



مدل جدید جمناي گوگل؛ Gemini ۲.۵ Computer Use

گوگل اوایل سال جاری میلادی اعلام کرده بود که قصد دارد قابلیت‌های «کاربری کامپیوتر» (computer use) را از طریق Gemini API در اختیار توسعه‌دهندگان قرار دهد. در تاریخ ۷ نوامبر ۲۰۲۵، این شرکت مدل **Gemini ۲.۵ Computer Use** را منتشر کرد؛ یک مدل تخصصی جدید که بر پایه قابلیت‌های درک بصری و استدلال Gemini ۲.۵ Pro ساخته شده و به «عامل‌ها» (Agents) قدرت می‌دهد تا با رابط‌های کاربری (UI) تعامل داشته باشند.

آنچه در ادامه می‌خوانید، ترجمه، تحلیل و بومی‌سازی مقاله‌ای است که مستقیماً توسط خود گوگل منتشر شده است. ما در این مقاله به بررسی عمیق مدل جدید جمناي گوگل می‌پردازیم و تمام جزئیات فنی و دموهای ارائه‌شده توسط گوگل را به زبان فارسی روان توضیح می‌دهیم.

عملکرد مدل جدید جمناي گوگل در چندین بنچمارک کنترل وب و موبایل (آزمون‌های استاندارد برای سنجش توانایی هوش مصنوعی در کنترل کردن اپلیکیشن‌ها)، از رقبای

پیشرو خود بهتر بوده است، آن هم در حالی که زمان پاسخ‌دهی کمتری دارد. توسعه‌دهندگان می‌توانند از طریق Gemini API در پلتفرم‌های [Google AI Studio](#) و [Vertex AI](#) به این قابلیت‌ها دسترسی داشته باشند.

اگرچه مدل‌های هوش مصنوعی می‌توانند از طریق API‌های ساختاریافته با نرم‌افزارها ارتباط برقرار کنند، بسیاری از وظایف دیجیتال همچنان نیازمند تعامل مستقیم با رابط‌های کاربری گرافیکی (GUI) هستند؛ به عنوان مثال، پر کردن و ارسال فرم‌ها.

برای تکمیل این وظایف، «عامل‌ها» باید درست مانند انسان‌ها در صفحات وب و اپلیکیشن‌ها پیمایش کنند؛ یعنی با کلیک کردن، تایپ کردن و اسکرول کردن.

گام بعدی و حیاتی در ساخت عامل‌های هوشمند قدرتمند و همه‌منظوره؛ هدفی که مدل Gemini ۲.۵ Computer Use برای آن طراحی شده است، اینجام اقدامات زیر است:

- توانایی بومی برای پر کردن فرم‌ها
- کار با عناصر تعاملی مانند منوهای کشویی و فیلترها
- فعالیت در پشت صفحات لاگین

مدل جدید جمناي گوگل، چگونه کار می‌کند؟

قابلیت‌های اصلی مدل جدید جمناي گوگل از طریق ابزار جدید `computer_use` در Gemini API ارائه می‌شود و طوری طراحی شده است که باید در یک «حلقه» (loop) عملیاتی شود.

ورودی‌های این ابزار شامل درخواست کاربر، اسکرین‌شات محیط و تاریخچه‌ای از اقدامات اخیر است. همچنین در بخش ورودی می‌توان مشخص کرد آیا توابعی از «لیست کامل اقدامات ال پشتیبانی‌شده» حذف شوند، یا اینکه توابع سفارشی اضافی برای گنجانده شدن، تعریف شوند.

توضیحات واضح‌تر

اگر مطلب بالا را درست متوجه نشدید، بیابید واضح‌تر توضیح دهیم.

بخش بالا کمی فنی است. به زبان ساده، گوگل توضیح می‌دهد برای استفاده از مدل جدید چمنای گوگل (Gemini ۲.۵ Computer Use)، توسعه‌دهنده باید سه چیز را به صورت همزمان به آن بدهد:

۱. درخواست کاربر: یعنی به مدل گفته شود دقیقا چه هدفی دارد. (مثلا: «در این صفحه ثبت‌نام کن» یا «قیمت این کالا را پیدا کن»).

۲. اسکرین‌شات محیط: یک عکس کامل از چیزی که همین لحظه روی صفحه نمایش (در مرورگر یا اپلیکیشن) دیده می‌شود.

۳. تاریخچه اقدامات اخیر: لیستی از کارهایی که هوش مصنوعی در چند ثانیه قبل انجام داده است. (مثلا: ۱. کلیک روی فیلد نام، ۲. تایپ کردن «علی»). این کار به مدل کمک می‌کند بفهمد الان کجای کار است و گیج نشود.

گوگل همچنین اشاره می‌کند این ابزار، دو قابلیت پیشرفته هم دارد:

۱. محدود کردن دسترسی (Exclude functions)

توسعه‌دهنده می‌تواند به مدل بگوید از برخی کارها استفاده نکند. مثلا: «اجازه نداری روی دکمه حذف کلیک کنی» یا «حق نداری چیزی را تایپ کنی، فقط کلیک کن».

۲. اضافه کردن قابلیت‌های جدید (Custom functions)

توسعه‌دهنده می‌تواند ابزارهای سفارشی خودش را به مدل اضافه کند. (مثلا: یک ابزار مخصوص به نام «دریافت کد تایید از ایمیل» تعریف کند تا مدل بتواند در صورت نیاز از آن هم استفاده کند).

فرایند تحلیل، تصمیم‌گیری و صدور دستور

سپس مدل جدید چمنای گوگل این ورودی‌ها را تجزیه و تحلیل کرده و پاسخی تولید می‌کند. این پاسخ، معمولا یک «فراخوانی تابع» (function call) است که یکی از اقدامات رابط کاربری مانند کلیک کردن یا تایپ کردن را نمایندگی می‌کند. این پاسخ

همچنین ممکن است حاوی درخواستی برای تایید از سوی کاربر نهایی باشد، که این تایید، برای اقدامات خاصی مانند نهایی کردن خرید، الزامی است. پس از دریافت پاسخ، کدی که در سمت کاربر (client-side) قرار دارد، اقدام مشخص شده را اجرا می‌کند.

چرخه ادامه می‌یابد...

پس از اجرای آن اقدام، یک اسکرین‌شات جدید از رابط کاربری گرافیکی (GUI) و همچنین URL صفحه فعلی، به عنوان «پاسخ تابع» به مدل Gemini ۲.۵ Computer Use بازگردانده می‌شود و به این ترتیب، حلقه (loop) مجدداً راه‌اندازی می‌گردد. این فرآیند تکرار شونده آنقدر ادامه می‌یابد تا:

- وظیفه به‌طور کامل انجام شود.
- یک خطا رخ دهد.
- تعامل توسط یک پاسخ ایمنی (safety response) یا تصمیم خود کاربر متوقف شود.

این مدل برای چه پلتفرم‌هایی بهینه شده است؟

مدل جدید جمناي گوگل در درجه اول برای مرورگرهای وب بهینه‌سازی شده است؛ اما پتانسیل بسیار قدرتمندی را برای وظایف کنترل رابط کاربری موبایل نیز از خود نشان می‌دهد.

با این حال، این مدل هنوز برای کنترل در سطح سیستم‌عامل دسکتاپ (مثلاً کنترل مستقیم ویندوز یا مک) بهینه‌سازی نشده است.

مثال واقعی: قدرت مدل Gemini ۲.۵ Computer Use در عمل

گوگل برای نمایش توانایی‌های این مدل، دمویی را منتشر کرده که در آن، یک دستور (Prompt) بسیار پیچیده به هوش مصنوعی داده می‌شود. این دستور نشان می‌دهد

مدل Gemini ۲.۵ Computer Use چگونه مانند یک دستیار انسانی واقعی عمل می‌کند.

مثال اول: هوش مصنوعی به عنوان یک عامل پردازش داده

دستور (Prompt) داده شده به مدل:

```
["message_box text_color="light]
```

«به این سایت ثبت‌نام حیوانات خانگی برو، اطلاعات تمام حیوانات ساکن کالیفرنیا را کپی کن. سپس به این سایت CRM (مدیریت مشتری) برو و آن‌ها را به‌عنوان مهمان جدید ثبت کن. بعد، برایشان در تاریخ ۱۰ اکتبر (بعد از ۸ صبح) با متخصص "Anima Lavar" نوبت ملاقات رزرو کن و دلیل ملاقات را همان درمانی که قبلاً درخواست کرده بودند، بنویس.»

```
[message_box/]
```

این یک دستور ساده نیست. مدل Gemini ۲.۵ Computer Use در این دمو نشان می‌دهد که می‌تواند:

- بین دو وبسایت مختلف جابه‌جا شود.
- اطلاعات را بر اساس یک شرط (ساکن کالیفرنیا) فیلتر کند.
- فرم‌ها را پر کند.
- با تقویم و سیستم رزرواسیون کار کند.
- اطلاعاتی مانند دلیل درمان را از سایت اول «به خاطر بسپارد» و در سایت دوم از آن استفاده کند.

مثال دوم: هوش مصنوعی به عنوان یک سازمان‌دهنده بصری

این مثال دوم، جنبه‌ی کاملاً متفاوتی از توانایی‌های این مدل را نشان می‌دهد:

دستور (Prompt) داده شده به مدل:

["message_box text_color="light]

«باشگاه هنری ما برای نمایشگاه پیش‌رو، یک جلسه طوفان فکری برگزار کرد. تخته یادداشت‌ها الان خیلی به‌هم‌ریخته است و من برای سازماندهی وظایف در دسته‌بندی‌هایی که قبلاً ایجاد کرده‌ام، به کمک تو نیاز دارم.

لطفاً به سایت sticky-note-jam.web.app برو و مطمئن شو همه‌ی یادداشت‌ها در بخش‌های صحیح خود قرار دارند. اگر در جای درست نیستند، آن‌ها را در جای درست بکش و رها کن.» (Drag and Drop)

[/message_box]

این مثال حتی از قبلی هم جالب‌تر است! مثال اول بر «تایپ کردن»، «کپی کردن» و «ناوبری بین صفحات» متمرکز بود؛ اما این مثال، توانایی «درک بصری و فضایی» را نشان می‌دهد. این دمو نشان می‌دهد مدل Gemini ۲.۵ Computer Use فقط یک ربات پرکننده فرم نیست؛ بلکه می‌تواند با عناصر بصری رابط کاربری، مانند یک انسان تعامل کند، آن‌ها را بفهمد و حتی اشیاء دیجیتال را روی صفحه جابجا کند.

مشاهده دموها در عمل

مقاله اصلی گوگل برای هر یک از این مثال‌ها، یک ویدیو دمو (نمایشی) نیز منتشر کرده است که فرایند کار و نتایج نهایی مدل Gemini ۲.۵ Computer Use را به‌صورت بصری، به‌خوبی توضیح می‌دهد. برای مشاهده این ویدیوهای جذاب، می‌توانید در «کانال تلگرام ترفندهای هوش مصنوعی» عضو شده و با جستجوی هشتگ «#Gemini_۲_۵_Computer_Use» آن‌ها را پیدا کرده و مشاهده کنید.

بررسی عملکرد مدل جدید چمنای گوگل

مدل Gemini ۲.۵ Computer Use عملکرد قدرتمندی را در چندین «بنچمارک کنترل وب و موبایل» (آزمون‌های استاندارد) به نمایش می‌گذارد.

گوگل اشاره می‌کند جدولی که در ادامه آمده است، شامل نتایجی از سه منبع مختلف است:

۱. امتیازاتی که توسط خود رقبا گزارش شده‌اند.
۲. ارزیابی‌هایی که توسط Browserbase (یک پلتفرم ثالث) انجام شده است.
۳. و ارزیابی‌هایی که خود گوگل مستقلاً اجرا کرده است.

جزئیات کامل این ارزیابی‌ها در سندی به نام «اطلاعات ارزیابی مدل Gemini ۲.۵ Computer Use» و همچنین در یکی از وبلاگ‌های رسمی «Browserbase» در دسترس عموم قرار دارد.

گوگل همچنین تاکید می‌کند مگر در مواردی که به صراحت خلاف آن ذکر شده باشد، تمام امتیازات نمایش داده شده در جداول، مربوط به ابزارهای «کاربری کامپیوتر» هستند که از طریق API (و نه رابط کاربری چت) در دسترس قرار گرفته‌اند.

Benchmark Performance					
Benchmark	Source/Harness	Gemini 2.5 Computer Use 10-2025	Claude Sonnet 4.5	Claude Sonnet 4	OpenAI Computer-Using Agent model
Online-Mind2Web	Official leaderboard	69.0%	—	—	61.3% (Operator)
	Measured by Browserbase	65.7%	55.0%	61.0%	44.3%
WebVoyager	Self-reported	88.9%	—	—	87.0%
	Measured by Browserbase	79.9%	71.4%	69.4%	61.0%
AndroidWorld	Measured by Google DeepMind	69.7%	56.0%	62.1%	Could not measure - No model access
OSWorld	Self-reported	OS control not yet supported	61.4%	42.2%	38.1%

این مدل، کیفیتی پیشرو را برای «کنترل مرورگر» با کمترین زمان پاسخ‌دهی ارائه می‌دهد؛ این ارزیابی بر اساس عملکردی است که در پلتفرم ارزیابی Browserbase برای Online-Mind2Web اندازه‌گیری شده است.

توضیح کوتاه

«Online-Mind2Web» نام یک آزمون (بنچمارک) بسیار سخت برای سنجش توانایی

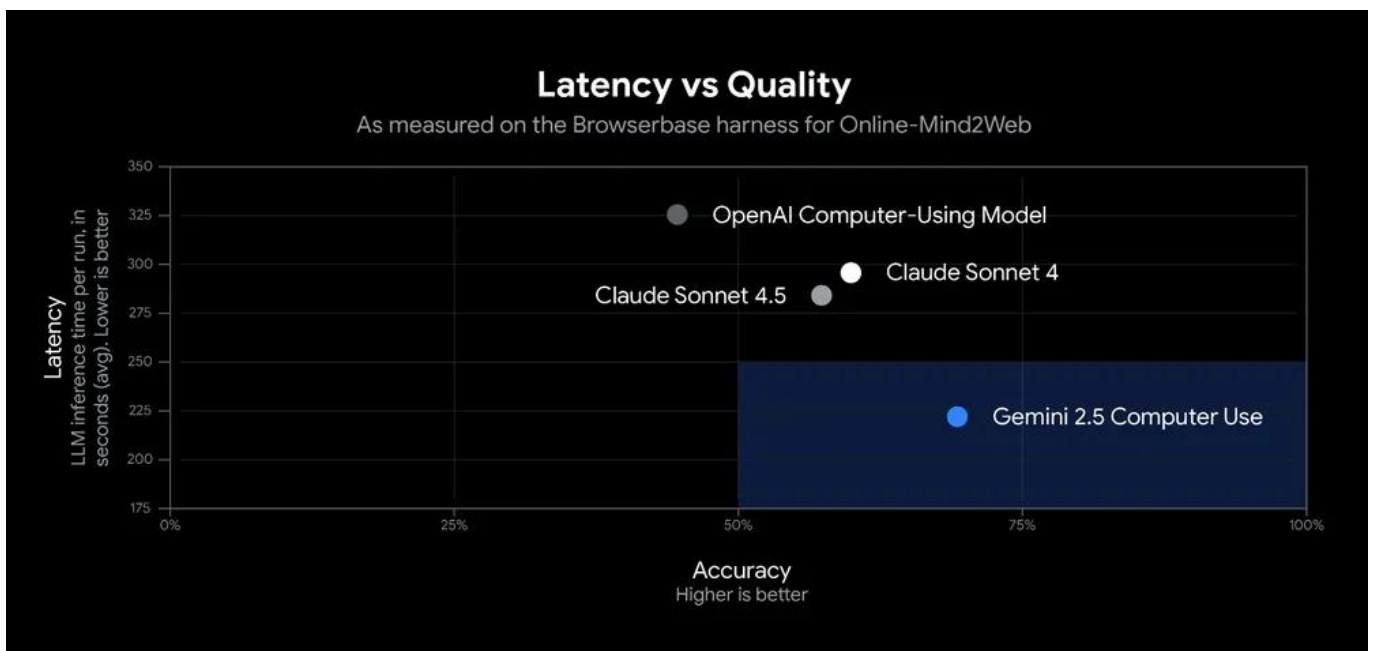
هوش مصنوعی در انجام کارهای واقعی در وبسایت‌های مختلف است. «Browserbase harness» نیز نام پلتفرمی است که این آزمون را اجرا می‌کند.

به زبان ساده، گوگل می‌گوید:

```
["message_box text_color="light]
```

در یک آزمون استاندارد و سخت که توسط یک شرکت ثالث (Browserbase) اجرا شده، مدل Gemini ۲.۵ Computer Use هم کارها را بهتر از رقبا انجام داده و هم سریع‌تر (با وقفه کمتر) پاسخ داده است.

```
[message_box/]
```



چالش ایمنی در «مدل جدید چمنای گوگل» و راهکارهای گوگل

گوگل معتقد است تنها راه برای ساخت «عامل‌های هوشمندی» (Agents) که به نفع همگان باشد، مسئولیت‌پذیری از همان ابتدا است. عامل‌های هوش مصنوعی که کامپیوترها را کنترل می‌کنند، ریسک‌های منحصر به فردی را به همراه دارند؛ از جمله:

- سوءاستفاده عمدی توسط کاربران.
- رفتار غیرمنتظره از سوی خود مدل.
- حملات «تزریق پرامپت» (prompt injections) و کلاهبرداری‌ها در محیط وب.

بنابراین، پیاده‌سازی دقیق «حفاظ‌های ایمنی» (safety guardrails) در این حوزه بسیار حیاتی است.

لایه‌های ایمنی مدل

گوگل اعلام کرده ویژگی‌های ایمنی را مستقیماً در خود مدل Gemini ۲.۵ Computer Use آموزش داده است تا این سه ریسک کلیدی را برطرف کند. علاوه بر این، گوگل کنترل‌های ایمنی را نیز در اختیار توسعه‌دهندگان قرار می‌دهد. این کنترل‌ها به توسعه‌دهندگان این قدرت را می‌دهند که از تکمیل خودکار اقدامات بالقوه پرخطر یا مضر توسط مدل جدید جمنای گوگل جلوگیری کنند.

نمونه‌هایی از این اقدامات پرخطر عبارتند از:

- آسیب رساندن به یکپارچگی سیستم (integrity).
- به خطر انداختن امنیت.
- دور زدن کپچاها (CAPTCHA).
- کنترل دستگاه‌های پزشکی.

انواع کنترل‌ها

این کنترل‌ها عبارتند از:

سرویس ایمنی گام‌به‌گام (Per-step safety service)

این یک سرویس ایمنی «خارج از مدل» است که در لحظه تصمیم‌گیری (inference-time) کار می‌کند. وظیفه آن، ارزیابی هر اقدامی است که مدل جدید جمنای گوگل پیشنهاد می‌دهد؛ پیش از آنکه آن اقدام واقعاً اجرا شود.

دستورالعمل‌های سیستمی (System instructions)

توسعه‌دهندگان می‌توانند به «عامل هوشمند» دستور دهند که یا از انجام برخی اقدامات پرخطر امتناع کند، یا حتماً قبل از انجام آن‌ها، از کاربر نهایی تاییدیه بگیرد.

توصیه‌های تکمیلی به توسعه‌دهندگان در مورد اقدامات ایمنی و بهترین شیوه‌ها، در «مستندات فنی گوگل» یافت می‌شود؛ درحالی‌که این حفاظها برای کاهش ریسک طراحی شده‌اند، گوگل اکیدا از همه توسعه‌دهندگان می‌خواهد سیستم‌های خود را قبل از راه‌اندازی، به طور کامل و دقیق آزمایش کنند.

کاربردهای «مدل جدید جمنای گوگل» توسط آزمایش‌کنندگان اولیه

تیم‌های داخلی خود گوگل، این مدل را برای موارد استفاده‌ای مانند تست رابط کاربری (UI testing) به مرحله تولید رسانده‌اند؛ (استفاده‌ای) که می‌تواند فرایند توسعه نرم‌افزار را به طور قابل‌توجهی سریع‌تر کند.

علاوه بر این، نسخه‌هایی از مدل جدید جمنای گوگل در حال حاضر به پروژه‌هایی مانند «پروژه مارینر» (Project Mariner)، «عامل تست فایربیس» (Firebase Testing Agent) و برخی از قابلیت‌های «عامل‌محور» (agentic) در «حالت هوش مصنوعی در جستجوی گوگل» (AI Mode in Search) قدرت می‌بخشند.

کاربرانی که در «برنامه دسترسی زودهنگام» گوگل (early access program) حضور داشته‌اند نیز مدل جدید جمنای گوگل را برای تقویت دستیارهای شخصی، اتوماسیون گردش کار (workflow automation) و تست ال، آزمایش کرده‌اند و به نتایج قدرتمندی دست یافته‌اند.

بازخورد آزمایش‌کنندگان اولیه

گوگل می‌گوید این کاربران به «نتایج قدرتمندی» دست یافته‌اند، اما بیایید این ادعا را از زبان خود آن‌ها بشنویم.

هیچ چیز بهتر از بازخورد مستقیم شرکت‌ها و توسعه‌دهندگانی نیست که واقعا مدل

جدید جمناي گوگل را در محصولات و گردش کارهای روزمره خود آزمایش کرده‌اند. در ادامه، تجربیات و دیدگاه‌های آن‌ها را مستقیماً می‌خوانیم تا ببینیم این فناوری چگونه به آن‌ها کمک کرده است.

Poke.com (یک دستیار هوش مصنوعی پیش‌فعال در iMessage، واتس‌آپ و SMS که دارای گردش‌های کاری متعدد شخص ثالث و عامل‌محور (agentic) است):

["message_box text_color="light]

«بسیاری از گردش‌های کاری ما نیازمند تعامل با رابط‌هایی هستند که برای انسان‌ها طراحی شده‌اند؛ (رابط‌هایی) که در آن‌ها سرعت اهمیت ویژه‌ای دارد. مدل جدید جمناي گوگل بسیار جلوتر از رقبای خود قرار دارد و اغلب ۵۰ درصد سریع‌تر و بهتر از بهترین راه‌حل‌های جایگزینی است که ما بررسی کرده‌ایم.»

[message_box/]

Autotab می‌گوید:

["message_box text_color="light]

«عامل‌های (Agents) ما کاملاً به‌صورت مستقل اجرا می‌شوند و کارهایی را انجام می‌دهند که در آن کوچک‌ترین اشتباه در جمع‌آوری و تجزیه داده‌ها غیرقابل قبول است. مدل جدید جمناي گوگل در زمینه تجزیه قابل اتکای مفهوم در موارد پیچیده، از سایر مدل‌ها عملکرد بهتری داشت و توانست عملکرد را در سخت‌ترین ارزیابی‌های ما تا ۱۸ درصد افزایش دهد.»

[message_box/]

تیم پلتفرم پرداخت گوگل، که مدل Gemini ۲.۵ Computer Use را به‌عنوان یک مکانیزم پشتیبان برای رسیدگی به تست‌های AI سرتاسری (end-to-end) شکننده پیاده‌سازی کرده بود؛ تست‌هایی که به تنهایی مسئول ۲۵ درصد از کل شکست‌های

تست بودند، می‌گوید:

```
["message_box text_color="light]
```

«زمانی که اسکریپت‌های مرسوم با شکست مواجه می‌شوند، مدل جدید جمناي گوگل وضعیت فعلی صفحه را ارزیابی کرده و به طور مستقل اقدامات مورد نیاز برای تکمیل گردش کار را تشخیص می‌دهد. این پیاده‌سازی در حال حاضر با موفقیت بیش از ۶۰ درصد از اجراها را «احیا» می‌کند؛ اجراهایی که قبلاً رفع کردن آن‌ها چندین روز طول می‌کشید.»

```
[message_box/]
```

نحوه شروع به کار با «مدل جدید جمناي گوگل»

گوگل اعلام کرده است مدل Gemini ۲.۵ Computer Use در حالت «پیش‌نمایش عمومی» در دسترس قرار گرفته است. توسعه‌دهندگان می‌توانند از طریق Gemini API در پلتفرم‌های Google AI Studio و Vertex AI به آن دسترسی داشته باشند.

گوگل برای شروع کار سه مسیر اصلی را مشخص کرده است:

۱. امتحان کردن فوری (Demo)

امکان آزمایش فوری مدل در یک «محیط دمو» (Demo Environment) فراهم شده است که توسط پلتفرم [Browserbase](#) میزبانی می‌شود.

۲. شروع به ساخت (Building)

توسعه‌دهندگان می‌توانند با مراجعه عمیق به «اسناد مرجع» و «مستندات فنی» گوگل، یاد بگیرند چگونه «حلقه عامل هوشمند» (agent loop) خود را بسازند. این کار هم به صورت محلی با ابزار Playwright و هم در یک ماشین مجازی ابری (Cloud VM) با استفاده از Browserbase امکان‌پذیر است. (گوگل همچنین مستندات Vertex AI را برای کاربردهای سازمانی ارائه کرده است).

۳. پیوستن به جامعه کاربری (Community)

گوگل از توسعه‌دهندگان دعوت کرده بازخوردها و نتایج کارهای خود را به اشتراک بگذارند و از این طریق به هدایت نقشه‌راه آینده این محصول در «انجمن توسعه‌دهندگان» گوگل کمک کنند.

آینده «عامل‌های هوشمند» گوگل

همانطور که در این مقاله خواندیم، مدل جدید جمنای گوگل فقط یک ابزار جدید نیست؛ این مدل، دست‌ها و چشم‌های هوش مصنوعی برای تعامل مستقیم با دنیای دیجیتال ما است. این مدل، شکاف بین «تصمیم‌گیری» و «اجرا» را پر می‌کند و به هوش مصنوعی اجازه می‌دهد کارهای پیچیده‌ای را که قبلاً فقط از عهده انسان برمی‌آمد، به صورت خودکار انجام دهد.

اما این «عامل اجرایی» قدرتمند، بخشی از یک پازل بسیار بزرگتر در استراتژی آینده گوگل است.

گوگل چندی پیش نیز در سکوت خبری، از برنامه‌های خود برای مغز متفکر و مدل زبان اصلی آینده‌اش، «جمنای ۳.۰ پرو»، خبر داد. آن مدل، به‌عنوان یک جهش استراتژیک به سوی «هوش مصنوعی محیطی» (Ambient AI) توصیف شده و قرار است به یک موتور استدلال مشترک برای کل اکوسیستم گوگل (از اندروید تا Workspace) تبدیل شود.

اگر مدل Gemini ۲.۵ Computer Use را «بازوهای اجرایی» بدانیم که امروز در دسترس قرار گرفته، Gemini ۳.۰ Pro «مغز متفکر» و استراتژیکی است که در آینده، این بازوها و کل اکوسیستم گوگل را هدایت خواهد کرد.

برای درک کامل این تصویر بزرگ و آشنایی با جزئیات این خبر مهم، پیشنهاد می‌کنیم حتماً مقاله دیگر ما با عنوان «غول جدید گوگل؛ جمنای ۳.۰ پرو» را در وبسایت مدیرسبز مطالعه کنید.