Anthropic周一发布了一款名为Sonnet 5.5的新人工智能模型，这是该公司CEO阿莫代伊呼吁放缓开发节奏以来的第二次发布。

Sonnet 5.5是Anthropic推出的更快、成本更低的产品，该公司表示，它在编程、完成范围明确的任务以及创建精致文档、幻灯片和电子表格方面优于前代产品。此次发布距离Anthropic推出更昂贵的模型Opus 5.5不到一周，公司周一还表示，该系列中最便宜的Haiku 5.5即将推出。

Anthropic研究产品经理Theo Chu在接受采访时表示：“Sonnet真正面向的是注重成本的客户，他们可能不需要那么高的智能。它可能用于只需要执行、但不需要Opus所能带来的那种判断力的常规任务。”

近几周，随着越来越多研究人员警告AI可能造成灾难性伤害，Anthropic及其竞争对手的安全与安保实践受到了严格审视。作为回应，阿莫代伊本月早些时候敦促AI公司放缓改进其最先进模型的速度，此举震惊了科技界和政府官员。

该公司周一表示，Sonnet 5.5并未推进其模型能力的前沿，因此其大部分对齐测试都集中在“适用于任何能力水平模型的一组针对性风险”上，根据一篇博客文章。在AI行业中，对齐模型是指行为符合人类利益和价值观的模型。

不过，Anthropic表示，Sonnet 5.5的网络安全能力较前代产品有“大幅提升”，因此它是首款以类似后备保护措施和网络安保措施推出的Sonnet模型，这些措施是Anthropic为其能力最强的模型开发的。

Chu说：“从一开始，关注对齐和安全就是我们使命的关键部分。这一直是我们所有模型都优先考虑的事情。”

Sonnet 5.5的定价为每百万输入token 2美元，每百万输出token 10美元，是Opus 5.5价格的一半。Anthropic表示，Sonnet 5.5每项任务所需的token也比上一版本更少，因此运行成本更低。

该模型周一已在所有平台上线，包括亚马逊云科技、谷歌云和微软Azure。