

为开放前沿 提供人工智能安全算力。

太平洋算力是一个位于新加坡、仅用于人工智能安全的算力集群，服务于太平洋两岸的开放权重实验室。我们用这些算力对新发布的模型进行评估，并与实验室合作，在安全问题出现时协助解决。

为什么

实验室在能力工作与安全工作之间左右为难，而多数时候，它们选择能力。后训练的回报如此惊人，以至于很难投入扎实的安全工作，因为每一次强化学习训练都直接关系到实验室的成败。

算力引导研究方向，一切都取决于算力如何分配。这些算力只能用于安全。在此范围内，没有人规定研究者应关注哪些风险、采用哪些方法。我们约束的是集群，而不是研究。

一次评估如何进行

发布前，评估针对实验室自行托管的接口运行；发布后，在公开权重上进行。发现问题时，实验室最先知晓，并决定下一步。

我们不发布评估结果，只提供让人们运行评估的基础。若评估过程中出现问题，我们会向实验室提供资源，协助彻底解决。

双向信任

对实验室

权重不会存放在我们的服务器上，实验室可以看到每一次查询。不接受政府资金；私人慈善资金是唯一的资金来源。

对资助方

默认阻止梯度更新；每个例外均单独批准并记录在公开记录上。不持有硬件，项目若停止也不会有资产搁浅。

一览

是什么	仅限安全用途的算力集群，供开放权重实验室、第三方评估机构和学术合作方使用。
在哪里	新加坡。首个 H200 节点已租用并上线。
权重	仅限开放权重。权重不会存放在我们的服务器上。
训练	默认阻止梯度更新。例外由我们的技术团队逐项批准并记录。
记录	每项任务都记录在公开记录上：谁运行、哪个模型、多少算力、哪一类测试。
结果	我们不发布评估结果。实验室最先知晓。
政治	不允许涉及政治议题的工作。任何安全测试都可以，唯独围绕政治的工作除外。
资金	不接受政府资金。私人慈善资金是唯一的资金来源。
费用	发布评估对实验室免费；研究项目按成本价或更低收费。每次发布约 20 万美元，为估算。
托管方	一家成熟的美国 501(c)(3) 机构已提出财务托管，协议正在起草中。
联系	info@pacificcompute.org

公开记录 · 截至 2026-09-01

完整表格、每一行均附来源：pacificcompute.org/zh/ledger

6 / 19

■ 具备独立评估 ■ 仅有模型卡

已审阅的发布中具备独立安全评估。没有任何一次附有开发者自己发布的危险能力评估。尚有 20 次待审阅。

42

近一年被广泛采用的开放权重发布（Hugging Face 点赞 1,000 以上）。

113

近一年全部开放权重发布（Hugging Face 点赞 250 以上）。