计算机专题研究：GPT4：模型能力提升推动应用升级

GPT-4：多模态确认，在专业和学术上表现亮眼北京时间 3 月 15 日 GPT-4 正式发布，支持图片、文字等多模态输入，以及文本输出。根据相关技术文档，1）模型架构（包括模型大小）、硬件、训练计算、数据集构造、训练方法等细节未公布；2）GPT-4 于 2022 年 8 月完成训练，此后 OpenAI 一直在评估、对抗性测试并迭代和改进模型；3）OpenAI 搭建了开源 OpenAIEvals 模型评估框架，支持现有准则和自定义准则。4）GPT-4API 已开放等待列表（waitlist），价格提升明显。实验结果表明，GPT-4 在各种专业和学术基准上表现出了人类的水平。

技术拆解：构建深度学习堆栈，新增奖励训练模型 GPT-4 项目重点之一是构建大范围可预测的深度学习堆栈。堆栈（stack）能够通过评估小计算量模型的性能，准确预测大计算量模型的性能，减少训练成本。训练方法上，预训练之后，GPT-4 采用了与 InstructGPT 同样的方法进行基于人类反馈的强化学习，并添加了基于规则的奖励模型来进一步引导模型产生人类预期的结果。多模态输入上，支持图片和文本的多模态输入，但是，OpenAI 未在技术文档中给出图片模态的相关技术细节。

安全性讨论：引入专家提高模型安全性和一致性 OpenAI 在技术文档中耗费大量篇幅讨论模型安全性问题。从目前结果看，GPT-4 仍然存在“幻觉”和推理错误，并在模型校准上表现不佳。为了进一步提高模型安全性，OpenAI 聘请了来自 AI 对齐风险、网络安全、生物风险和国际安全等领域的 50 多名专家对模型进行对抗性测试，涉及幻觉、有害内容、虚假信息、

武器扩散、隐私、网络安全等 11 个方面。我们认为，OpenAI 对模型安全性的关注，或是为未来大规模商业化应用做铺垫。

模型能力提升，应用或进一步升级

GPT4 相比 GPT3.5 在多模态、推理能力、支持文本长度方面有了较明显的提升，有望推动应用进一步升级。对多模态的支持有望加速 PDF、图像等领域的生产力应用升级，或将推动生产力应用效率的进一步提升。相关公司包括：万兴科技、福昕软件、金山办公。更强的推理能力与语言理解能力有助于进一步优化服务型应用的使用效果，包含垂类信息的搜索引擎、客服等产品的功能有望进一步升级。相关公司包括：三六零、同花顺。

风险提示：宏观经济波动，技术进步不及预期。本报告内容均基于客观信息整理，不构成投资建议。

关键词：网络安全