精选快讯　软件快讯　2026年11月23日 03:14

OpenAI发布文章提议将困惑度确立为跨模型比较指标并公开预训练数据

要点

- OpenAI于11月22日发布由首席科学家撰写的文章，正式提议将困惑度确立为跨模型开发者比较及形成共同安全标准的指标。
- 为支持这一提议，OpenAI公开了其内部预训练阶段的困惑度测量数据。
- 尽管主张建立统一指标，OpenAI在文中承认将困惑度做成共享指标存在实际困难。

OpenAI于11月22日发布由首席科学家撰写的文章，正式提议将困惑度确立为跨模型开发者比较及形成共同安全标准的指标。文章指出，现代人工智能的智能主要源于训练压缩器（预测器）的过程，而困惑度正是衡量模型在留出数据集上平均预测下一个词元能力的核心指标，数值越低代表预测越准确。

为支持这一提议，OpenAI公开了其内部预训练阶段的困惑度测量数据。数据显示，GPT-3.5在长期跟踪的主要数据集上对数困惑度约为0.99；随后一代模型将该数值降至约0.76，并在此阶段开始产出突破性的数学结果。基于此趋势，文章推算当对数困惑度进入约0.68至0.7的区间时，应当预期出现进一步的能力跃迁。

尽管主张建立统一指标，OpenAI在文中承认将困惑度做成共享指标存在实际困难。主要挑战包括跨实验室比较时的标准差异，以及该指标与具体可执行任务之间缺乏直接对应关系。

报道来源：itsdw

来源网站：https://www.inews.today/