است که اغلب تحت سلطه تعداد محدودی شرکت قرار دارند.

مشارکت در این اکوسیستم‌ها همراه با قواعد و استانداردهای از پیش تعبیه‌شده است، در حالی که ایجاد قابلیت‌های مستقل نیازمند منابع عظیم و زمان طولانی است. نتیجه، محدودیتی ساختاری است که خطر تعمیق شکاف‌های فناوری جهانی را افزایش می‌دهد.

در حوزه حکمرانی جهانی هوش مصنوعی نیز، جایی که گفت‌وگوهای بین‌المللی پیش‌تر عمدتاً بر اخلاق، شفافیت و ایمنی متمرکز بود، اکنون

رقابت ژئوپلیتیکی بیش از پیش تعیین‌کننده شده است.
حکمرانی دیگر صرفاً کاهش ریسک نیست، بلکه مدیریت مزیت راهبردی نیز هست.
امنیتی‌سازی افراطی خطر چندپاره کردن چشم‌انداز جهانی فناوری به بلوک‌های رقیب را به همراه دارد؛ روندی که هزینه‌ها را برای همه افزایش داده و نوآوری را کند می‌کند.
این تحول همکاری بین‌المللی را پیچیده‌تر می‌کند، حتی در حالی که گفت‌وگو در حوزه‌های حساس مانند کاربردهای نظامی هوش مصنوعی و امنیت زیرساخت را ضروری‌تر می‌سازد.
برای چین، این تحولات هم چالش و هم فرصت ایجاد می‌کنند. از یک سو، تشدید کنترل‌های خارجی مسیرهای سنتی دستیابی به فناوری‌های پیشرفته را محدود می‌کند. از سوی دیگر، این روند تلاش‌ها برای تقویت اکوسیستم‌های نوآوری داخلی را تسریع می‌کند.
در بلندمدت، ایجاد یک اکوسیستم جامع هوش مصنوعی شامل داده، توان محاسباتی، مدل‌ها و کاربردها، برای تقویت تاب‌آوری فناورانه ضروری خواهد بود.
در عین حال، چین در حفظ یک محیط بین‌المللی باز و فراگیر برای توسعه هوش مصنوعی نیز ذی‌نفع است.
رویکرد سیاستی چین با تأکید هم‌زمان بر توسعه و مدیریت ریسک شاید در این زمینه چشم‌انداز مفیدی ارائه کند. چالش صرفاً پاسخ دادن به محدودیت‌های خارجی نیست، بلکه مشارکت در شکل‌دهی به چارچوب حکمرانی‌ای است که میان امنیت و گشودگی توازن برقرار کند.
در نهایت، جنجال پیرامون تقطیر مدل رویدادی منفرد نیست؛ بلکه

بازتاب تغییری گسترده‌تر در منطق رقابت فناورانه است؛ تغییری که در آن ملاحظات امنیتی بیش از پیش در راهبردهای بازار و نوآوری ادغام می‌شوند.

بنابراین، پرسش اصلی برای جامعه بین‌المللی فقط نحوه حکمرانی ریسک‌های هوش مصنوعی نیست، بلکه این است که چگونه می‌توان مانع از تبدیل روایت‌های امنیتی به ابزارهای حذف و طرد شد.

اگر استانداردها، مقررات و کنترل قابلیت‌ها عمدتاً برای افزایش موانع به کار روند، توسعه جهانی هوش مصنوعی ممکن است به سمت چندپارگی ساختاری حرکت کند: نظام‌های پیشرو با کنترل قابلیت‌های حیاتی مزیت خود را تثبیت می‌کنند، در حالی که دیگران با محدودیت‌های فناورانه و نهادی فزاینده روبه‌رو می‌شوند.

وظیفه پیش‌رو انتخاب میان امنیت و گشودگی نیست، بلکه یافتن توازنی پایدار میان این دو است.

چگونگی برقراری این توازن تعیین خواهد کرد که آیا هوش مصنوعی به موتور مشترک پیشرفت جهانی تبدیل می‌شود یا به شکافی دیگر در جهانی که از نظر فناوری هرچه بیشتر دچار تقسیم‌بندی است.

رونمایی دیپ‌سیک از V۴ با حمایت تراشه‌های هواوی

دیپ‌سیک سرانجام مدل پایه نسل جدید هوش مصنوعی مورد انتظار خود،V۴ متن‌باز، را عرضه کرد؛ مدلی که به گفته این شرکت، با مدل‌های بسته پیشروی آمریکایی از شرکت‌هایی مانند OpenAI و Google DeepMind رقابت می‌کند.

به گزارش ساوت چاینا مورنینگ پست، این استارتاپ هوش مصنوعی مستقر در هانگژو دو نسخه از این مدل را منتشر کرد. مدلV۴-pro با ۱.۶ تریلیون پارامتر، بزرگ‌ترین مدل تاریخ این شرکت از نظر تعداد پارامتر محسوب می‌شود، در حالی که مدل کوچک‌ترV۴-flash دارای ۲۸۴ میلیارد پارامتر است. به‌طور کلی، تعداد پارامتر بالاتر معمولاً با

توانمندی بیشتر مدل همبستگی دارد، هرچند همزمان نیاز محاسباتی آموزش و استقرار آن را نیز افزایش می‌دهد.

هر دو مدل دارای پنجره متنی (context window) یک میلیون توکنی هستند؛ قابلیتی کلیدی که میزان اطلاعات قابل پردازش توسط سیستم هوش مصنوعی را تعیین می‌کند. دیپ‌سیک اعلام کرد این دستاورد با بهره‌وری هزینه‌ای «در سطح جهانی پیشرو» محقق شده است. مدل پرچمدار قبلی این شرکت تنها پنجره متنی ۱۲۸ هزار توکنی داشت.

اندکی پس از انتشار مدل جدید، هواوی اعلام کرد مجموعه تراشه‌های Ascend این شرکت به همراه سامانه‌های سوپرنود (supernode) آن، «پشتیبانی کامل» برای اجرای مدل‌های V۴ در مرحله استنتاج (inference) ارائه خواهند کرد. شرکت تولیدکننده تراشه هوش مصنوعی کمبریکن نیز به سرعت از سازگاری محصولات خود با مدل‌های جدید دیپ‌سیک خبر داد.

در عرضه V۴ به‌صراحت به سازگاری با تراشه‌های داخلی اشاره شده است. می‌توان انتظار داشت توانمندی کارت‌های گرافیک داخلی امسال به‌طور قابل‌توجهی ارتقا یافته و استفاده از آن‌ها فراگیر شود..

با این حال، اندازه بسیار بزرگ pro-V۴ باعث می‌شود اجرای محلی آن روی سخت‌افزارهای مصرفی عملاً غیرممکن باشد. با این وجود، گزارش فنی مفصل منتشرشده درباره معماری مدل و روش‌های آموزشی V۴ احتمالاً برای توسعه‌دهندگان جهانی هوش مصنوعی ارزشمند خواهد بود.

مدل flash-V۴ نیز یکی از ارزان‌ترین مدل‌های پیشرفته موجود در بازار است و قیمت‌گذاری توکن آن مشابه مدل V۲ دیپ‌سیک است که در ژوئن ۲۰۲۴ عرضه شده بود.

دیپ‌سیک اعلام کرد ظرفیت پردازشی V۴-pro در حال حاضر به دلیل کمبود عرضه توان محاسباتی محدود شده است و با عرضه گسترده سوپرنودهای ۹۵۰PR Ascend هواوی، قیمت‌ها در نیمه دوم سال به‌طور قابل توجهی کاهش خواهد یافت.

دیپ‌سیک به‌عنوان شناخته‌شده‌ترین شرکت هوش مصنوعی چین، به‌نوعی شاخص سنجش پیشرفت صنعت نیمه‌هادی این کشور محسوب می‌شود؛ صنعتی مرتبط که توسعه آن تحت تأثیر محدودیت‌های صادراتی آمریکا بر تجهیزات پیشرفته تولید تراشه قرار گرفته است.

پیش از عرضه V۴، مقام‌های آمریکایی دیپ‌سیک را متهم کرده بودند که برای آموزش مدل‌های خود از تراشه‌های ممنوعه Blackwell انویدیا استفاده کرده است.

دیپ‌سیک جزئیاتی درباره سخت‌افزار مورد استفاده برای آموزش V۴ منتشر نکرد. با این حال، در گزارش فنی خود به توسعه «کرنل»ها - کدهایی که عملکرد واحدهای پردازش گرافیکی (GPU) را تعیین می‌کنند - اشاره کرد که برای تراشه‌های انویدیا و هواوی بهینه‌سازی شده‌اند.

انتظار می‌رود زنجیره تأمین گسترده‌تر تراشه نیز از این تحولات منتفع شود. مدل جدید احتمالاً هم تقاضا برای توان محاسباتی داخلی را افزایش می‌دهد و هم تطبیق تراشه‌های بومی با مدل‌های پیشرو - از جمله CPUها - را تسریع می‌کند.

با وجود گمانه‌زنی‌ها مبنی بر اینکه V۴ مدلی چندوجهی (multimodal) خواهد بود، یعنی علاوه بر متن، توان پردازش تصویر و ویدئو نیز خواهد داشت، مدل پرچمدار دیپ‌سیک همچنان فقط متنی است. این شرکت اعلام کرد «در حال کار روی افزودن قابلیت‌های چندوجهی» است.

آخرین عرضه بزرگ دیپ‌سیک، مدل R۱، با کاهش فاصله توسعه مدل‌های چینی و آمریکایی به چند ماه، شوک بزرگی در سطح جهانی ایجاد کرده بود. همچنین هزینه آموزشی اعلام‌شده این مدل که ۶ میلیون دلار بود، باعث شد انویدیا در یک روز نزدیک به ۶۰۰ میلیارد دلار از ارزش بازار خود را از دست بدهد.

با این حال، این بار دیپ‌سیک برخلاف مدل‌های قبلی V۳ و V۲ که اعلام کرده بود با GPUهای Nvidia H۸۰۰ آموزش دیده‌اند، درباره هزینه آموزش و سخت‌افزار مورد استفاده سکوت کرده است.

دستورالعمل‌های پکن برای کنترل و توسعه عامل‌های هوش مصنوعی

پکن اخیرا در راستای تلاش‌های خود برای پیشبرد طرح «ای‌آی پلاس» (AI Plus)، مجموعه‌ای از دستورالعمل‌ها را به منظور ترویج کاربرد مدیریت شده عامل‌های هوش مصنوعی و توسعه نوآورانه آنها صادر کرد.

به گزارش چاینا دیلی، دستورالعمل‌های مذکور به طور مشترک توسط اداره فضای مجازی، کمیسیون توسعه و اصلاحات ملی و وزارت صنعت و فناوری اطلاعات این کشور منتشر شده‌اند و هدف آنها، اجرای خط‌مشی تعیین شده از سوی شورای دولتی در مورد پیشبرد هر چه بیشتر طرح ای‌آی پلاس و پشتیبانی از توسعه سالم و ایمن عامل‌های هوش مصنوعی است.

اصطلاح عامل هوش مصنوعی به سامانه‌های هوشمندی اطلاق می‌شود که از قابلیت ادراک، به حافظه سپاری، تصمیم‌گیری، تعامل و اجرای خودکار برخوردارند.

در این اسناد آمده که عامل‌های هوش مصنوعی روز به روز بیشتر با فضای مجازی و جهان فیزیکی ادغام می‌شوند و تغییراتی عمیق در تولید، زندگی روزمره و حکمرانی اجتماعی ایجاد می‌کنند؛ و در توسعه آنها باید اصول ایمنی و کنترل‌پذیری، توسعه استاندارد و نظام‌مند، رشد مبتنی بر نوآوری، و پیشرفت کاربردمحور را مد نظر قرار داد.

اسناد مذکور اقداماتی را شرح می‌دهد که باید در چهار حوزه کلیدی زیر انجام شوند:

تلاش برای تقویت زیرساخت‌های توسعه عامل‌های هوش مصنوعی از طریق بهبود ظرفیت‌های فناورانه و تدوین استانداردها و پروتکل‌ها

حفظ امنیت از طریق تعریف روشن و شفاف الزامات محصول، جلوگیری از ریسک‌های مربوط به ایمنی، ارتقای سامانه‌های حکمرانی و افزایش خودکنترلی یا خودانضباطی صنعت

تقویت فرایند توسعه کاربردمحور (۱۹ سناریو هم برای کاربردهای مختلف در بخش‌هایی مانند تحقیقات علمی، توسعه صنعتی، رفاه عمومی و حکمرانی اجتماعی ارائه شده‌اند)

تلاش برای خلق یک زیست‌بوم نوآوری، ارتقای همکاری صنعتی و گسترش کاربردهای عامل‌های هوش مصنوعی

سه نهاد منتشر کننده این اسناد اعلام کرده‌اند که با طرف‌های ذی‌ربط برای برنامه‌ریزی بهتر، بهبود سیاست‌های حمایتی، افزایش هماهنگی بین مقامات و اطمینان از اجرای وظایف کلیدی همکاری خواهند کرد.

رقابت هوش مصنوعی؛ تراشه با آمریکا، امتیاز با چین

تحولی آرام اما اثرگذار در حال بازتعریف رقابت جهانی هوش مصنوعی است؛ تغییری که ارتباط چندانی با این ندارد که کدام کشور قدرتمندترین مدل را می‌سازد.

به گزارش ساوت چاینا مورنینگ پست، ینسن هوانگ قصد نداشت یک راهبرد ژئوپلیتیکی را توصیف کند. اما زمانی که مدیرعامل انویدیا گفت: «بار کاری شما استنتاج (inference) است، توکن‌های شما کالای شما هستند و توان محاسباتی درآمد شماست»، در واقع از سمت عرضه، همان چیزی را بیان می‌کرد که چین از زاویه‌ای دیگر به آن رسیده بود.

برای درک چرایی این موضوع، باید از یک مفهوم پایه شروع کرد. توکن‌ها واحدهای بنیادی‌ای هستند که مدل‌های هوش مصنوعی برای پردازش و تولید زبان از آن‌ها استفاده می‌کنند: هر کلمه، هر پاسخ و هر وظیفه خودکار به این واحدها تجزیه می‌شود. ارائه‌دهندگان خدمات ابری همان‌گونه که شرکت‌های خدماتی بر اساس کیلووات‌ساعت هزینه دریافت می‌کنند، بر اساس توکن قیمت‌گذاری می‌کنند. هر کس بتواند این توکن‌ها را با کمترین هزینه و در بزرگ‌ترین مقیاس تولید کند، در اقتصاد هوش مصنوعی مزیتی به دست می‌آورد؛ همان‌طور که روزگاری فولاد ارزان برتری صنعتی را تعیین می‌کرد.

در پلتفرم **OpenRouter**، که یکی از پلتفرم‌های پرکاربرد دسترسی توسعه‌دهندگان به هوش مصنوعی است، مدل‌های چینی برای نخستین بار از مدل‌های آمریکایی در مصرف توکن پیشی گرفته‌اند: ۵.۱۶ تریلیون توکن در برابر ۲.۷ تریلیون توکن تنها در یک هفته. تا اواسط مارس، مدل‌های چینی ۳۶ درصد از حجم جهانی **OpenRouter** را در اختیار داشتند و مصرف هفتگی آن‌ها به ۷.۳۶ تریلیون توکن رسیده بود.

واشنگتن سه سال گذشته را صرف ایجاد جامع‌ترین نظام کنترل صادرات فناوری از زمان جنگ سرد کرده است؛ نظامی که برای جهانی طراحی شده که مزیت رقابتی در سخت‌افزار جابه‌جا می‌شود: تراشه‌هایی که قابل شمارش‌اند، محموله‌هایی که می‌توان جلوی آن‌ها را گرفت و زنجیره‌های تأمینی که می‌توان بر آن‌ها فشار وارد کرد. اما اکنون رقابتی متفاوت در کنار آن در حال شکل‌گیری است؛ رقابتی که ابزارهای موجود برای آن هیچ معماری مشخصی ندارند.

دوره‌های صنعتی تنها با فناوری‌های مسلط تعریف نشده‌اند، بلکه با

واحدهایی که برای سنجش آن‌ها به کار می‌رفتند نیز تعریف شده‌اند. تُن فولاد نشان می‌داد چه کسی در حال صنعتی‌شدن است. بشکه‌های نفت نشان می‌داد چه کسی اهرم نفوذ دارد. تولید ناخالص داخلی (GDP)، که در دهه ۱۹۳۰ ساخته شد، معنای فعالیت اقتصادی را تعریف کرد و دولت‌ها نیز معماری‌های سیاستی خود را حول حداکثرسازی آن بنا کردند.

رئیس اداره ملی داده چین در مجمع توسعه در پکن، توکن‌ها را «لنگر ارزش» عصر هوشمند و «واحدهای تسویه» توصیف کرد. او گزارش داد مصرف روزانه توکن در چین به ۱۴۰ تریلیون رسیده است؛ رقمی که تنها دو سال پیش ۱۰۰ میلیارد بود.

هیچ دولت دیگری تاکنون یک مفهوم مهندسی هوش مصنوعی را به شاخص اقتصادی ملی ارتقا نداده است. چین عملاً یک واحد حسابداری جدید برای عصر هوشمند پیشنهاد کرده است.

این ایده اکنون در زیرساخت‌های فیزیکی در حال شکل‌گیری است. علی بابا در شنژن یک خوشه محاسباتی ۱۰ هزار کارتی مبتنی بر تراشه‌های بومی Zhenwu راه‌اندازی کرده است. خوشه ۹۱۰C Ascend شرکت هوآوی، که آن هم در شنژن قرار دارد، ظرفیتی معادل ۱۱ هزار پتافلاپس ارائه می‌دهد. این‌ها صرفاً نمایش توان تراشه نیستند.

این پروژه‌ها در واقع بیانیه‌هایی درباره مقیاس تولید هستند؛ ظرفیت‌هایی که پیش از شکل‌گیری تقاضایی که این نظام قصد ایجاد آن را دارد ساخته شده‌اند. از این منظر، منطق آن‌ها بیشتر به توسعه شبکه راه‌آهن آمریکا در دهه ۱۸۷۰ شباهت دارد تا کشوری که با شتاب در تلاش برای عقب نماندن است.

مزیت هزینه‌ای تولید توکن در چین ساختاری است. چین «هم‌افزایی محاسبات و برق» را به یک اولویت ملی تبدیل کرده و به‌صراحت انرژی تجدیدپذیر ارزان را به رقابت‌پذیری هوش مصنوعی پیوند داده است. مدل‌های چینی برای هر یک میلیون توکن خروجی حدود ۲ تا ۳ دلار دریافت می‌کنند، در حالی که معادل‌های پیشروی غربی حدود ۱۵ دلار قیمت دارند.

این شکاف با ظهور اپلیکیشن‌های عامل‌محور هوش مصنوعی (AI agents) ، نرم‌افزارهایی که وظایف پیچیده را به‌طور خودکار انجام می‌دهند، تشدید شده است، زیرا این برنامه‌ها ۱۰ تا ۱۰۰ برابر گفت‌وگوهای عادی توکن مصرف می‌کنند. زمانی که هزینه هر نشست افزایش یافت، توسعه‌دهندگان به سمت گزینه‌های ارزان‌تر مهاجرت کردند. این مهاجرت ایدئولوژیک نبود؛ اقتصادی بود.

پروژه نیمه‌هادی یکپارچه‌سازی در مقیاس بسیار بزرگ(VLSI) ژاپن آموزنده‌ترین نمونه تاریخی و در عین حال روشن‌ترین محدودیت را ارائه می‌دهد. در اواخر دهه ۱۹۷۰، تلاش مشترک دولت و صنعت ژاپن به تولید رقابتی از نظر هزینه و صادرات تهاجمی انجامید. تراشه‌های حافظه DRAM ژاپن بازار جهانی را تصاحب کردند و سهم آمریکا بین سال‌های ۱۹۷۸ تا ۱۹۸۶ از ۷۰ درصد به ۲۰ درصد سقوط کرد.

پاسخ آمریکا، توافق نیمه‌هادی آمریکا-ژاپن در سال ۱۹۸۶ بود که کف قیمتی و نظارت بر سهم بازار را تحمیل کرد. این سازوکارها جواب دادند، زیرا صادرات قابل مشاهده بود: تراشه‌ها در مرزها شمارش می‌شدند، در گمرک قیمت‌گذاری می‌شدند و مشمول اجرای معاهدات بودند.

اما صادرات توکن چنین آسیب‌پذیری مشابهی ندارد. هیچ سازوکار

حداقل قیمت‌گذاری برای استنتاجی که از طریق فیبر نوری زیردریایی تحویل داده می‌شود وجود ندارد.

البته این مزیت سقف هم دارد. بارهای کاری سازمانی که داده‌های حساس را دربر می‌گیرند، هزینه‌های مقرراتی و اعتباری به همراه دارند که صرفاً با قیمت پایین قابل جبران نیست. اینکه این روند از توسعه‌دهندگان فراتر رفته و به پذیرش سازمانی در جنوب‌شرق آسیا و جنوب جهانی برسد یا نه، تعیین خواهد کرد که آیا این مزیت ساختاری است یا صرفاً پدیده‌ای مربوط به پذیرندگان اولیه.

ارقام داخلی چین نیز نیازمند دقت‌اند. تنها موتور Volcano شرکت بایت‌دنس روزانه ۱۲۰ تریلیون توکن گزارش می‌کند.

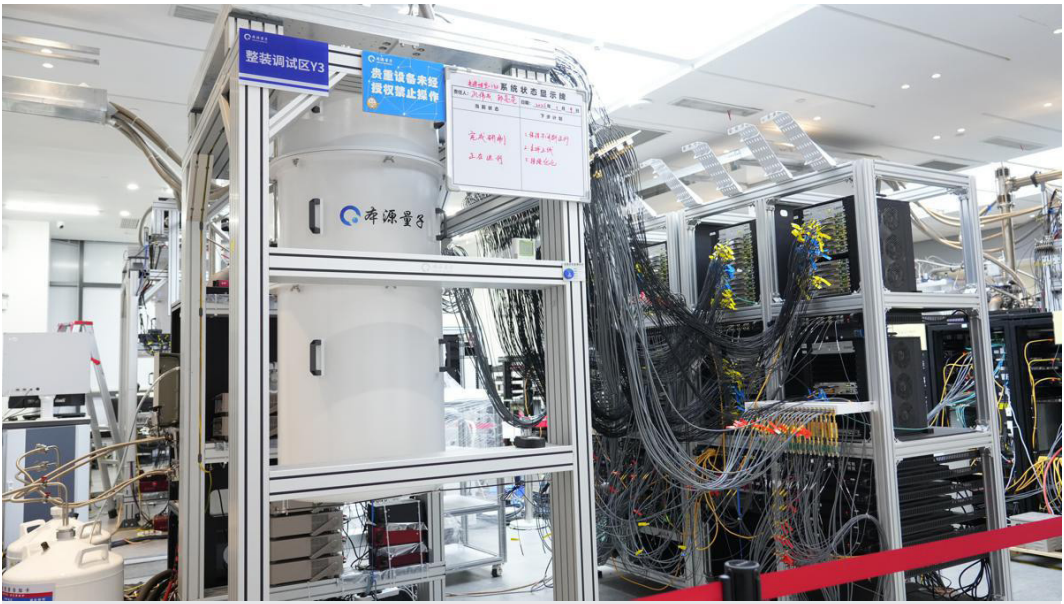
با این حال، حجم توکن لزوماً معادل ارزش اقتصادی نیست. بخش زیادی از این رشد ناشی از یارانه‌هاست. اتحاد جماهیر شوروی نیز تولید فولاد را با اشتیاق مشابهی دنبال می‌کرد؛ اما بهینه‌سازی حول این شاخص در نهایت به تحریف اقتصاد منجر شد.

واحد حساب معمولاً توسط نوآورترین اقتصاد تعیین نمی‌شود؛ بلکه اقتصادی آن را تعیین می‌کند که بیشترین کاربران، عمیق‌ترین زیرساخت و کمترین هزینه پذیرش را دارد.

چین اکنون پیش‌شرط‌های یک ادعای موازی را می‌سازد: زیرساخت داخلی در مقیاس بزرگ، استنتاج کم‌هزینه، سازوکار صادراتی نامرئی برای چارچوب‌های تجاری و نظام حسابداری ملی‌ای که توکن را در مرکز خود قرار داده است.

دیگر پرسش فقط این نیست که چه کسی قدرتمندترین هوش مصنوعی را می‌سازد. پرسش این است که چه کسی واحد سنجش خروجی هوش

مصنوعی را تعریف می‌کند و زمانی که این استاندارد تثبیت شود، چه کسی از آن واحد استفاده خواهد کرد.



عرضه چهارمین نسل رایانه کوانتومی ابررسانای بومی توسط شرکت چینی

شرکت چینیOrigin Quantum از چهارمین نسل رایانه کوانتومی ابررسانای بومی خود با نام Origin Wukong-۱۸۰ رونمایی کرد.

به گزارش چاینا دیلی، این رایانه کوانتومی به یک تراشه کوانتومی ابررسانای توسعه‌یافته داخلی مجهز است که نسبت به نسل قبلی قدرت بیشتری دارد و پذیرش وظایف محاسبات کوانتومی از سراسر جهان را آغاز کرده است.

بر اساس بیانیه شرکت، Origin Wukong-۱۸۰ که به‌طور مستقل توسط Origin Quantum مستقر در شهر هفی، مرکز استان آنهویی در شرق چین توسعه یافته، از خودکفایی و قابلیت کنترل کامل در سراسر زنجیره صنعتی برخوردار است.

طبق اعلام این شرکت، چهار سامانه کلیدی این رایانه شامل سامانه تراشه

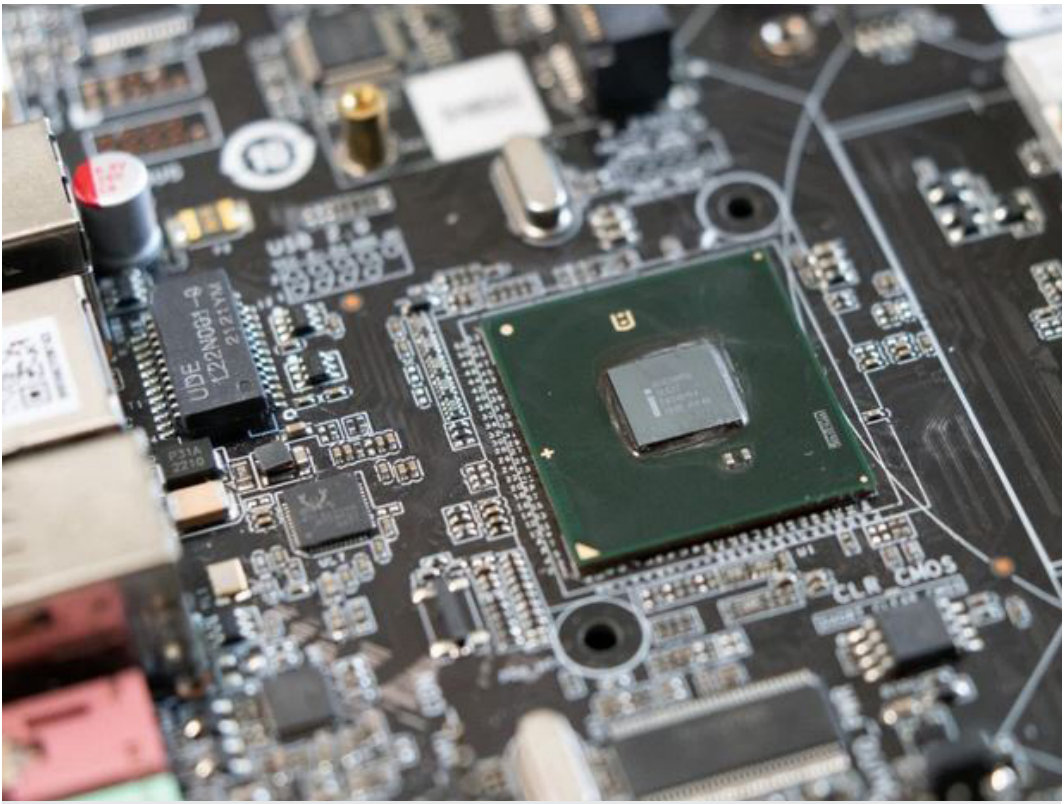
محاسبات کوانتومی، سامانه اندازه‌گیری و کنترل محاسبات کوانتومی، سامانه پشتیبانی محیطی محاسبات کوانتومی و سیستم‌عامل رایانه کوانتومی، همگی به‌صورت کامل و در سطح «فول‌استک» توسط Origin Quantum توسعه یافته‌اند.

این شرکت در سال ۲۰۲۴ نسل سوم رایانه کوانتومی ابررسانای بومی خود با نامOrigin Wukong را در سطح جهانی عرضه کرد.

این رایانه که تاکنون بیش از دو سال به‌صورت پایدار فعالیت کرده، حدود ۵۰ میلیون دسترسی از راه دور از بیش از ۱۶۰ کشور جهان ثبت کرده، بیش از ۹۰۰ هزار وظیفه محاسبات کوانتومی جهانی را تکمیل کرده و در سال ۲۰۲۵ نخستین صادرات و فروش تجاری توان محاسبات کوانتومی بومی چین را محقق ساخته است.

شرکتOrigin Quantum به‌عنوان نخستین شرکت محاسبات کوانتومی چین، نخستین خط تولید تراشه کوانتومی این کشور را راه‌اندازی کرده و نخستین رایانه کوانتومی ابررسانای چین را تحویل و مستقر کرده است.

این شرکت همچنین نخستین سیستم‌عامل رایانه کوانتومی چین را که برای دانلود عمومی در دسترس است توسعه داده و با همکاری نزدیک به ۱۰۰ دانشگاه در سراسر کشور، برنامه‌های آموزشی محاسبات کوانتومی را راه‌اندازی کرده است.



برتری تراشه‌سازان چینی در نسبت هزینه تحقیق و توسعه

صورت‌های مالی سه‌ماهه نخست نشان می‌دهد شرکت‌های پیشروی تراشه‌سازی چین، در مقایسه با همتایان آمریکایی خود، سهم بیشتری از درآمدشان را به تحقیق و توسعه اختصاص داده‌اند؛ موضوعی که همزمان با پیشبرد راهبرد خودکفایی فناوری پکن در بحبوحه رونق هوش مصنوعی رخ داده است.

به گزارش ساوت چاینا مورنینگ پست، شرکتMoore Threads در سه‌ماهه منتهی به مارس ۲۰۲۶، نیمی از درآمد خود را صرف تحقیق و

توسعه کرد؛ در حالی که شرکتMetaX در همین دوره ۴۵ درصد از درآمد خود را به این بخش اختصاص داد.

در مقابل، شرکت‌های آمریکایی تولیدکننده تراشه مانندAMD و اینتل در سال‌های اخیر بین ۲۰ تا ۳۰ درصد از درآمد خود را صرف تحقیق و توسعه کرده‌اند.

طبق اسناد بورسی، نسبت هزینه تحقیق و توسعه به درآمد در انویدیا از ۲۷.۲ درصد در سال ۲۰۲۲ به ۸.۶ درصد در سال ۲۰۲۵ کاهش یافت؛ زیرا درآمد این شرکت در سال مالی منتهی به ۲۵ ژانویه ۲۰۲۶، تحت تأثیر جهش تقاضا برای تراشه‌های پیشرفته هوش مصنوعی، به ۲۱۵.۹ میلیارد دلار رسید.

R&D as % of Revenue: US vs. China Chip Champions

	Country	2025	2024	2023	2022
MetaX	China	**62**	121	1,318	151,858
Moore Threads	China	**87**	310	1,076	2,423
Cambricon	China	**18**	91	158	209
Nvidia	US	**9**	10	14	27
AMD	US	**23**	25	26	21
Intel	US	**26**	31	30	28

Note: for mobile user, scroll horizontal for more information

Source: Exchange filings SCMP Graphics

با وجود نسبت‌های بالاتر، طراحان تراشه چینی همچنان از نظر حجم مطلق هزینه‌کرد تحقیق و توسعه فاصله زیادی با غول‌های آمریکایی دارند.

بر اساس اسناد بورسی، انویدیا در سال مالی منتهی به ۲۵ ژانویه ۲۰۲۶، ۱۸.۵ میلیارد دلار صرف تحقیق و توسعه کرد؛ در حالی که Advanced Micro Devices حدود میلیارد دلار و اینتل ۱۳.۸ میلیارد دلار در سال مالی منتهی به ۲۷ دسامبر ۲۰۲۵ هزینه کردند.

در مقایسه،MetaX در سال ۲۰۲۵ حدود ۱ میلیارد یوان (۱۴۶ میلیون دلار) وMoore Threads حدود ۱.۳ میلیارد یوان صرف تحقیق و توسعه کردند.

هزینه‌های تحقیق و توسعه تراشه‌سازان چینی همچنان رو به افزایش است؛ موضوعی که نشان‌دهنده تلاش آنها برای کاهش فاصله با رقبای آمریکایی است. هزینه تحقیق و توسعهMoore Threads در سه‌ماهه نخست ۲۰۲۶ نسبت به سال قبل ۵۰ درصد افزایش یافت و به ۳۶۹ میلیون یوان رسید، در حالی که این رقم برایMetaX در همین دوره با رشد ۱۶.۳ درصدی به ۲۵۳ میلیون یوان رسید.

همچنین انتظار می‌رود با رشد درآمد، نسبت‌های تحقیق و توسعه شرکت‌های چینی به‌تدریج به سطوح مشابه شرکت‌های آمریکایی نزدیک شود؛ روندی که هم‌اکنون در میان طراحان تراشه داخلی باسابقه‌تر قابل مشاهده است.

شرکت که چهار سال زودتر از MetaX و Moore Threads تأسیس شده، در سه‌ماهه منتهی به مارس ۲۰۲۶ معادل ۱۱.۲ درصد درآمد خود را صرف تحقیق و توسعه کرد؛ رقمی که نسبت به ۲۴.۵ درصد در دوره مشابه سال قبل کاهش یافته است.

کمبریکن در سه‌ماهه نخست ۲۰۲۶ عملکرد مالی قدرتمندی ثبت کرد؛ به‌طوری که درآمد آن ۱۶۰ درصد افزایش یافت و به ۲.۹ میلیارد یوان

رسید، در حالی که سود آن نیز با رشد ۱۸۵ درصدی به ۱ میلیارد یوان رسید. این عملکرد در ادامه نخستین سود کامل سالانه شرکت در سال ۲۰۲۵ از زمان عرضه اولیه در ۲۰۲۰ بود.

این شرکت علت رشد را «افزایش مستمر تقاضای صنعت هوش مصنوعی برای توان محاسباتی» عنوان کرد. ■

دفتر همکاری فناوری سفارت جمهوری اسلامی ایران در پکن

با همکاری:

گروه مطالعاتی چین نگار

چین نگار 中国

www.chinnegar.com

@chinnegar

www.techchina.ir

info@techchina.ir

@fanavarichin

@fanavarichin

ماهنامه‌های
گروه مطالعاتی چین نگار:

ماهنامه
چین | انرژی‌های نو و تجدیدپذیر

ماهنامه
فناوری | چین

ماهنامه
چین | هوش مصنوعی و صنعت تراشه

ماهنامه
چین | صنعت خودرو

سفارت جمهوری اسلامی ایران - پکن

Embassy of the I.R. of Iran—Beijing