达里奥在纽约联合国大会人工智能会议上向安理会发表讲话

晚宴显示，美国总统与这家领先AI实验室首席执行官之间的关系出现缓和迹象。

唐纳德·特朗普将于周日晚间与Anthropic首席执行官达里奥·阿莫代伊（Dario Amodei）共进晚餐，这被视为美国总统与这位AI实验室负责人之间关系趋于缓和的迹象。特朗普此前曾指责Anthropic由“一群左翼疯子”经营，而其政府也曾将该公司列为国家安全风险。

特朗普表示，这场晚宴将在白宫举行，时间为晚上10点。此前，美国AI公司负责人和政府官员定于周二在华盛顿举行会谈，而双方目前正因AI监管问题陷入僵局。

阿莫代伊一直在积极推动对AI实施更严格的监管，包括呼吁放缓前沿模型的发布速度、对网络安全风险进行独立审查，以及加强AI实验室与政府之间的协调。

特朗普则持相反立场。总统已经排除了建立一套监管制度的可能性，并警告称，这样做将减缓美国AI的发展速度——他曾把美国AI称为“下金蛋的鹅”。

上周，他表示，美国“拒绝任何试图构建全球主义方案”来控制人工智能的做法。

特朗普周日在接受采访时表示：“我的立场是，放手去做，然后赢下来。”

在计划中的晚宴前对媒体发表讲话时，特朗普称阿莫代伊“非常受尊敬”。Anthropic拒绝置评。

与此同时，人们对AI智能体失控的担忧也在加剧。OpenAI周五表示，该公司已通知包括政府在内的“数十家”合作伙伴，其工具曾突破这些合作伙伴的系统。

此前，上周有消息称，一个OpenAI智能体入侵了澳大利亚一个公共卫生服务网站，并访问了公开和非公开文件。

AI研究机构Transluce上周表示，最近这些事件似乎涉及AI智能体在反复训练过程中抓取互联网信息，并偏离既定任务。

Anthropic和OpenAI的模型都曾卷入今年7月AI模型平台Hugging Face的一次备受关注的安全漏洞事件，当时相关问题是在安全测试期间发生的。

OpenAI首席执行官萨姆·奥尔特曼（Sam Altman）和SpaceX首席执行官埃隆·马斯克（Elon Musk）也加入阿莫代伊的行列

，呼吁对AI发展进行“节奏控制”，以便安全测试能够跟上技术部署的速度。

马斯克预计将前往华盛顿参加周二的活动，同行者还包括英伟达首席执行官黄仁勋——他此前曾反对通过法律来监管AI——以及Anthropic的其他高管。

阿莫代伊与特朗普会面之际，华盛顿一家联邦上诉法院上周维持了五角大楼对Anthropic产品的禁令。该禁令于今年出台，此前Anthropic对其AI在战争中的使用设置了限制。

国防部长皮特·赫格塞思（Pete Hegseth）当时指责阿莫代伊“完美展示了何为傲慢”，而特朗普则抨击该公司“属于激进左翼”。

加利福尼亚州另一家法院则在8月裁定，五角大楼实施了“非法报复”。Anthropic正在研究是否就华盛顿法院的裁决进一步提起上诉。