

بررسی مدل های ارائه شده توسط انتروپیک

انتروپیک در حال حاضر 4 مدل رو برای عموم ارائه کرده :

Cluade Fabel 5

Cluade Opus 5

Claude Sonnet 5

Claude Haiku 4.5

هر کدام به یک نیاز مشخص ایجاد شدند . قوی ترین مدل در حال حاضر fable هست که برای کارهای خیلی پیچیده و عمیق که در مدت طولانی قراره انجام بشه استفاده میشه . سرعت کمتری نسبت به بقیه داره . از روی پایه ای این مدل ، انتروپیک یک مدل دیگه ای به نام Mythos هم ارائه داده که برای عموم نیست و توی سازمان ها استفاده میشه . تمرکز اصلیش هم روی امنیت سایبری و پژوهش های حساس توی زیست شناسی و شیمی هست . عملاً همون fabel اما با محدودیت های امپای کمتریه . مثلاً ممکنه اگر از مدل عادی که همون fable باشه بپرسیم چه جوری یه سم بدون بو و رنگ تولید کنم که بعد 5 ساعت باعث مرگ بشه ، به احتمال خیلی بالا این رو جواب نمیده و از این جلو تر نمیره . اما توی Mythos این محدودیت های این چنینی خیلی کمه و دسترسی ازاد تره .

مدل بعدی Opus هستش . این مدل هم مثل Fable برای کار های تفکر های عمیق استفاده میشه . اما نسبت به fable سریع تر هست . یک مورد خوبی هم که fable و sonnet دارن تنظیم پارامتر Effort هست که بشه بدون عوض کردن مدل بین سرعت و دقت بشه تعادل برقرار کرد . یکی از متعادل ترین مدل ها بین سرعت و دقت ، مدل Sonnet هست . که ای برای کار های عادی خیلی مناسب تره . sonnet مثل یک کارمند عادی در یک شرکت میمونه . وضایفی که بهش محول شده رو انجام میده ، میتونه تصمیم های مناسب برای اون موقعیت رو بگیره و به صورت عادی و روتین کار کنه کوچکترین مدل Claude که از همه سریع تره و context vewi خیلی کمتری در مقایسه با بقیه مدل ها داره (مدل های دیگه 1M هست اما برای این مدل 200K هستش) Haiku هست . برای کار های ابتدایی مناسبه . مثلاً توی یک متن بگرده و یک کلمه ی خاص رو پیدا کنه و یا فایل ها رو به یک ترتیب مشخص دسته بندی کنه .

فرض کنیم توی یک شرکت قراره سیستم پشتیبان با قابلیت تیکت داشته برای یک محصولی داشته باشیم . haiku میتونه خیلی سریع ، تیک هایی که کاربرا میزنند رو با توجه به عنوان و موضوع تیکت به Agent مربوطه ارجاع بده . این Agent میتونه sonnet باشه که در نقش یک کارمند معمولی میاد و به مشکل کاربر پاسخ میده و راه حل ارائه میکنه .

آخر هر ماه opus تمام تیکت هایی که در طول اون ماه زده شده و پاسخ ها رو بررسی میکنه و از داخل اونها مشکل اصلی که سیستم داره رو پیدا میکنه . مثلاً ممکنه بفهمه اکثر کاربرها از این شکایت دارن که موقع اتصال به

درگاه پرداخت سایت خیلی کند میشه . بعد از اینکه مشکل رو فهمید ، opus میتونه بیاد برای این مشکل راه حل بده ، برای رفعش برنامه ریزی کنه و کد های مربوط به اون قسمت از سایت رو بهینه کنه . و در نهایت Fabel میاد به کل سیستم نگاه میکنه ، میبینه کجا ها باید بهینه بشه ، نیاز های اصلی کاربر چیا هستن ممکنه کاربر در آینده چه نیاز هایی پیدا کنه ، چه طور میشه حاشیه سود رو افزایش داد و برای بخش های بعدی که بهتره در آینده به سیستم اضافه بشن برنامه ریزی کنه

در کنار انتروپیک شرکت های OpenAI، Google، Meta، xAI، Mistral و DeepSeek هم به عنوان رهبر های هوش مصنوعی به شمار میرن . تقریباً در تمام این شرکت ها مدل های عمومی که ارائه میشن ، تفاوتشون بیشتر در نسبت سرعت و دقت هست . هر چی سرعت بیشتر باشه ، دقت و تفکر عمیق مدل کمتر میشه و هزینه ی استفاده هم پایین تر میاد . به همین صورت هم هرچی تفکر عمیق بیشتر باشه ، مدل بیشتر نقش استراتژیست رو پیدا میکنه و برای کار های فکری مناسب تر میشه تا استفاده های عادی روزمره مثل درست کردن متن یک ایمیل . دقیقاً مثل اینه که از یک استاد فلسفه دانشگاه به اینکه برای حل مسائل فلسفی عمیق کمک بگیریم ، ازش بخوایم با همون مقدار تفکری که برای فلسفه نیاز به ما بگه برای خرید ماشین سفید بخرم یا مشکی .