谷歌新Gemini AI模型在基准测试中击败GPT-4o

作者：Tristan Greene，CoinTelegraph；编译：陶朱，本站

生成式人工智能基准测试领域又出现了一位新霸主，它的名字是 Gemini 1.5 Pro。

之前的冠军 OpenAI 的 ChatGPT-4o 终于在 8 月 1 日被超越，当时谷歌悄然发布了其最新模型的实验版本。

Gemini 的最新更新没有大张旗鼓地发布，目前被标记为实验性的。但它很快引起了社交媒体上人工智能社区的关注，因为有报道称它在基准测试分数上超越了竞争对手。

## 人工智能基准

自 GPT-3 发布以来，OpenAI 的 ChatGPT 一直是生成式 AI 的标杆。过去一年左右，其最新模型 GPT-4o 和最接近的竞争对手 Anthropic 的 Claude-3 在大多数常见基准测试中都遥遥领先于大多数其他模型，几乎没有遇到任何竞争对手。



来源：大型模型系统组织。

最受欢迎的基准测试之一是 LMSYS Chatbot Arena。它测试各种任务的模型并分配总体能力分数。GPT-4o 的得分为 1,286，而 Claude-3 获得了可观的 1,271 分。

Gemini 1.5 Pro 的先前版本得分为 1,261。但 8 月 1 日发布的实验版本 (Gemini 1.5 Pro 0801) 得分高达 1,300。

这表明它总体上比竞争对手更强大，但基准测试并不一定能准确反映 AI 模型能做什么和不能做什么。

## 社区兴奋

在没有更深入的比较的情况下，我们正进入一个 AI 聊天机器人市场已经足够成熟，可以提供多种选择的时代。最终由用户来决定哪种 AI 模型最适合他们。

据传，Gemini 的最新版本引起了一波兴奋，社交媒体上的用户称它“非常好”。一位 Redditor 甚至写道，它“完全胜过 4o”。

目前尚不清楚 Gemini 1.5 Pro 的实验版本是否会成为未来的默认版本。虽然截至本文发表时，它仍然普遍可用，但它处于早期发布或测试阶段这一事实表明，出于安全或协调原因，该模型可能会被撤销或更改。