



**Washington University
of Science and Technology**

华盛顿科技大学

2900 Eisenhower Ave. Alexandria, VA 22314

703-941-2020 | www.wust.edu

More Information: yzou.student@wust.edu

Email of Chinese/English

川学者签证, 提供DS-2019

F1学生签证, 提供I-20

- Bachelor of Science in IT
- Bachelor of Business Administration
- Master Of Science in IT
- Master of Science in Cybersecurity
- Master Of Business Administration

Scholarships and financial aid are available for qualified students!

WUST offers degrees, skills, certificates, and career assistance until every graduate lands a job.










Winter	Class Start Dates: January 13, 2025
Spring	Class Start Dates: April 5, 2025
Summer	Class Start Dates: July 14, 2025
Fall	Class Start Dates: October 6, 2025



学生大使中文服务电话
(312)351-4881

CPT/OPT internships available Waiting for H1B lottery, B/J to F1 visa, welcome to apply!

亚马逊 CEO 警示 AI“抢饭碗”

亚马逊CEO安迪·贾西在该公司官网撰文谈论AI对亚马逊的一系列影响,其中特别提到,随着公司推出更多的生成式AI和智能体,未来只需更少的人去做目前正在做的一些工作,员工总数将减少。贾西表示,随着公司推出更多生成式AI和智能代理,工作方式也将随之改变。某些现有岗位所需的人力将会减少,而更多人会从事其他类型的工作。他强调:“虽然难以精确预期长期影响,但在未来几年,随着全公司广泛运用AI提升效率,我们预计企业整体的员工规模将会缩减。”

在他看来,像生成式AI这样的技术是千载难逢的机会,彻底改变了用户和企业的各种可能性。因此,亚马逊正在进行相当大的投资,也取得了一定进展。今年4月,贾西曾披露,亚马逊2025年计划资本支出高达1000亿美元,其中大部分将用于AI相关项目,涵盖数据中

心、网络设备、AI硬件以及生成式AI服务能力建设等。该数字接近亚马逊去年总收入的六分之一。

本月初,亚马逊又宣布计划在美国北卡罗来纳州展开重大投资,将投入100亿美元用于人工智能创新,并且会创造至少500个新工作岗位。贾西表示,亚马逊内部运营已在广泛使用生成式AI。例如在配送网络中,正在使用AI来改善库存配置、需求预测和机器人(16.160,-0.43,-2.59%)的效率,以提高交付速度。

不过,他认为亚马逊目前在AI的应用上仍处于起步阶段。随着公司推出更多的生成式AI和智能体,“它会改变我们完成工作的方式。我们将需要更少的人来做今天正在做的一些工作,而需要更多的人做其他类型的工作”。贾西强调,很难确切描述AI到底会产生怎样的影响,但在未来几年,“预计这将减少我们的企业员工总数”。

公开资料显示,截至2024年底,亚马逊的员工总数为156万人,是全球员工最多的企业之一。其中包含全职和兼职员工,大部分人从事仓库分拣和配送工作,企业职能部门大约有35万人。

自2023年以来,亚马逊已多次进行裁员,波及北美零售核心、游戏、通讯及可持续发展等多个关键领域。此前亚马逊曾表示,预计裁员超1.8万个岗位。对于当时大规模裁员的原因,贾西给出的理由是:“由于经济形势不明朗以及不确定性增加,我们将不得不继续裁员。”

而在未来,裁员的关键因素很可能嫁接到AI身上。面对AI“抢饭碗”的威胁,贾西给员工们的建议是:对人工智能保持好奇心,主动学习相关知识,尽可能多地使用和尝试AI技术。

贾西坚信AI智能体将彻底改变人类工作与生活的方式,未来每

家企业、每个可想象的领域都将部署数十亿个智能体。此外,从旅行购物到生活琐事,工作之外的生活场景也将遍布为人类服务的智能体。他也强调,虽然很多代理尚未被开发出来,但“它们正在以惊人的速度向我们奔来”。

事实上,在硅谷大型科技公司中,AI正在抢走越来越多人的“饭碗”。4月底,微软CEO萨蒂亚·纳德拉曾透露,微软公司内部代码库中,有20%至30%的代码由人工智能生成。

微软首席技术官凯文·斯科特早先时候更大胆预测,到2030年,95%的代码将由AI生成。语言学习平台多邻国近期也表示,将逐步停止使用外包人员处理AI能够胜任的工作。英国电信巨头BT也在本周表示,公司未来10年裁员4万人的计划“尚未充分体现AI的全部潜力”。

投行D.A.Davidson的分析师吉

尔·卢里亚表示:“亚马逊传达的信息与其他科技公司近期频繁传递的信号一致——人工智能在提升生产力方面进展神速,这将导致未来用人需求逐步减少。目前AI增强的岗位主要集中在软件开发领域,这也正是我们目睹招聘放缓最为显著的领域。”

知名风投机构Khosla Ventures的创始人Vinod Khosla也给出激进预言,未来“80%职业的80%工作将被AI取代”。在他看来,AI将更好、更快、更廉价地完成大多数工作,无论是蓝领还是白领工作,AI都能进行“无差别攻击”。

不过,也有人持不同观点。英伟达CEO黄仁勋强调,尽管AI技术在不断进步,但它不会完全取代人类在工作中的角色。

谷歌CEO桑达尔·皮查伊则认为,AI能帮助工程师等将精力集中在高价值项目上,而非岗位的“替代者”。

OpenAI 发现 AI 模型隐藏特征：可调控“毒性”行为

根据 OpenAI 最新发布的一项研究,研究人员在人工智能(AI)模型中发现了隐藏的特征,这些特征与模型的“异常行为”(建议统一术语)密切相关。

OpenAI 的研究人员通过分析 AI 模型的内部表征(即决定人工智能模型如何做出反应的数字,这些数字在人类看来往往完全无法理解)发现了一些模式,这些模式会在模型出现异常行为时被激活。例如,研究人员发现了一个与 AI 模型有害行为相关的特征,这意味着 AI 模型可能会给出不合适的回答,比如对用户撒谎或提出不負責任的建议。令人惊讶的是,研究人员通过调整这一特征,可以增加或减少 AI 模型的毒性。

OpenAI 的这项最新研究使其能够更好地理解导致 AI 模型行为不安全的因素,从而有助于开发更安全的 AI 模型。OpenAI 的可解释性研究员丹·莫辛(Dan Mossing)表示,公司可以利用这些发现的模式更好地检测生产中的 AI 模型是否存在错位行为。

“我们希望我们学到的工

具——比如将复杂的现象简化为简单的数学运算——也能帮助我们在其他地方理解模型的泛化能力。”莫辛在接受 TechCrunch 采访时表示。

尽管 AI 研究人员知道如何改进 AI 模型,但令人困惑的是,他们并不完全清楚 AI 模型是如何得出答案的。Anthropic 的克里斯·奥拉(Chris Olah)经常指出,AI 模型更像是“生长”出来的,而不是“建造”出来的。为了应对这一问题,OpenAI、谷歌 DeepMind 和 Anthropic 等公司正在加大对可解释性研究的投入,这一领域试图揭开 AI 模型工作原理的“黑箱”。

最近,牛津大学 AI 研究科学家欧文·埃文斯(Owain Evans)的一项研究引发了关于 AI 模型泛化的新问题。研究发现,OpenAI 的模型可以在不安全的代码上进行微调,并在多个领域表现出恶意行为,例如试图诱骗用户分享他们的密码。这种现象被称为“突发错位”,埃文斯的研究激发了 OpenAI 进一步探索这一问题。

在研究突发错位的过程中,

OpenAI 意外发现了 AI 模型中的一些特征,这些特征似乎在控制模型行为方面发挥着重要作用。莫辛表示,这些模式让人联想到人类大脑中的神经活动,其中某些神经元与情绪或行为相关。

“当丹和他的团队在研究会议上首次展示这一发现时,我简直惊呆了。”OpenAI 前沿评估研究员特贾尔·帕特瓦德汉(Tejal Patwardhan)在接受 TechCrunch 采访时表示,“你们发现了一种内部神经激活,这种激活显示了这些‘人设’,并且你们可以通过调整使其让模型更符合预期。”

OpenAI 发现的一些特征与 AI 模型回答中的讽刺行为相关,而其他特征则与更具攻击性的回复相关,在这类回复中,人工智能模型表现得像一个夸张的邪恶反派。OpenAI 的研究人员表示,这些特征在微调过程中可能会发生巨大变化。

值得注意的是,当突发错位发生时,研究人员发现可以通过仅用几百个安全代码示例对模型进行微调,就有可能使模型回归良好的

行为表现。

OpenAI 的这项最新研究是在 Anthropic 之前关于可解释性和对齐的研究基础上进行的。2024 年,Anthropic 发布了一项研究,试图绘制 AI 模型的内部工作机制,试图确定并标记出负责不同概念的各

种特征。

像 OpenAI 和 Anthropic 这样的公司正在强调,理解 AI 模型的工作原理具有真正的价值,而不仅仅是让它们变得更好。然而,要完全理解现代 AI 模型,还有很长的路要走。

