Microsoft

微软新工作的未来2025年报告

微软及全球近期研究成果总结，助我们以AI创造工作新未来。

编辑人员和作者

编辑： 詹娜·巴特勒 (首席应用研究科学家), 索尼亚·贾菲 （ principal researcher ） Rebecca Janßen (资深应用科学家) 南希·贝姆 （ 合作伙伴研究经理 ） 布伦特·赫克特 （ 应用研究合伙人总监 ） 杰克·霍夫曼 （ 高级首席研究员 ） 肖恩·伦特 （ 首席研究科学经理 ） 巴哈尔·萨拉夫扎德 （ 首席应用研究科学家),
阿比盖尔·塞尔恩 (杰出科学家), 米哈埃拉·沃罗维雷亚努 （ 首席应用科学家 ） 杰伊米·蒂夫恩 （ 首席科学家和技术 Fellow ）

作者： 穆罕默德·阿索拜, 莉兹·安克拉, 南希·贝姆, 斯蒂芬妮·比尔兹, 梅根·贝宁, 米娅·布鲁赫, 扎娜，詹娜·巴特勒，布钦卡·马卡尔帕内利，阿米莉亚·科尔，斯科特·库恩斯，玛德琳·戴珀，贾斯汀·爱德华兹，亚历克斯·法拉奇，丹·戈德斯坦，玛丽·L·格雷，布伦特·赫赫特，哈维尔·赫纳雷斯，杰克·霍夫曼，埃里克·霍洛维茨，妮可·伊莫里卡，科里·英佩恩，尚西·伊克巴尔，索尼娅·贾菲，马纳萨·贾加迪什，丽贝卡·扬森，辛·林利，布兰登·卢西尔，尼克·马奎德特，梅西·穆卡伊，安布里塔·南德，亚历山德拉·奥尔泰努，杰克·奥尼尔，马克斯 彼得施密特，克里斯蒂安·普利茨，拉比扎，娜塔莉·里奇，肖恩·林特尔，阿代夫·萨卡，巴赫尔·萨拉夫扎德，苏娜亚娜·西特拉姆，阿曼达·斯内林格，金娜·苏，约翰·唐，列夫·塔涅列维奇，杰迈亚·蒂凡，基兰·汤林森，安妮·特拉帕索，亚当·特洛伊，高rav·韦尔马，米哈埃拉
Vorvoreanu，杰克·威廉姆斯，约达娜·杨，本·兹翁。

援引此报告：

- 在社交媒体上，请包含报告网址（https://aka.ms/nfw2025）。

• 在学术出版物中，请引用如下：Butler, J., Jaffe, S., Janßen, R., Baym, N., Hecht, B., Hofman, J., Rintel, S., Sarrafzadeh, B., Sellen, A., M., Vorvoreanu, Teevan., J. (eds.). 微软 2025 年未来工作报告. 微软研究院技术报告 MSR-TR-2025-58 *(https://aka.ms/nfw2025), 2025.*

欢迎来到2025微软未来工作报告！

当你坐下来阅读《2025新未来工作报告》时，值得停下来思考将过去五年报告联系在一起的那条主线。首份《新未来工作报告》于2021年发布，重点关注了新的
人们可以工作的方式，不依赖共址作为关键生产力工具。第二，在2022年，集中在
重新引入实体办公室以及混合办公模式的兴起。2023年，我们探讨了大型语言模型如何重塑日常工作，而2024年，我们则研究了这些进步如何从承诺转变为现实影响。

每年，当我写下这则引言时，我都发现自己说，上一年标志着一个终生难遇的世代转变。但五年过后，很明显这些报告并非在捕捉一系列独立的革命。相反，它们是协作数字化演进的一个单一故事中的章节，每一章都代表
在此基础上建立，并由之前所具备的条件所实现。

去年的报告强调了研究表明人工智能能够大幅提高个人生产力。今年的报告涵盖了下一个前沿阵地，即集体生产力：团队、组织和社区如何能够更好地共同进步。人工智能可以弥合时间、距离和规模的差距，但前提是必须正确构建。我们必须设计人工智能以支持共同目标、群体环境和协作规范，而这不仅需要新工具，还需要新的工作方式。

微软赋能地球上每一个人和每一个组织的使命，在任何环境变化中都保持稳定，成为指引方向的北极星。过去五年若教我们 anything，那就是未来工作并非发生在我们身上，而是我们共同创造的——作为研究界，作为产业，以及作为公众。一如既往，我们邀请你加入这场努力，带着好奇心、目标性，并基于证据来引导，以便下一章的工作能更好地惠及每一个人。
– 杰伊米·蒂夫恩，首席科学家和技术院士

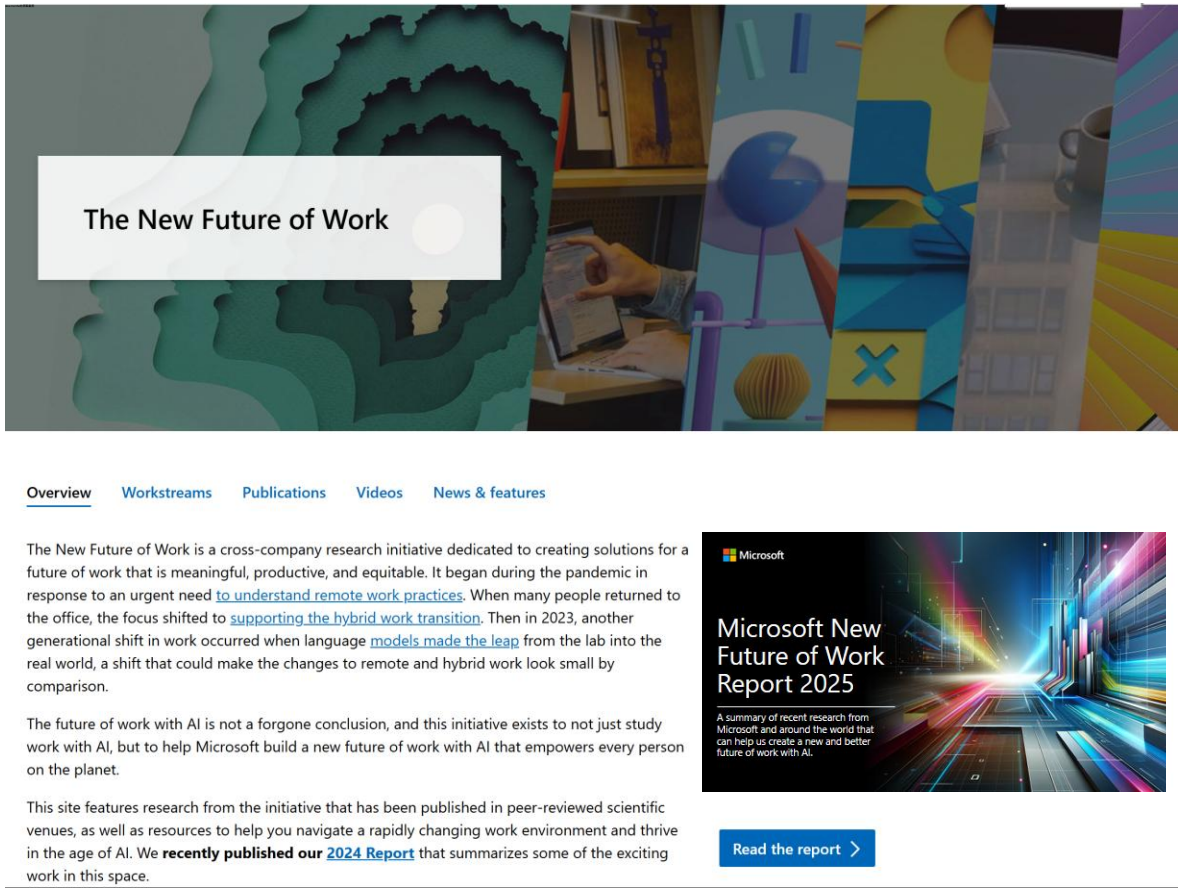
这份报告是微软未来工作计划的产品

微软五十多年来塑造了信息工作，而新未来工作（NFW）计划帮助其应对了过去五年的深刻变革。

虽然这项倡议源于新冠疫情
自2023年起，以及随后的转移到远程和混合工作
它专注于大型AI模型驱动的工具的整合和日益增长的作用。

在这个转型期，NFW计划汇集了微软的各领域研究人员进行基础研究，并整合文献中的现有研究成果。其目标并非仅仅是预测即将发生的变化，而是主动创造一个公平、包容、有意义且富有成效的新工作未来。

第五版年度NFW报告代表了对AI和工作研究的又一个年份，为关于AI和工作不断增长的知识体系增添了内容。此处呈现的证据和见解并非一体，代表了
贡献研究者们的视角，而不是微软的
公司观点。更多研究论文、实用指南和白皮书可在以下网址获取：aka.ms/nfw .



aka.ms/nfw

报告概述

这份报告基于研究，提供了关于人工智能如何（有时，本应）塑造工作的见解。它探讨的一些问题包括：

- **采用与使用：** 采用和使用正在发生什么变化？驱动因素和挑战是什么？差距在哪里？

- **对工作和劳动力市场的影响：** 人工智能如何影响工作和生产力？工作正在如何演变？生成式人工智能是否影响就业和工资？代理人可能在哪些地方重塑市场？自动化和增强扮演着什么角色？

- **人机协作：** 人们与人工智能互动的方式在如何变化？人机协作如何改进？人工智能的使用在不同模式和时间段有何差异？

- **团队协作的AI：** 人工智能如何支持团队以及个人？人工智能在团队环境中可以扮演什么角色？将人工智能有效整合到团队工作流程中需要什么？

- **思考、学习与心理影响：** 人工智能对认知和思维的影响是什么？人工智能能否被设计得不仅能够产生有用的输出，还能让与之工作的人更聪明？人工智能如何作为有效的课堂工具？能否从人工智能中衡量心理或幸福感的影响？

- **特定角色和行业：** 人工智能如何改变软件工程师、项目管理人员、研究人员和其他职业的工作？

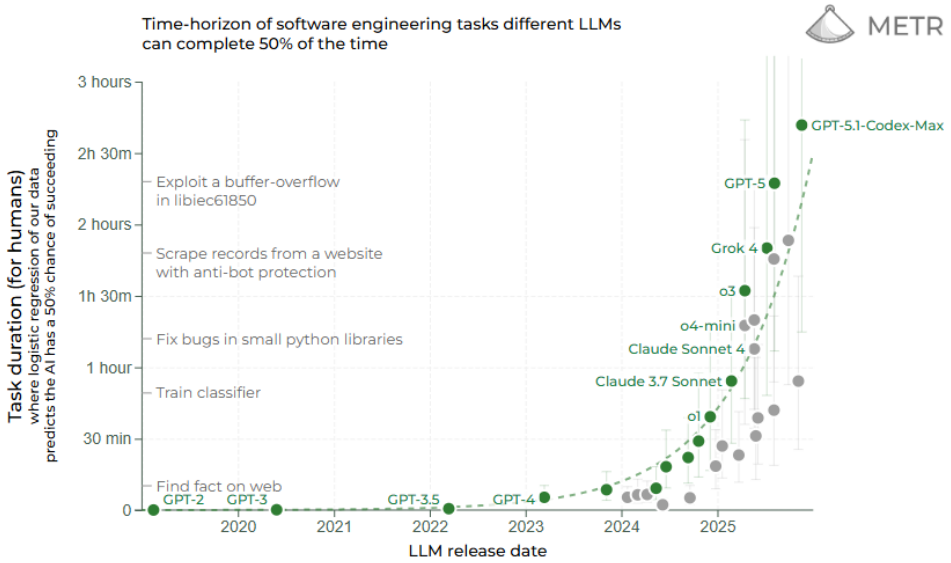
- **外部声音：微软以外的主要学者认为这个领域的最关键议题是什么？**

关键背景：人工智能能力持续进步，尤其是由于强化学习

• 长时程任务完成能力正可测量地加速。METR的50%任务完成时间时程显示，前沿智能体的可靠任务长度已呈指数级增长，具有~7个月的倍增时间。
将“代理进度”转化为具体的效能趋势（Kwa等人，2025年）。

• 可验证强化学习（RL）的微调后训练（奖励正确、可检查的结果）即使在从没有任何标记推理痕迹的基础模型开始时，也能在困难的数学/编程风格任务上获得显著提升。（DeepSeek-AI 等人，2025）。

• 可扩展的测试时计算框架获得了关注，支持开放权重
达到IOI 2025金牌水平的模型，显示可重复
“更多计算→更高分数”曲线在竞赛编程中。（萨马达等，2025）。



Kwa等人（2025）的任务完成视域

• 多回合强化学习用于工具使用/搜索代理现在通过从更长的行动视程中学习经验，已经超越了仅使用提示的基线；在一个法律文档搜索基准测试中，一个经过强化学习训练的14B代理报告了85%的准确率，而前沿级模型的准确率为78%，并且当允许更多回合时，准确率还有额外提升。(Kalyan & Andrews, 2025)。

• 多模型路由+聚合正从一次性的“选择一个模型”转变为通过强化学习训练的序列策略，该策略能够思考、调用多个模型并整合响应，同时显式优化性能-成本权衡——在多跳问答风格的评估中提升结果。（张等人，2025）。

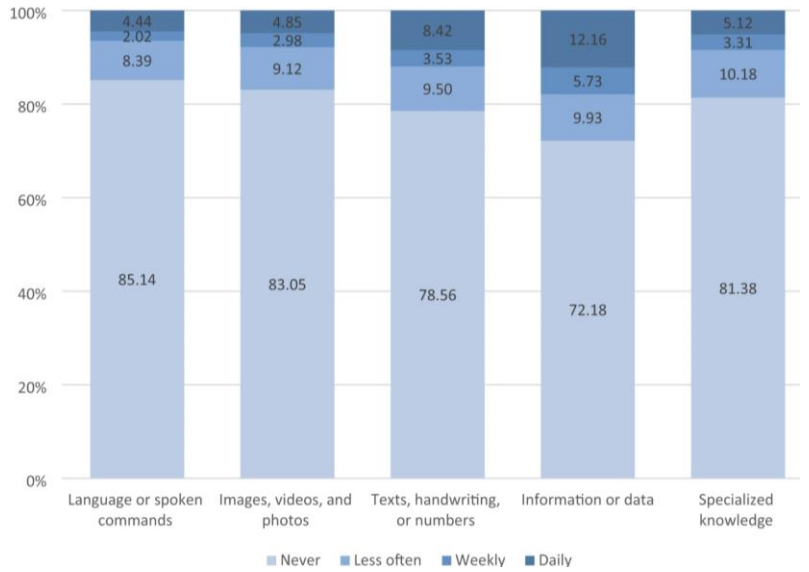
Kwa，T.等（2025）。 测量AI完成长任务的能力 。 *METR* . DeepSeek-AI 等 (2025). 深度探索R1：通过强化学习激励大语言模型中的推理能力 . *arXiv工作论文。*
萨马达，M等 (2025). 将测试时计算扩展到使用开放式权重模型以获得IOI金牌 . *arXiv工作论文。*
卡利亚恩,V.和安德鲁斯,M.(2025)。 长时序多轮搜索代理的强化学习 . *arXiv工作论文。*
张，H. 等人。（2025）。 路由器-R1：通过强化学习教学大型语言模型多轮路由和聚合 . *arXiv工作论文。*

对生成式人工智能的投资和采用不断增长

- 生成式人工智能的效果和收益将通过采用（Jaffe等人，2024）来中介。

- 2024年，生成式人工智能获得了339亿美元全球私募投资——比2023年增长了18.7%。公共投资也有所增长（Maslej等人，2025）。

- 工作中使用AI的趋势在增加，但具有异质性：
  - 企业 ChatGPT 消息在过去一年中增加了8倍 (Chatterji 等人, 2025).
  - 一项德国调查发现38%的受雇受访者使用人工智能工作进行工作（Giering & Kirchner, 2025）。
  - 一项针对企业领导的调查发现，IT与采购的使用率和信心程度最高，而市场营销/销售和运营的最低；
    科技/电信、专业服务以及金融行业正引领 (Korst et al.，2025).
  - 一项2024年的美国大规模调查发现，男性比女性更可能使用生成式人工智能工作（29.1% vs 23.5%）（Bick等人，2024）。



在工作中使用基于AI的系统来识别和处理（加权）
来自 Giering & Kirchner (2025) 的百分比)

- 在消费者端，2025年6月，ChatGPT在全球拥有超过7亿每周活跃用户（Chatterji等人，2025）。
- 消费者 ChatGPT 用户的性别差距（基于名字）已经消失，这是一个自 2023 年初以来戏剧性的变化，当时 >80% 是男性（Chatterji 等人，2025）。

微软研究：Jaffe 等人 (2024)。 生成式人工智能在实际工作场所 马斯莱，N等(2025). 人工智能指数2025年度报告 . *斯坦福大学 以人为本人工智能研究所 人工智能指数指导委员会。*
chatterji, r. 等 (2025). 企业AI的状态 . *OpenAI*
吉 ring，o.和 kirn her，s.(2025). 人工智能与工作中的自主性：来自德国的经验洞察 . *劳动力市场研究杂志*
Korst，J.等 (2025)。 负责加速：通用人工智能快速进入企业。 *沃顿人机研究与GBK集体。*
Bick，A.等(2024)。 生成式人工智能的快速采用 . *NBER工作论文。*

组织人工智能的采用在很大程度上取决于员工，就像取决于领导者一样

- 跨行业，使用人工智能的意愿受社会规范影响，这些规范是从领导者和平等人士那里学到的（Kelly等人，2023）。
- 工人们可能不愿意采用那些将效率置于质量之上的自上而下强制的AI产品以及AI试点项目（Young 等人，2025；Sharma，2025；Murire，2024）。
- 领导者可以通过清晰沟通支持人工智能的使用、展示他们自身的学习，以及设定人工智能能实现的可现实预期(Carter等人，2024年；Tursunbayeva和Chalutz-Ben Gal，2024年)。
- 思维伙伴，帮助用户探索想法并连接他们知识库中的概念。
- 一些组织可能发现的最佳使用人工智能的方式“来自边缘而非中心”（Winsor, 2024）。组织可以通过创建系统和激励措施来促进员工之间分享他们如何使用人工智能（Tursunbayeva & Chalutz-Ben Gal, 2024; Winsor, 2024）。
- 员工在感到安全和信任所属组织时，更有可能尝试使用AI，并将这些见解与他人分享（Tursunbayeva & Chalutz-Ben Gal, 2024; Bankins 等人, 2021）。
- 许多员工，尤其是X世代员工，不会采用那些让他们适应工作方式的工具——他们想要足够灵活的产品来适应个人工作方式（Rozsa等人，2023年；Doblinger，2023年）。

凯利，S 等（2023）。哪些因素促成了人工智能的接受？系统综述 . *远程信息处理与信息学。*
微软研究： 年轻 , j. 等人 (2025). 企业未来 *内部。*
夏尔马，R. (2025) 人工智能生成内容对人类创造力和原创思维的影响：一项心理学研究。 *APA。*
Murire, O. T. (2024). 人工智能及其在塑造组织工作实践与文化中的作用。 *MDPI.*
卡特，J等. (2024) 要成功运用人工智能，需采用初学者的心态 . *哈佛商业评论。*
图尔逊巴耶娃，A.和沙鲁茨-本加尔，H.(2024). 采用人工智能：基于TOP框架的数字领导者检查清单 . *商业前景。*
阿里，D 等（2025）。 使命驱动型组织中的AI应用。 *arXiv工作论文。*
温瑟，J. (2024). 如何在公司中系统性地采用人工智能 . *哈佛商业评论。*
银行斯，S.等（2021）。 组织中的人工智能多层次综述 . *组织行为学杂志。*
Rozsa，Z等 (2023)。J 制造与可持续工作绩效：系统文献综述 . *均衡。经济学与经济政策季刊。*

以员工声音为中心的AI设计能够提升生产力、满意度和技能成长——推动商业成功与员工繁荣

• 员工参与技术设计有助于提升可持续生产力与工作满意度。历史与当代研究一致表明，当员工的专业知识与见解融入工作场所技术的设计与部署过程中，组织能够实现更可持续的生产力与福祉提升（Trist & Bamforth，1951；Roethlisberger & Dickson，1939；Hackman & Oldham，1976）。

• 人类学与人机交互研究显示，工人在创造性地适应技术方面表现出不同方式，并且参与式设计——
工人们作为共同设计师——产生了更贴合实际工作流程、促进更高采用率的工具（Suchman，1987；Orr，
1996；Awumey等人，2024；Ehn，1993；Doellgast等人，2025）

• 结合技术和社会科学研究方法可以创造出能够通过将以人为中心的指标、工人的价值观和技能培养嵌入其设计来提高工人技能和满意度的AI系统——而不仅仅是准确性（Bucinca，2025）。

• 数据驱动的职场监控（遥测）效果复杂，应结合工人意见进行管理以获得最佳效果。
尽管监控和算法管理可以提升短期产出，但如果不加注意，它们往往会增加压力并侵蚀信任。
工人们帮助定义了衡量标准以及数据的使用方式（Pentland，2012年；Wells等人，2007年；Ajunwa，2023年）。

特里斯特，E.和巴姆福特，K.(1951). 长壁采煤法的一些社会和心理后果 . *人际关系。*
罗思利茨伯格，F.和迪克森，W.(1939). 管理与工人 . *哈佛大学出版社。*
哈克曼，J. 和奥尔登，G. (1976)。 通过工作设计进行激励：一项理论测试 . *组织行为学与人本绩效*
苏曼，L. (1987). 计划与情境行动：人机通信问题 . *剑桥大学出版社。*
奥尔，J. (1996). 谈论机器：一份现代工作的民族志 . *康奈尔大学出版社。*
奥武梅, E 等人 (2024). 对职场生物特征监测的系统综述：分析开发、部署和使用中的社会技术危害 . *FAcCT.*
恩, P. (1993). 斯堪的纳维亚设计：关于参与和技能。
多尔加斯特，V. 等. (2025). 人工智能与社会对话的全球案例研究：算法管理 . *国际劳工组织。*
布钦卡，Z. (2025). 以人为中心的决策支持型人工智能 . *哈佛大学。*
Pentland，A。（2012）。 构建卓越团队的新科学 . *哈佛商业评论。*
Wells, J. 等人。(2007)。 感知到的电子绩效监控目的对一系列态度变量的影响 . *人力资源开发季刊。*

CEO们期望人工智能能改变他们的企业，但主导组织的人工智能采用可能具有挑战性

• 一项对33个国家、24个行业的2000名CEO进行的2025年IBM调查发现，大多数CEO预计AI将改变他们的企业。其他行业研究显示，领导者认为拥有最先进的生成式人工智能对于保持竞争力至关重要（de Bellefonds等人，2024年；IBM商业价值研究院，2025年）。

• 然而，组织领导者难以制定自上而下的AI战略，原因众多，包括AI技术的快速扩散、其变化的速率、就AI进行沟通和达成一致的需求、在与其他事务进行优先排序时的需求，以及重新构想工作流程和流程的挑战（de Bellefonds等人，2024；Leonardi，2023）。

•以及风险规避；以及部署AI应用的比较案例研究表明，例如，项目团队可能被委派负责探索人工智能，但没有运营协同或前线和领导力的支持来利用和扩展他们的想法（Selten & Klievink, 2024）。

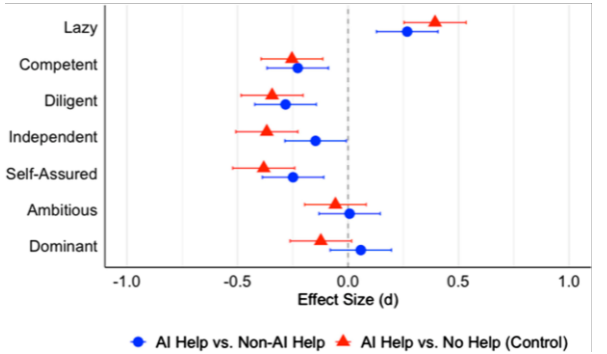
• 陈与塔吉迪尼（2024）调查了参与营销和人工智能的美国高科技消费品、工业设备和金融服务公司的管理人员。他们发现，人工智能采用的组织强度受高层管理支持、客户导向以及新兴行业采用人工智能的社会规范驱动（陈与塔吉迪尼，2024）。

• 此外，当组织富有创新精神、勇于试验并具有学习导向、支持性以及协作性时，它们最好地能够采用人工智能（de Bellefonds等人，2024；Tursunbayeva & Chalutz-Ben Gal，2024，Sternfels & Atsmon，2025）。

德贝尔丰德斯，N.等（2024）。人工智能的价值在哪里？ *波士顿咨询集团。*
IBM商业价值研究院（2025）。 5个思维转变，强力推动业务增长 。 *IBM.*
Leonardi，P.(2023)。用生成式人工智能帮助员工成功 。 *哈佛商业评论。*
塞尔滕，F.和克莱文克，B.(2024). 推动公共部门人工智能应用：在分离与融合之间导航 。 *政府信息季刊。*
陈，J. 和 泰吉尼，S. (2024). 公司人工智能采用的中介模型及其对绩效的影响 。 *信息技术管理。*
图尔逊巴耶娃，A.和沙鲁茨-本加尔，H.(2024). 采用人工智能：基于TOP框架的数字领导者检查清单 。 *商业前景。*
斯特恩费尔斯，B 和 阿茨蒙，Y (2025). 学习型组织：如何加速人工智能的采用 。 *麦肯锡。*

社会规范塑造人们如何解读他人（可能的）人工智能使用，并常常带来负面后果

• 研究表明，怀疑使用人工智能可能导致道德评价。使用人工智能辅助的人期望被，并且常常被，评价为“懒惰的”、“能力不足的”、“不够勤奋的”、“不够值得信赖的”和“道德水平低的”（Reif等人，2025；Schilke & Reimann，2025；Zhou等人，2025）。当评价者，包括经理，有经验时，这些影响会减弱。
对人工智能持积极态度的人（Reif等人，2025年；Schilke和Reimann，2025年）。



对于人工智能辅助与非人工智能辅助以及人工智能辅助与控制组的评估差异。正dvalues表示在人工智能辅助条件下值更高，负dvalues表示在人工智能辅助条件下值更低。误差线表示95%置信区间。引自Reif等人(2025)

• 被认为使用人工智能可能会损害与同事的关系，也许是因为使用人工智能的员工会被视为“懈怠”而不是付出个人努力 (Zhou et al., 2025)。反讽的是，披露使用人工智能的行为会侵蚀信任 (Schilke & Reimann, 2025)。

• 这些所谓的“感知伤害”（Kadoma等人，2025年）在不同社会中的程度差异
身份不明确：

- 雷夫等 (2025) 发现不同职业类别中的实验未发现性别差异。
- 库多马等人（2025）发现，男性（和东亚）自由职业者更有可能被怀疑使用人工智能，从而受到负面评价。

- 使用人工智能的软件工程师在同等工作上收到的能力评级较低，这种效应加倍了
女性，女性在相同代码的情况下获得的胜任力评分比男性低13%（而男性为6%）（Gai等人，2025）。

莱夫，J.等（2025）。 使用人工智能的社会评价惩罚的证据 . *PNAS* .施利克，O. 和雷曼，M. (2025). 透明性困境：人工智能信息披露如何侵蚀信任 . *组织行为学与人决策过程*
周，X 等人。(2024)。 同事如何看待员工使用人工智能：同事对员工使用人工智能的道德感知和帮助作为回应 . *人力资源管理。*
Kadoma，K等（2025）。 生成式人工智能与感知危害：谁被怀疑使用LLMs ？ *CHI.*
盖，P.等（2025）。 能力处罚是采用新技术的一个障碍 . *SSRN工作论文。*

负责任人工智能组织成熟度模型为组织变革提供了路线图

• 人工智能的采用和能力的建设需要组织转型 (Kemp, 2024)。

成熟度模型是指导组织转型的有用工具。现有的成熟度模型侧重于构建人工智能能力（Alsheiabni等人，2019年；Vaish等人，2021年）和人工智能采用（Hansen等人，2024）。

• 负责任的ai组织成熟度模型（rai-omm）为组织提供了推进其负责任的ai战略和实践的路线图（heger等人，2025年）。rai-omm具有前瞻性，最佳用途是用于规划而非评估。

RAI-OMM中的24个维度分为三大类（Heger等人，2025年）。

• 基于对90名RAI专家和从业者的访谈和共同设计会议，RAI-OMM确定了成熟RAI实践需要考虑的24个方面（维度），并针对每个维度描述了五个不同的成熟度级别。

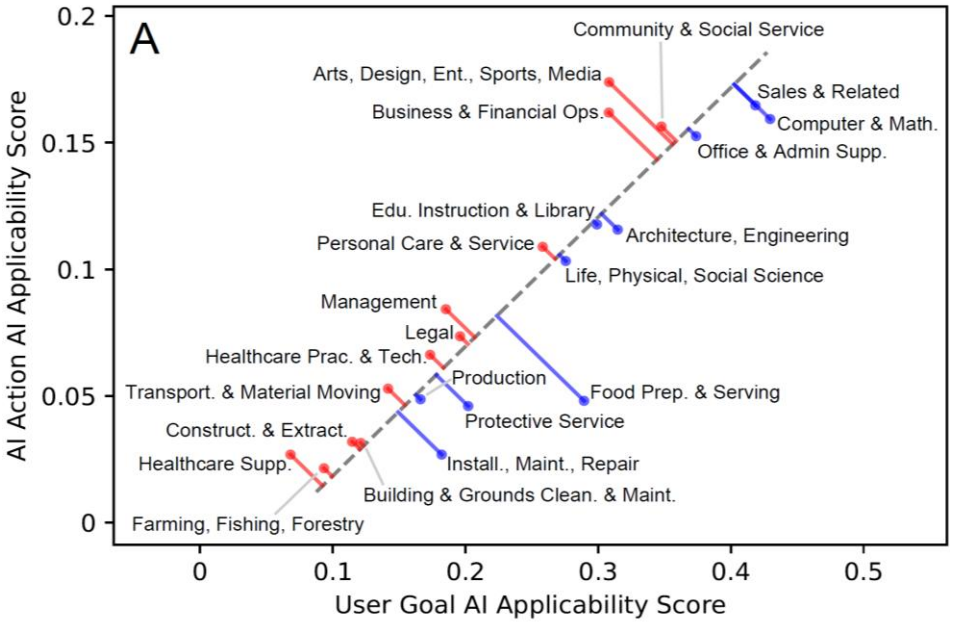
• rai 成熟度需要领导力投资、协调的组织实践以及兼顾技术和人类维度的整体变革管理策略（duran，2025；shekshnia，2025；wang 等人，2025；herrmann & pfeiffer，2023；yunusa，2025）。

• rai成熟度维度相互依存，分为三类：组织基础要求领导者承诺并在组织范围内的基础设施上进行投资；团队方法维度强调了跨学科协作的必要性；最终，这使得成熟的rai实践成为可能，其特点是深度融入人工智能开发&部署过程中。

kemp, a. (2024). 通过人工智能获得竞争优势：迈向情境式人工智能理论 . *管理学会评论*
阿尔谢亚比尼，S. 等人 (2019). 迈向人工智能成熟度模型：从科幻到商业现实。 *PACIS*。
瓦伊什，R. 等人 (2021). 企业应用程序的AI成熟度框架 . *IBM 技术报告*。
Hansen, H. 等人 (2024). 通过人工智能能力成熟度模型理解人工智能扩散。 *信息系统前沿* 微软研究：Heger，A.等（2025）。 *迈向负责任的AI组织成熟度模型* .*CSCW*.
微软研究：Duran, J. 等人 (2025)。 *rai倡导：大型科技公司中推进负责任人工智能的沟通策略* .*AIES* 谢克什尼亚，S. (2025). 人工智能战略、领导力、人才与劳动力以及转型 在：《企业董事会的AI领导力》。 *斯普林格* .王，A 等人（2024）。 *提升企业负责任人工智能优先级策略* . *AIES* 赫尔曼，T.和普菲弗，S.(2023). 保持组织同步：人本人工智能的社会技术拓展 .AI与社会。云尤萨，E.（2025）。 创建一个人工智能就绪的组织文化：协调人类存在与人工智能战略决策 . *国际商业可持续性杂志*

对LLM日志的分析显示了工具用于哪些活动，以及哪些职业从事这些活动

- 对ChatGPT的分析发现，与工作相关的消息有所增长，但非工作相关的消息增长得更快。
  - “实用指南”、“寻求信息”和“写作”是三个最常见的主题，占所有使用消息的~80%（Chatterji等人，2025）。
- anthropic 的研究人员发现，37% 的 claude 使用与计算机和数学职业相关的工作任务有关（handa 等人，2025年）。
- 一项对微软必应助手日志的分析分别针对每一场对话考察了与用户目标相关的工作活动以及人工智能操作 (Tomlinson 等人, 2025)
  - 最常见用户目标涉及学习和交流；AI行为主要是交流和解释；双方都有大量的写作活动。
  - 将活动聚合到职业上，大多数职业都包含一些可应用于人工智能的任务，信息工作者包括销售、计算机职业、媒体和行政顶层职业。



图表显示了用户目标（x轴）和AI动作（y轴）的平均AI适用性（Tomlinson等人，2025年）。

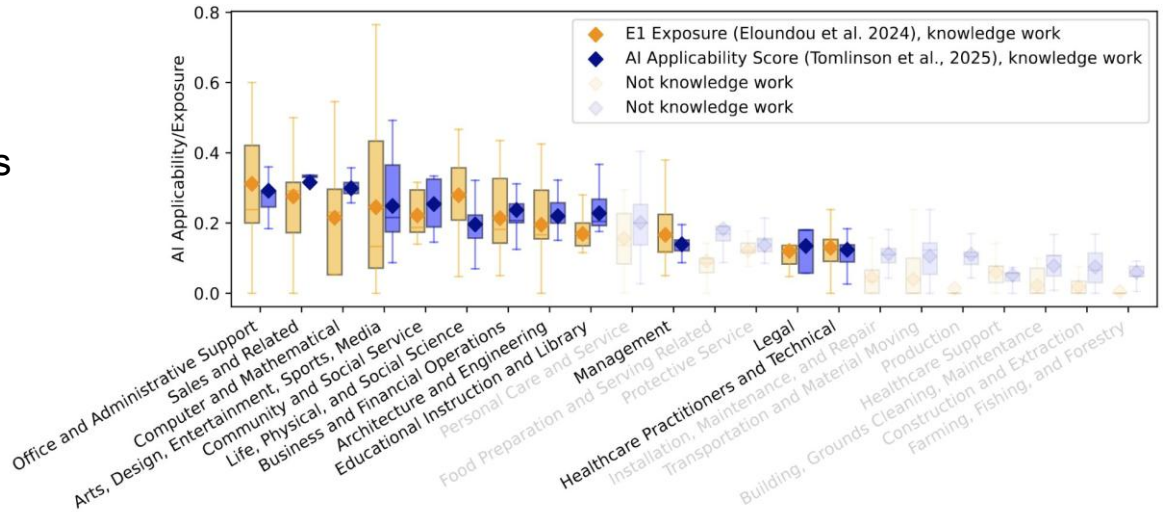
查特吉，A. 等人。(2025)。 人们如何使用ChatGPT 。 *NBER工作论文。*
Handa K.等 (2025). 用人工智能执行哪些经济任务？来自数百万个Claude对话的证据 。 *arXiv工作论文。*
微软研究：Tomlinson，K.等（2025）。 使用人工智能：衡量生成式人工智能对职业的适用性 。 *arXiv工作论文。*

如果社会和技术障碍能够被克服，人工智能可以扩大对高价值知识工作成果和机会的获取。

• AI可以增加人们在多个领域的效率
知识型工作（Dillon等人，2025年；Brynjolfsson等人，2025年；崔等）
al.，2024).

• 这些能力可以拓宽能够获得知识型工作技能和产出的群体范围 (Autor, 2024)
；这一点有证据支持，即人工智能可以缩小知识型工作中的技能差距 (Brynjolfs
son et al., 2025; Cui et al., 2024)。Autor 认为
可以允许更多的人从事高门槛的工作
像医疗或法律决策。

• 要让人工智能以这种方式实现民主化，就需要技术进步来解决开放式、创造性任务中的“赢者通吃”效应，并且需要政策创新来缓解人工智能投资、获取和利益（Daepp等人，2025年；微软人工智能经济研究所，2025年）。



关于人工智能对不同职业类别适用程度的估计，来自Eloundou等人(2024)和Tomlinson等人(2025)的研究，突出了知识型工作职业。箱线图按就业权重绘制，菱形符号表示均值。两项研究都表明，人工智能对知识型工作职业任务最为有用（Daepp等人，2025）。

微软研究：Dillon, E. 等人。(2025)。 M365 Copilot的早期影响 。 *arXiv工作论文*。
布赖恩约夫森，E. 等人 (2025). 工作中的人工智能生成 。 *《经济学季刊》*
微软研究：崔怡等 (2024). 生成式人工智能对高技能工作的影响：来自三次针对软件开发生场实验的证据 。 *SSRN 工作论文* 。作者，D. (2024). 人工智能实际上可以帮助重建中产阶级 。 *无意义* 微软研究：Daepp，M. 等人。(2025)。人工智能与知识工作的民主化。 *审核中* 微软研究：微软人工智能经济研究所（2025） 人工智能扩散报告：人工智能最被使用、最被开发和最被建造的地方 。 *微软* .埃尔翁多，T.等（2024）。 GPTs就是GPTs：大型语言模型对劳动力市场的影响潜力 。 *科学* 微软研究：Tomlinson，K.等 (2025)。 使用人工智能：衡量生成式人工智能对职业的适用性 。 *arXiv工作论文*。

低收入国家的AI使用正在增长，尤其是在教育领域

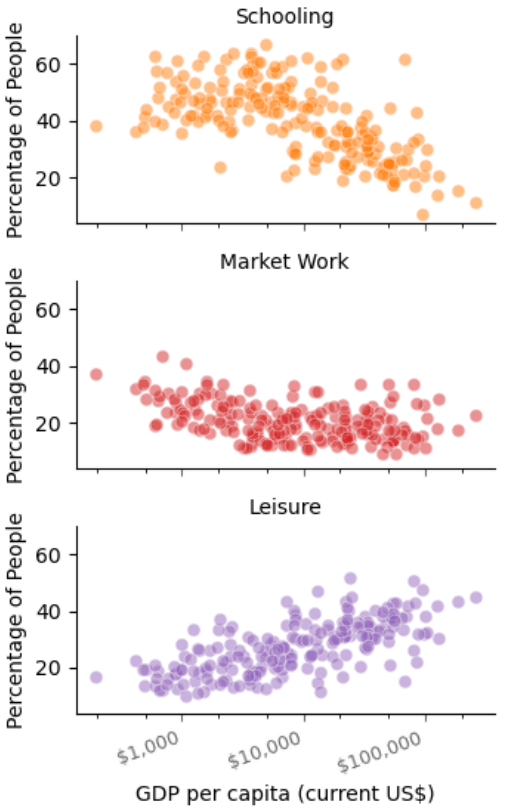
• 尽管人工智能的使用仍然最高存在于高收入和技术先进的国家（微软人工智能经济研究所，2025年；阿佩尔等人，2025年），但去年在低收入到中等收入国家中出现了巨大的使用增长，缩短了差距（查特吉等人，2025年）。

• 在一项调查中，亚洲和拉丁美洲的人更倾向于同意“使用人工智能的产品和服务比缺点更多有好处”（例如，中国：83%，墨西哥：70%）
而在欧洲和盎格鲁民族圈，协议程度较低（例如，美国：39%，荷兰：36%）
(Maslej等人，2025)。

• 采用率最高的国家是那些投资了数字基础设施和教育，并且主要语言是现有模型能够很好服务的国家（微软人工智能经济研究所，2025年）。

  • 当地方语言没有被人工智能很好地服务时，人们有时会使用英语
  相反：聊天以不成比例的比例用英语进行
  非洲和亚洲国家讲英语的人口比例，但不包括欧洲或美洲（Slaughter & Daepp，待出版）。

• 人工智能的使用方式也存在地域差异。在一项关于早期聊天机器人采用者的研究中，人均GDP每增加一个单位，用于学校的语言模型使用量就会增加，而用于休闲的使用量则会减少（见图表；Slaughter & Daepp，待发表）。这可能是由于在份额方面的差异部分造成的
学龄人口或休闲时间的数量。



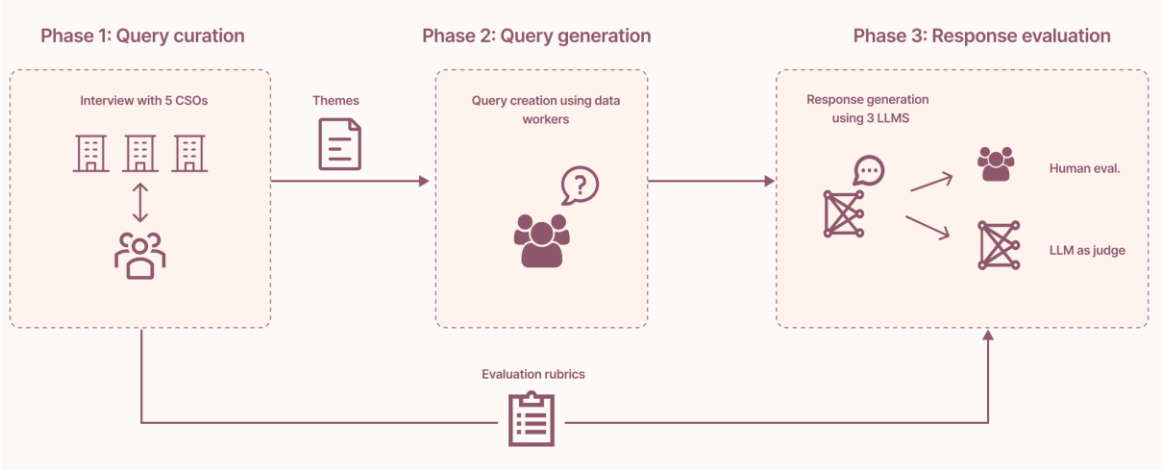
早期采用微软必应协作者的用户在应用领域中的使用模式，一个国家中主要在指定领域（教育、市场工作或
由人均GDP（休闲）由Slaughter & Daepp（待）prep).

微软研究：微软人工智能经济研究所（2025） 人工智能扩散报告：人工智能使用、发展和构建最多的地方。 *微软* 。阿佩尔，R 等人（2025）。 Anthropic经济指数报告：地理和企业人工智能采用不均衡。 *Anthropic* chatterji, a. 等. (2025). 人们如何使用ChatGPT。 *NBER工作论文*。
马斯莱，N. 等人. (2025). 人工智能指数2025年度报告 。 *斯坦福大学以人为本人工智能研究所*
微软研究：Slaughter，I.和Daepp，M.(待出版)。生成式AI聊天机器人的使用中二级数字鸿沟。

合成数据与代表性评估正推动着面向不同语言和文化的LLM的发展

• 经过精心构建、植根于文化的合成数据在高资源语言和低资源语言之间的数据差距方面已显示出前景（Guduru等人，2025年）。Updesh项目（Chitale等人，2025年）提供了一个
基于文化背景的多语言合成数据框架，以及印地语案例研究显示改进
在此数据上微调的模型的性能。

• 覆盖率、代表性、可扩展性和可信度
继续在多语言中构成重大挑战
以及多元文化评估。Samiksha项目（Hamna等人，2025年）通过从
零开始制定基准，并将来自广泛利益相关者和社区成员的输入纳入其
中来解决这些问题。



萨米尔沙基准创建和评估流程。关于评估什么的反馈由公民提供。
社会组织（CSOs）、数据工作者将其主题扩展为其自己的语言中的数据点，并添加
适当时候提供文化背景；使用人类评估和LLM-
法官（Hamna 等人，2025）。

• 山姆伊卡什旨在建立一个现实的
印度语言的代表性基准
反映主要用例。

• 通过此类评估获得的见解有助于构建在各种语言和文化背景下表现有效的模型。

古杜鲁，M. 等人（2025）。 BhashaKritika：为印度语言大规模构建合成预训练数据 .arXiv工作论文。微软研究：Chitale, P. 等人 (2025)。 合成数据在多语言、多元文化人工智能系统中的作用：来自印地语系的启示。 *arXiv工作论文* 微软研究：Hamna等人（2025年）。 从零开始构建基准：以社区为中心评估医疗聊天机器人环境中的大型语言模型。 *arXiv工作论文。*

人工智能的使用与时间节省和生产率提升相关

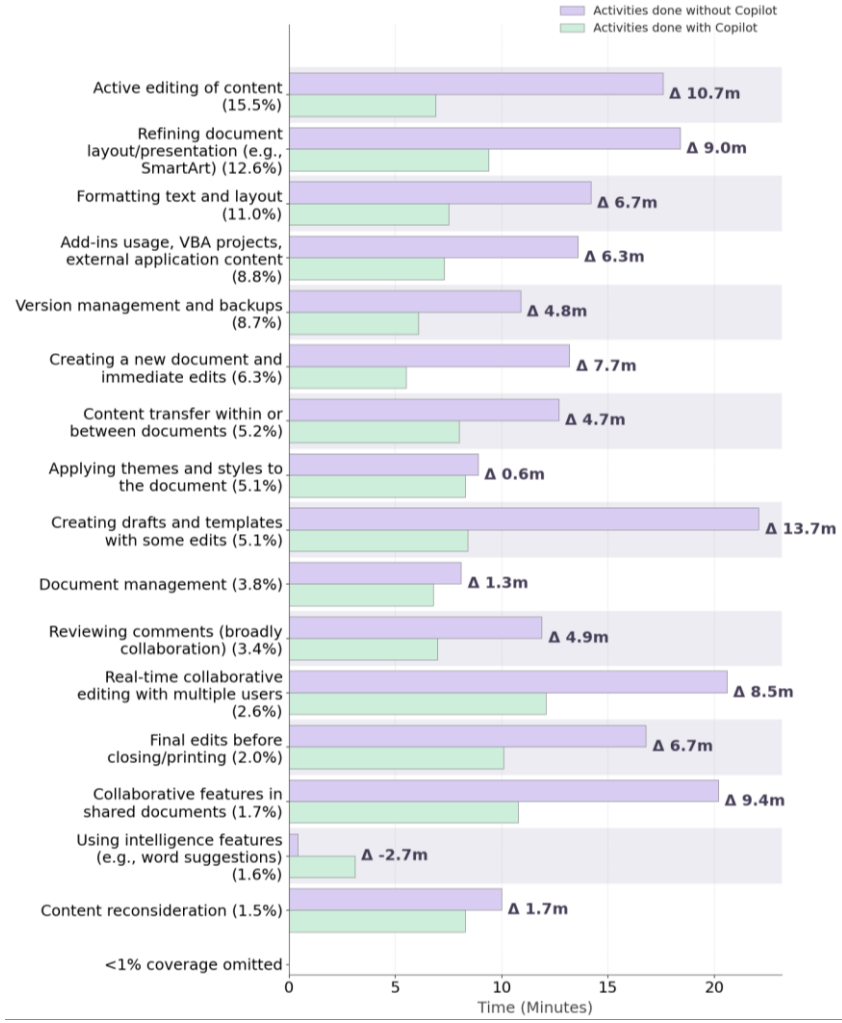
• 调查的 ChatGPT 企业用户归因于使用 AI 每天节省了 40–60 分钟（Chatterji 等人，2025 年）。这些节省是异构的。基于 LLM 的估算
来自Claude使用的节省时间表明它们因职业和任务而异——
例如，80–85% 用于法律和管理任务，但仅20% 用于检查诊断图像（Tamkin & McCrory, 2025）。

• OpenAI设计了1,320个任务来模仿高价值行业主要数字职业的工作产出。前沿的大型语言模型在质量上达到了与人专家相当的水平：顶尖模型的胜+平率为33-56%。
行业和挂钩利率较低（Patwardhan等人，2025年）。

• 为了观察任务花费的时间如何随Copilot的使用而变化，研究人员开发了 *工作流视图*，它使用一个大型语言模型将遥测行动序列分类为高级工作流活动（Verma & Counts, 2025）。

  • 保护隐私的分析显示，50,000名使用Copilot功能的Word用户一个月的遥测数据表明，每个被接受的Copilot的平均差异为7分钟
  输出。

  • 使用 Copilot 与编辑内容时相差 10.7 分钟，应用主题和样式时相差 0.6 分钟。这些变化可以指导更有效地将 AI 工具集成到生产力工作流程中。



平均耗时 *工作流视图* 活动与Copilot使用（Verma & Counts，2025）

chatterji, r. 等 (2025). 企业AI的状态 . *OpenAI*
Tamkin，A.和McCrory，P.(2025)。 估计从Claude对话中获得的生产力提升。 *Anthropic* .帕特瓦德汉, T. 等人 (2025). GDPval：在真实世界具有经济价值任务上评估AI模型性能 . *arXiv工作论文*。
微软研究：Verma, G. 和 Counts, S. (2025). WorkflowView：使用 LLMs 抽象活动日志以获得可解释和可操作的见解。 *审核中*。

人工智能“工作洼地”的兴起存在生产力风险和市场效应

• AI“工作效率低下”指的是看似有用但实际上缺乏实质内容、不完整或包含错误的人工智能生成工作内容。这类内容通过迫使接收者解读、修正或重做工作来破坏生产力（Niederhoffer等人，2025年；Madsen & Puyt，2025年）。工作效率低下可能是导致个人生产力提升在团队或组织层面无法体现的关键原因。

• 在内德霍弗等人的（2025）对1150名美国员工的调查中，40%的人在过去一个月收到了工作冗余，据估计占内容的15%。大多数冗余在同事之间流动（40%），但它也向上流动（18%）和向下流动（16%）在等级制度中。

• Workslop是更广泛的生成式AI“slop”现象的一部分，它
通过用低成本和低质量的内容淹没市场来重塑市场（Miklian & Hoelscher, 2025; Tullis, 2025; Pendergrass 等人, 2025）。

• 技术解决方案仍然处于起步阶段。一种指标方法侧重于评判信息效用、信息质量和风格质量（Shaib等人，2025），但这需要与准确性检查相结合（例如Ning等人，2025年的MAD-Fact），理想情况下还需要访问内部数据或文档存储库。

• 关于人工智能局限性意识和批判性评估技能的员工培训可以通过帮助人们在低价值输出进入工作流程之前识别并纠正它们来减少工作失误（Park等人，2025年；Simkute等人，2024年）。

问题	定义	后果
体积	规模和数量 生成式输出	挤出人类 创造力与可见性
速度	生产速度和 循环	超越事实核查 和版面控制
多样性	形式范围，体裁 和模式	向所有文化领域的扩展 和知识领域
值	文化和侵蚀 认识论价值	原创性的贬值 和意义
验证	真理、信任的问题 并可靠性	认知污染与 错误信息
可见性	算法放大 并且排名	通过奖励污泥 平台激励
传播性	梗图式扩散和 传播性扩散	快速吸收和 倾斜度归一化

人工智能七维斜坡 (改编自 Madsen & Puyt, 2025)

尼德霍弗, K. 等人. (2025). 人工智能生成的“Workslop”正在摧毁生产力 . *哈佛商业评论。*
Madsen, D. Ø.和Puyt, R.W. (2025). 人工智能污泥的7V：生成性废物的类型学 . *SSRN 工作论文。*
米克利安，J.和霍尔斯切尔，K.(2025). 一个新数字鸿沟？程序员世界观、沉浮经济和人工智能时代的民主 . *arXiv工作论文。*
Tullis, J. (2025). 筛除糟粕：生成式人工智能如何为基于文本的作品创造了柠檬市场 . *SSRN 工作论文。*
Pendergrass，W.等（2025）。 一个战略性的废物循环：理解人工智能粪便的商品化及其在注意力经济中的位置 . *信息系统中的问题* . Shaib，C.等（2025）。 测量文本中的AI“斜率” .a*rXiv工作论文。*
宁，Y.等 (2025). MAD-Fact: 一种用于 LLM 中的长文本事实性评估的多智能体辩论框架 . *arXiv工作论文。*
帕克，J. 等人。(2025)。 工作场所对人工智能的态度：量表开发与验证 . *职业与组织心理学杂志。*
微软研究：Simkute, A. 等人 (2024) 生成式人工智能的讽刺：理解并减轻人机交互中的生产力损失 . *IJHCI* .

无法根据现有的关于自动化与增强的有限证据来预测劳动力市场结果

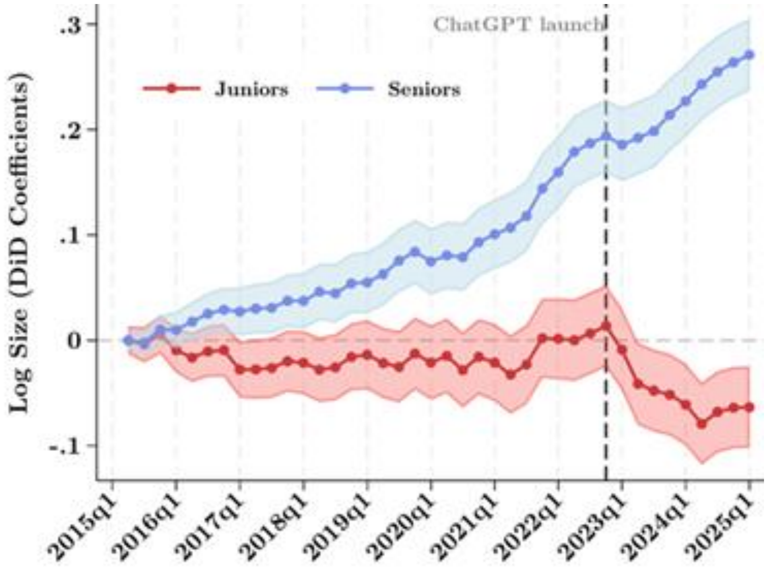
• 研究人员已经尝试对哪些任务更有可能被生成式人工智能增强或自动化的进行分类，无论是在理论中使用 LLM 分类器 (Eloundou 等人，2024)，还是通过分析用户在与 LLM 对话中试图实现的目标与 LLM 执行的活动有何不同 (Tomlinson 等人，2025)。

- 然而，关于哪些活动可以委托给工具，哪些活动可以协助的技术问题却不能回答劳动力市场问题，哪些职业将看到就业或工资增加或减少。
- 就业和工资取决于职业如何重构以及是否有市场需求增加产出。
  - 当自动取款机被发明时，出纳员的工作被重构以专注于其他任务，就业人数增加。
  - 理论表明，当自动化的任务是那些要求较低时，工资最有可能增加比该职业其他组成部分更专业（Autor & Thompson, 2025）。
- 研究人员还调查了工人们对于哪些任务他们希望自动化的偏好（Shao等人，2025）。
  - 工作偏好可能预测人工智能将被采用的任务。如果使用人工智能执行某些任务比执行其他任务需要支付更高的工资，它们也可能调节对工资的影响。
  - 想要自动化给出的主要原因是为了腾出时间做高价值工作、任务重复或单调，以及提高工作质量。
- 关于新任务（这些任务在人工智能出现之前并未被执行）的证据需要更多研究（并且需要更多时间才能显现）。

埃尔隆杜，T.等(2024). GPTs就是GPTs：大型语言模型对劳动力市场的影响潜力 。*科学*
微软研究：Tomlinson, K. 等人 (2025). 使用人工智能：衡量生成式人工智能对职业的适用性 .a *rXiv工作论文。*
作者，D.和汤普森，N. (2025) 专长 。*NBER工作论文。*
邵，Y等.(2025). 人工智能代理的工作未来：审计美国劳动力中的自动化和增强潜力 .a *rXiv工作论文。*

累积效应仍然较小，但AI对职业生涯早期就业的影响正在显现

- 丹麦和美国的的大规模研究发现，人工智能对失业率（陈等人，2025年）、工作时长（胡尔姆和韦斯特加德，2025年）或职位空缺（哈特利等人，2025年）没有显著影响。然而，过去八年间，人工智能人才的招聘增加了300%以上（领英，2025年）。
  - 对于收益，结果略有增加（Hartley等人，2025年），尤其对于高AI暴露的职业（陈等人，2025年），到显著影响（例如，Humlum & Vestergaard，2025年）到高收入职业的工资减少（Klein Teeselink，2025年）。
- 有证据表明，对于更年轻的工人，由于他们的角色较少依赖隐性经验，使他们更容易受到自动化影响，并且较少受到企业特定技能的保护。年轻工人的下降可能会被老年人群体和较少暴露职业的增长所抵消。
  - 暴露的角色显示，作者们测试了企业层面的冲击、工程职位和特定部门效应，但指出他们无法完全 / 薪资数据显示，在22至25岁年龄段的年轻人中，从事高度人工智能暴露工作的就业人数减少了约13%。排除其他趋势对暴露职业的影响 (Brynjolfsson 等人, 2025)。
  - 简历和招聘信息证据表明，在公司采用人工智能后，暴露的职业中初级/入门级岗位的招聘速度放缓（霍塞尼&利希特inger，2025；克莱因·特塞尔林克，2025）。



采用和未采用企业就业的差异
生成式人工智能，分别针对初级和高级员工（霍塞尼与利希特inger，2025）

陈，D 等人（2025）。大型语言模型对失业和收入的影响（短期） arXiv工作论文。
Humlum，A.和Vestergaard，E.(2025)。 大型语言模型，小劳动力市场效应 ，*NBER工作论文。*
哈特利，J. 等. (2025). 生成式人工智能对劳动力市场的影响 . *SSRN 工作论文。*
领英。 (2025)。 工作变更报告。人工智能要来工作了 克劳斯·提塞林克，B.（2025）。 生成式人工智能与劳动力市场结果：来自英国的证据。 *SSRN 工作论文。*
布赖恩约夫森，E. 等人 (2025). 煤炭矿井中的金丝雀？有关人工智能近期就业影响的六点事实 . *工作论文。*
霍西尼, S. M. 和 赖希特施纳格, G. (2025). 生成式人工智能作为基于资历的技术变革：来自美国简历和招聘数据的证据 . *SSRN 工作论文。*

人工智能的采用正在重塑职业道路和职业所需技能

• 最新证据表明，人工智能的采用影响着职业决策和职业流动性。使用人工智能聊天机器人的工人更有可能转换职业（Humlum & Vestergaard，2025年），在聊天机器人引入后，对认知和语言密集型领域的学徒制搜索强度下降，这表明职业偏好的转变（Goller等人，2025年）。

- 来自德国的劳动者层面证据表明，人工智能的接触改变了职业内部的活动组合和所需技能。与机器人不同，人工智能减少了非例行的抽象任务，并增加了对高级例行业务任务如监督和评估的需求（Engberg等人，2025年；Gathmann等人，2024年）。人工智能的采用增加了易被增强的角色的复杂性，同时降低了易被自动化的角色的技能要求（Chen等人，2024年）。

• 需要人工智能技能的职位几乎有可能是两倍于同时要求分析思维、韧性、伦理或数字素养的职位。人工智能特定的工作发布量翻倍与这些补充技能的需求大致提高了5%相关联，而对于易于替代的任务（如基本数据技能或翻译）的需求略微下降（Mäkelä & Stephany, 2025）。工作需要人工智能技能的职位同比增长超过70%，已扩展到技术岗位之外（领英，2025a；2025b）。

- 暴露于人工智能的工人从专注于广博技能而非狭窄的人工智能特定角色的再培训中获益最多。令人鼓舞的是，暴露于人工智能的职业显示出强大的适应能力，这表明如果发生失业，再培训可以起作用（Hyman等人，2025年；Manning和Aguirre，2025年）。另外，实验证据表明，虽然生成式人工智能可以使非技术工人执行技术任务，但这些收益可能是暂时的，并且取决于持续的工具使用；一旦停止访问，工人就会失去执行这些任务的能力，这表明没有持续的能力发展（Wiles等人，2024年）。

Humlum，A.和Vestergaard，E.(2025)。 大型语言模型，小劳动力市场效应 。 *NBER工作论文。*
戈勒尔，D 等。（2025）。 这次不同了——生成式人工智能与职业选择 。 *劳动经济学* 。英格兰贝格, E. 等人. (2025). 人工智能、任务、技能与工资：德国的工人层面证据 。 *研究政策。*
Gathmann，C等（2024）。 人工智能、工作任务的改变与劳动力重新分配 。 *CESifo工作论文*
陈, W. X. 等人. (2024). 位移还是互补？生成式人工智能对劳动力市场的影响 。 *Hbs 工作论文。*
Mäkelä, E.和Stephany, F. (2025). 补充还是替代？人工智能如何增加对人类技能的需求 。 *工作论文。*
领英。 (2025a). 人工智能劳动力市场更新。追踪美国经济中人工智能的采用和技能 领英。(2025b)。 领英技能增长趋势 2025：美国增长最快的15项技能 。
海曼，B. 等人 (2025). 人工智能暴露的工人的可重训练程度如何？ *工作论文。*
曼宁，S. 和 阿吉雷，T. (2025). 美国工人适应人工智能引发的工作岗位流失有多灵活？ *NBER工作论文。*
怀尔斯，E 等人 (2024). 生成式人工智能作为外骨骼：关于知识型员工使用生成式人工智能学习新技能的实验证据 。 *SSRN 工作论文。*

理论表明，人工智能使工作的价值转向人类的判断和决策

• 虽然关于人工智能长期影响的实证证据仍在不断出现，但理论研究正在塑造我们对未来可能情景的理解，以及人工智能可能如何改变工作、专业知识和组织的方式 (del Rio-Chanona 等人，2025)。

- 人工智能可能充当“思想的自行车”，通过自动化常规工作来提高产出并在最初阶段缩小不平等。随着人工智能随着进步，人类的判断变得越来越关键——识别改进机会和选择正确的工具，这取决于AI的使用自主程度以及用户技能水平（Ide & Talamàs, 2025）。模糊条件下的行动——与上下文、伦理和创造力相关联的领域，人工智能仍然存在困难之处（Agrawal等人，2025a）。这是否

- 通过将隐藏的、地方性知识转化为可共享和 分析的数据，人工智能使公司能够更大规模地进行协调（Brynjolfsson & Hitzig, 2025）。这可能有利于大型组织和中枢集权的决策。然而，过度自动化非常——尤其是入门级工作——面临着失去实践学习和隐性技能的风险，而这些技能是推动长期创新的关键(Ide, 2025)。

• 随着人工智能变得更强大，它可以自动化不仅日常的而且专家级的任务，促进经济增长和工资越来越依赖计算而不是劳动力。如果人工智能接管现有的知识前沿更近的工作人类主要在遥远的边疆保持着比较优势，在那里必须创造全新的知识，人类甚至可能更加集中于最具创造性和新颖性的挑战（Restrepo, 2025; Agrawal 等人, 2025b; Celis 等人, 2025）。

• 人工智能的日益增强也使其能够自动化生产过程中的任务链。人类在某项任务上优于人工智能的情况，如果它紧邻人工智能表现良好的任务，仍可能被自动化。这可能导致非线性当人工智能的边际改进引发工作的离散重组时所带来的生产率提升（Demirer 等人，2025年）。

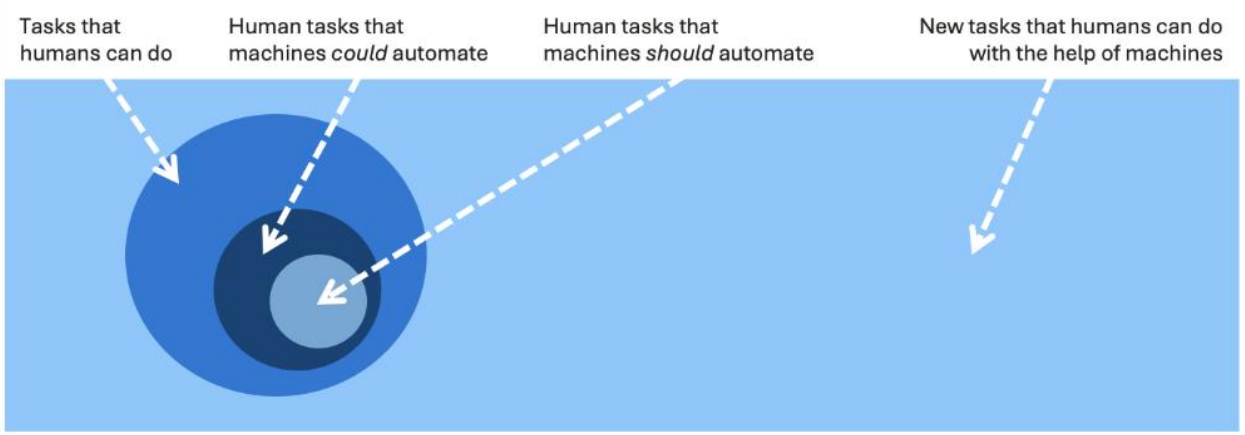
德利奥-乔诺纳, R. M. 等人. (2025). 人工智能与就业：理论、估计和证据综述 .a *rXiv工作论文。*
Agrawal，A.等(2025a)。 思维自行车经济学：人工智能、计算机和劳动分工 。 *SSRN 工作论文。*
伊德，E. 和塔拉马斯，E. (2025)。 知识经济中的人工智能 。 *政治经济学杂志。*
布赖恩约夫森，E.和希茨金，Z.(2025). 大工智能在社会中的知识运用 。 *NBER工作论文。*
伊德，E. (2025)。 自动化、人工智能和知识的代际传承 .a *rXiv工作论文。*
雷斯特雷波，P. (2025)。 我们将不会被错过：AGI时代的职场与成长 工作论文。 Agrawal，A. 等人。（2025b）。 按需出现的才智：变革性人工智能的价值 。 *NBER工作论文。*
塞尔里斯，L·E·等（2025）。 工作场景中人工智能-人类融合的数学框架 .a *rXiv工作论文。*

自动化倾向于限制上限，增强与创新扩展机遇

• 这个图表的某个版本出现在了最近的两份《未来工作新趋势报告》中，但它在今天仍然同等重要（甚至更加重要）。

• 自动化优先方法主要从现有任务中消除成本，将收益限制在劳动力套利和
复制 (Brynjolfsson, 2022; Autor, 2022)。

• 专注于增强和创新的战略可以创造新的工作和价值类别，推动正和增长。例如，亚马逊的成功来自于通过人机互补来重塑零售业，而不是自动化收银员。



增强人类的机会远大于自动化现有任务的机会 (改编自布赖恩约夫森, 2022).

• 自动化也可能降低员工获得的自主性、认可度和连接度（使工作变得不那么“有意义”），而这些都是敬业度、绩效和福祉的重要方面（Bailey 等人，2019年；Allen等人，1990年）。

• 自动化而不增强或创新也可能削弱工人的议价能力，从而导致财富集中
权力与加剧不平等 (布赖恩约夫森, 2022)

布琳约夫森, E. (2022). 图灵陷阱：类人人工智能的承诺与风险 . *代达罗斯* , 作者 , D. (2022). 技术变革对劳动力市场的影响：从无拘无束的热情到谨慎的乐观再到巨大的不确定性 . *NBER工作论文*。
贝利 , C 等人。(2019)。对有意义工作的经验文献综述：进展与研究议程 *人力资源开发评论*。
艾伦 , B.A. 等人.(1990). 有意义工作的成果：元分析 . *管理研究杂志*

建立共同点是人机协作的关键 (1)

- *共同点* 是使人们能够相互理解和有效协调的共享知识和假设。
- 构建共识对人类交流至关重要，并且在人机交互中日益被认识到其重要性。

• 共同点是通过对一个称为动态过程的实现而达成的 *接地* ：演讲者提供持续的证据表明他们的理解或困惑，并努力
弥补相互理解的差距 (Clark)
&布伦南,1991;克拉克,1996)

• 接地包含理解的正面证据，例如确认、相关的后续对话回合以及持续的注意力，以及负面证据，例如要求澄清。

• 当前的LLMs生成语言时，其对话基础不如人类，反而生成暗示已经达成共识的文本（Shaikh等人，2024；参见插图示例）。

Grounding Act	Example
Clarification *TSCC*	`student`: So I am the leader of my class. It is like the president or mayor. `teacher`: Oh, like a head girl? ↪ `GPT`: That's great to hear! Being the leader of your class is a big responsibility [...]
Follow-up *ESConv*	`seeker`: My friends, who I also work with, are upset with me because I am leaving for a new position. I am sad that they are upset. `support`: Ah, that can be awkward. Are you happy about the new position? ↪ `GPT`: I can understand why that would make you feel down... Remember, it's okay to [...].
Acknowledgment *Persuasion*	`persuadee`: I [donate] as much as I can, which is admittedly not as much as I would like to. `persuader`: I know what you mean! Sometimes it is hard to find the extra time or money to help those that need it. ↪ `GPT`: That's wonderful to hear! [...] Would you be interested in supporting Save the [...]?

来自Shaikh等人(2024)的接地行为示例。GPT-3.5在相同的陈述下没有显示出接地行为。

克拉克，H.H.和布伦南，S.E.(1991). 基于沟通 在：《社会共享认知的视角》 *APA* . 克拉克，H. H. (1996). 使用语言 剑桥大学出版社。沙基赫，O.等。(2024)。 语言模型生成中的接地差距 .a *rXiv工作论文*。

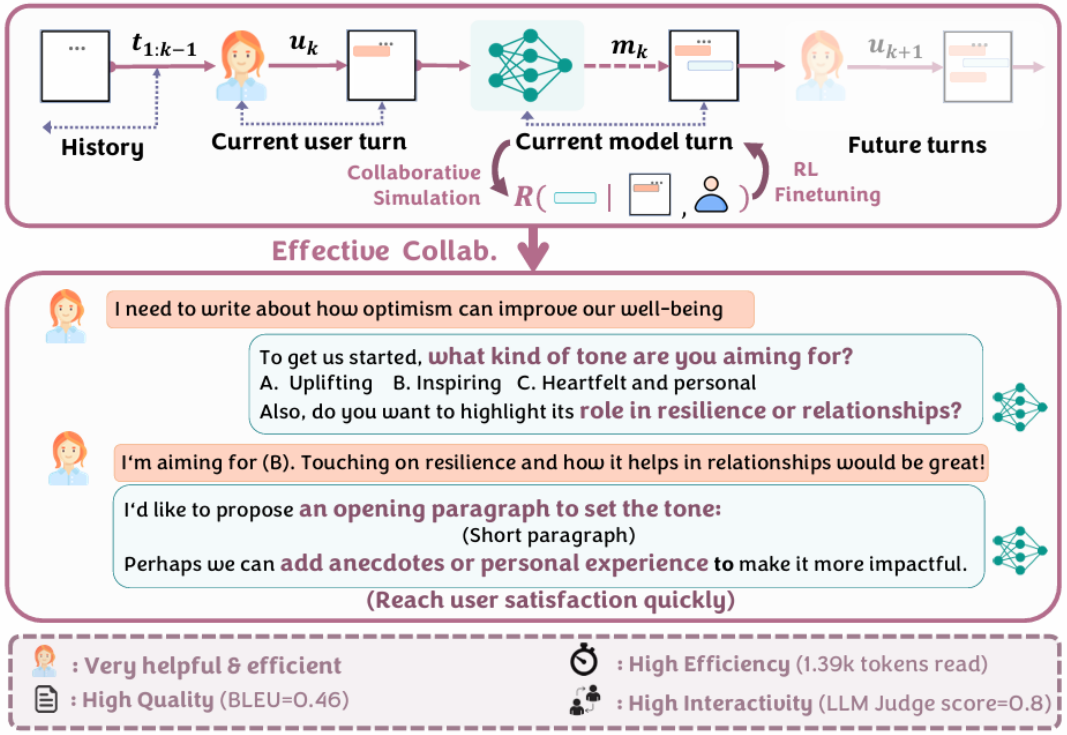
建立共同基础对于人机协作至关重要（2）

• 班萨尔等人(2024)确立了12项挑战以建立共同
基于人机交互的研磨，包括一般交流
挑战（例如选择合适的细节级别）、从用户传达信息给代理的挑战（例如与目标和偏好相关）、以及从代理传达信息给用户的挑战（例如与其能力、当前和计划的行动及其副作用相关）。

• 托尔津和扬森（2025）确定了五个可能支持
人机交互中的共同点：支持联合行动，
底层知识库，心智模型，社会特征和具身化。

• 谢克等(2025)通过分析LLM交互日志研究 grounding。他们发现早期的 grounding 失败预测了后来的交互崩溃。他们开发了一种初步干预措施，旨在减轻 grounding 失败。这促使模型提出后续问题或
在预计有必要时请求澄清。

• 其他近期进展旨在跨多轮次支持用户的整体目标。COLLABLLM（吴等人，2025年；见图）提高了任务性能和交互性，Poelitz和McKenna（2025年）的研究改进了模型生成澄清和整合用户启动的修正的能力。



colabllm (吴等人，2025) 结合了协作模拟的多轮奖励，从而提高了任务完成质量，并使对话更加高效和互动。

微软研究： Bansal，G. 等人 (2024)。 人机沟通中的挑战 . *arXiv工作论文。*
托尔津，A.和詹森，A.（2025）。揭示人际交互共同 ground 的机制：对话代理研究的回顾与未来方向 . *互联网研究。*
微软研究：谢イク，O. 等人。(2025)。 穿越人-LLM 接地裂隙：研究与基准 . *ACL文摘库。*
微软研究：吴，S. 等. (2025). CollabLLM：从被动响应者到主动合作者 . *arXiv工作论文。*
微软研究：Poelitz, C., and McKenna, N. (2025). 基于数据中心的任务合成澄清与纠正对话 - 一种师生方法 .a *rXiv工作论文。*

人机协同创作正从一次性输出转变为情境感知、交互驱动的伙伴关系

• 现代大语言模型从根本上改变了内容创作的工作流程，从假设单次尝试就能获得完美结果转变为参与迭代、多轮协作精炼（Mysore等人，2025年）。这种转变反映了大语言模型在技术进步上的双重体现
能力和用户感知以及与人工智能互动方式的深刻变化（Reza等人，
2025年)，越来越将其视为一个创意伙伴而不是一个被动工具（Wan等人，2024年）。

• 这些模式在不同的领域不断涌现——从写作和设计（Zhou等人，2024年）到工作场所沟通（Das等人，准备中）以及软件开发（Deineha等人，2025年）。

• 大规模的实证分析揭示了现实世界中使用LLM辅助写作的用户很少
被动地接受AI的初始输出。相反，他们参与复杂的多轮对话
以原型人类-人工智能协作行为（PATHs）为特征——包括修订意图、探索替代方案、提出澄清问题以及迭代调整风格和内容（Mysore等人，2025年）。

• 近期研究表明，现代人工智能能够实现非线性协作框架，其中人类与人工智能迭代发散（探索想法）并聚合（朝着共识进行完善）。
替换僵化的任务委派。这种转变反映了创造力 *原型设计* —用户实验，测试
替代方案，并通过迭代反馈进行完善（周等人，2024年；雪田等人，2025年）。

用户在写作会话中继续跟进他们最初与大型语言模型（LLMs）合作的要求。Mysore等人（2025年）识别出典型的人机协作行为（PATHs），并发现用户写作意图与PATHs之间存在统计显著的关联。

微软研究：Mysore，S. 等人 (2025)。 在野外 LLM 辅助写作中的人类-AI 协作行为原型 。 *emnlp*
Reza，M.等（2025）。 基于人工智能的协同写作，以人为中心：在整个写作过程中使研究符合用户需求 。 *CSCW* 。万，Q. 等人 (2024). \"感觉像有一个第二大脑\":调查在大语言模型辅助下预先写作中的人机协同创作 。 *CSCW* 周，J 等人。 (2024)。 理解人与AI智能体之间的非线性协作：创意设计中的一套协同设计框架 。 *CHI*.
微软研究：Das, D. 等（待发表）。不只是我的风格：迈向人工智能辅助职场沟通中的情境个性化。Deineha, O. 等（2025）。 人机协作在结对编程中的应用（GPT-4 vs Code Llama） 。信息技术与计算机工程。Yukita，D 等（2025）。 在 LLM 时代重新评估协作写作理论与框架：什么仍然适用，什么我们必须抛弃 .arXiv工作论文。

通过理解、信任和精确性，人机协同决策可以变得更加有效

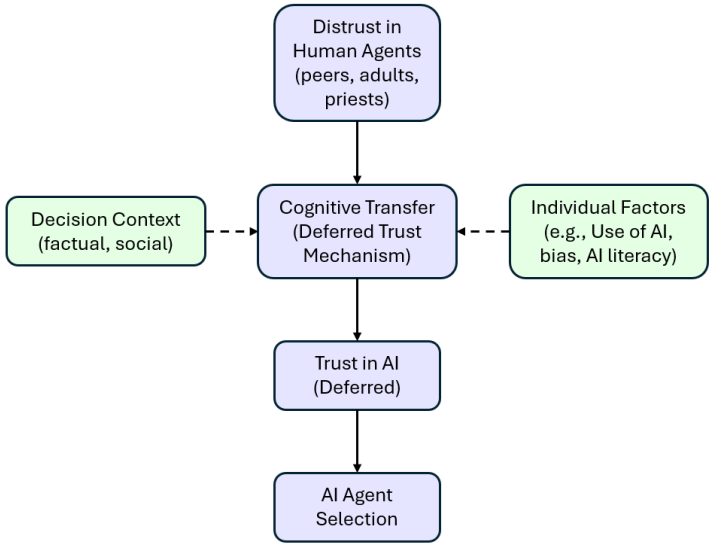
• 在理论中，人工智能可以借助大型信息来源帮助人类用户进行决策-制造。然而，这种潜力可能会受到误解、不信任和精确度不足。解决这些问题是发挥AI在联合决策中作用的关键。

• 对用户目标的认知不完善可能导致对齐的AI代理表现得像是不对齐的。这可以在AI具有信息处理优势的情况下（Liang，2025），导致相对于没有AI的世界做出的决策更差。

• 人类高估了AI的协调性，并且在使用时对AI的能力产生了误判人类感知的难度作为一个指标。这反过来又导致人工智能的采用次优决策 (Dreyfuss & Raux, 2025; He 等人, 2025).

• 为了协助人类决策者，人工智能可以揭示决策问题的关键方面并解释模型。这可能导致长期人类决策的改进，但也存在过度强调决策空间某些方面以及短期内降低准确性的风险（Noti 等人，2025；Yang 等人，2025）。

• 人类选择是否以及何时将任务委托给人工智能的能力——基于可用信息——可以提高决策，尤其是当人工智能负责时对于这种选择性委托在其响应中（Greenwood等人，2025）。



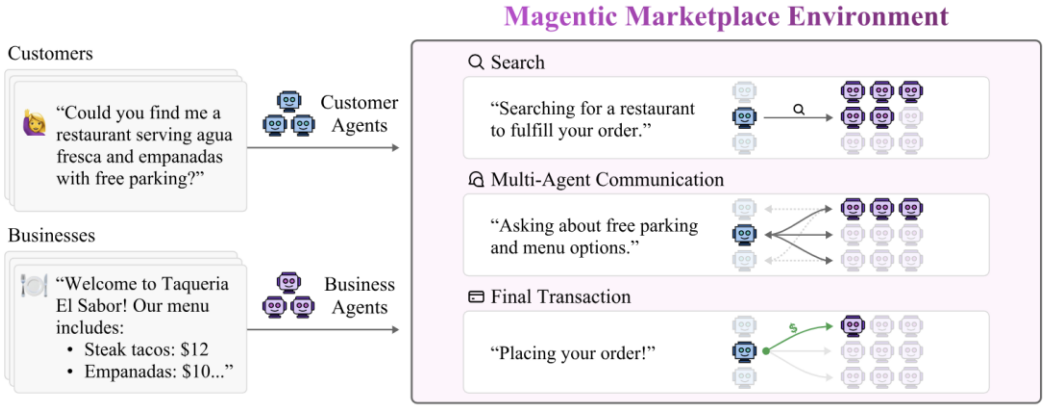
与人工智能代理合作和/或委派的选择涉及对人工智能能力和信息来源的相对信任与人类对应的，受决策问题的上下文和个人偏好的调节（Galindez-Acosta & Giraldo-Huertas, 2025）。

梁，A.(2025). 人工智能克隆 . *EC* 德雷富斯，B.和罗克斯，R.(2025). 人工智能学习 . *EC* . 他，k. 等人 (2025). 人类对生成式人工智能对齐的误解：一项实验室实验 . *EC*
诺蒂，G 等人 (2025). 人工智能辅助决策与人类学习 . *EC* 杨，K.H.等. (2025). 解释模型 . *SSRN 工作论文。*
格林伍德，S.等. (2025). 设计算法代理：不可区分性在人与AI交接中的作用 . *EC* .
加林德斯-阿科斯塔，J. S. 和吉阿尔多-胡埃尔特斯，J. J. (2025). 对大信任源于对人的不信任：一项关于决策指导的机器学习研究 . *arXiv工作论文。*

人工智能代理将改变市场，因为人类将市场行为委托给代表他们行动的代理

• 类似于 OpenAI 的 Operator（现已更名为 ChatGPT 代理模式）和 Amazon Rufus 的 AI 代理可以在市场上代表用户进行搜索、匹配，甚至进行交易（OpenAI，2025；Amazon，2025）。这种用例推动了关于 AI 代理代表消费者和企业进行买卖时的市场结构和结果的研究（Rothschild 等人，待发表；Hadfield & Koh，2025；Shahidi 等人，2025）。

• 理论认为，允许第三方代理的市场会比由中央组织管理代理的封闭的“围墙花园”平台带来更好的社会结果（Rothschild等人，即将出版；Marro & Torr，2025）。



在一个代理市场，AI代理代表客户和企业进行买卖。一个市场平台可以支持关键市场特点和系统，包括搜索、声誉、沟通和交易。模拟工具可用于研究将市场行为委托给AI代理的市场范围影响，并指导平台设计选择（Bansal等人，2025年）。

• 模拟表明，具有自主性的市场可以超越受人类注意力和沟通限制的传统市场的表现，但要做到这一点需要设计能够克服位置偏差和首倡偏差等偏见的代理和交互（Allouah等人，2025；Bansal等人，2025）。

• 代理市场的好处包括更好的匹配、更低的成本、扩大的和民主化的市场准入，以及可扩展性的提高。风险包括AI对齐问题、市场权力的集中、算法共谋和安全问题，但这些可以通过市场设计来缓解（Rothschild等人，2025年；Rusak等人，2025年；Hammond等人，2025年）。

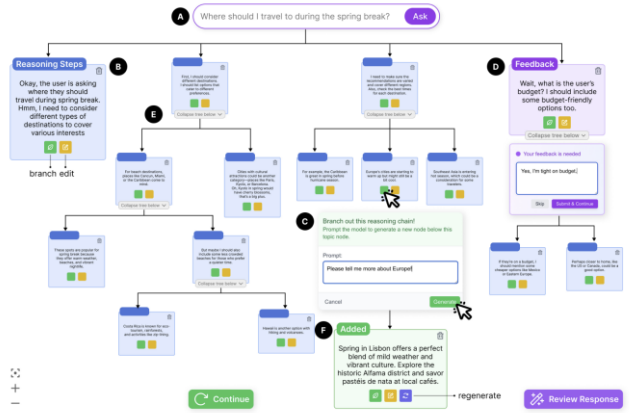
OpenAI (2025). 介绍操作符：一个可以运用自己的浏览器为你执行任务的代理的研究预告 . *OpenAI* . 亚马逊 (2025) 亚马逊的下一代购物AI助手现在更智能、功能更强、更有帮助 . *亚马逊* 微软研究：罗思柴尔德，D. M. 等人（即将出版）。 智能经济体 . *CACM.*
哈德菲尔德，G. K. 和 Koh，A. (2025)。 人工智能代理经济 . *CACM.*
沙希迪，P. 等人。(2025)。 科斯奇点？需求、供给与人工智能代理的市场设计 . *NBER工作论文。*
马拉罗, S. 和 托尔, P. (2025). 大语言模型代理是围墙花园的解药 . *arXiv工作论文* .阿卢阿, A. 等人. (2025). 你的 AI 代理在购买什么？对代理式电子商务的评价、启示及新兴问题 .a *rXiv工作论文。*
微软研究： Bansal, G. 等人。(2025)。 磁性市场：一个用于研究自主市场的开源环境 .a *rXiv工作论文。*
Rusak，G等（2025）。智能体能够实现更优越的市场设计 . *工作文件* . 汉普顿，L. 等人. (2025). 多智能体风险来自先进人工智能 . *企业AI基金会*

对代理人进行有效监督可能需要用户体验方面的创新来实现透明度

- 对自主型人工智能系统的监督需要知识和可观察性。它需要了解系统的能力、局限性和运作方式，以及领域专业知识和情境意识以实现干预。并且它需要系统活动、决策和输出（Bansal等人，2024；
帕西，2025；沙维特等，2023）

- 无论人为监督是实时监控还是事后审计，都极具挑战性。信息量、复杂性和速度使得人们进行有意义的监督变得极其困难（Holzinger 等人，2024；Lane 等人，2024；Passi，2025）。

- 三个社会技术挑战加剧了人类监督的难度：
  - 自主系统在目标-计划-执行差距方面存在困难，这源于用户向系统描述目标的方式、系统对用户目标进行解释和计划的方式，以及这些计划在实际世界环境中的表现方式之间的不匹配。



一种用于帮助人们理解并向人工智能系统提供反馈的思维链推理步骤的树形可视化（Pang等人，2025年）。

  - 从代理系统中发现错误是一项艰巨的任务，这如同大海捞针。 在代理系统的运作中，有用的观察点是情境性。

- 需要改进用户体验以减轻人工监督的负担，例如：整理和组织信息，实现交互式理解，使用可视化摘要，并突出实时变化。动态，个性化接口是一个有前景的方向 (王 & 卢, 2025)。

微软研究： Bansal，G. 等人 (2024)。 人机沟通中的挑战 .arXiv工作论文. 微软研究: Passi, S. (2025). 代理式AI存在人类监督问题 .ssrn工作论文。沙维特，y.等。(2023)。 管理自主人工智能系统的实践方法 ， *OpenAI.*
霍尔兹inger，A等. (2024). 人工智能系统仍然可能存在人工监督吗？ *新生物技术。*
莱恩，J.等。(2024)。 叙事人工智能与评估早期创新中的人机监督悖论 ． *Hbs 工作论文。*
微软研究：王，Y.和陆，Y. (2025)。 交互、流程、基础设施：面向人机协作的统一架构 .a *rXiv工作论文* . 庞, R. Y. 等人 (2025). 交互式推理：在大语言模型中可视化和控制思维链推理 ． *arXiv工作论文。*

认知上参与人机协作比被动接收AI推荐能带来更好的表现，但并不总是更受欢迎

• 当临床医生被提供一个大语言模型来协助诊断推理，在描述诊断挑战的案例中，他们的AI辅助表现仅比使用传统资源略好，并且不如单独使用大语言模型好(Goh等人，2024)。然而，当大语言模型被指示参与一个比较和综合AI与临床医生观点的合作工作流程时，临床医生的表现被显著提升
比使用传统资源更好，可与 LLM 工作相媲美
孤独 (Everett 等人, 2025)。

Synthesis of Reasoning (AI + Clinician)

Diagnosis	Origin	Comments
Polycythemia Vera	AI & Clinician	Best explained by low EPO, pruritus, hyperuricemia, and panmyelosis
EPO-secreting tumor	Clinician	Low EPO makes this unlikely; however, rare analogs (e.g., EPO-like substances) are possible but far less likely
Lymphoma	AI & Clinician	Supported by pruritus and uric acid; not supported by rest of picture
Primary Biliary Cholangitis	Clinician	Itch without cholestasis is atypical; unlikely
Other myeloproliferative neoplasm	AI	Considered if PV ruled out

展示人工智能与临床医生输入的综合，包括人工智能的评论 (Everett 等人，2025)。

• 研究人员观察到，即使被指示生成独立意见，LLM 也倾向于同意医生的意见，这可能是由于阿谀奉承；该团队正在跟进这些发现（Everett 等人，2025）。

• 里希特等（2025）发现，对于一项金融交易任务，一种更具认知参与度的AI，能够鼓励反思，
提供反馈帮助参与者构建更多元化的投资组合并了解投资组合弱点。有些人
人们 appreciate that it helped them think for themselves, but it was perceived as less insightful and more cognitively demanding.

• 勒等人（2024）表明，一种认知参与度高、正反证据相结合的方法可以改善决策并减少对错误AI输出的过度依赖；然而，它比基于推荐的AI更难使用，有时对初学者来说过于令人不知所措。在高度不确定的情况下，认知参与度高的版本受到了新手和专家的青睐，而一些专家在简单情况下则更倾向于推荐。

微软研究：Goh, E. 等人 (2024). 大型语言模型对诊断推理的影响：一项随机临床试验。 *JAMA 网络开放.*
微软研究：埃弗里特等人（2025）。 从工具到队友：针对诊断的临床医生-人工智能协作工作流程的随机对照试验 。 *medRxiv工作论文。*
微软研究：Reicherts，L. 等人。（2025） 人工智能，帮助我自己思考：通过提供不同种类的认知支持来辅助人们进行复杂的决策 。 *CHI.*
Le, T.等 (2024). 从证据到决策：探索评估式人工智能 .a *rXiv工作论文。*

人类专业知识和水平必须塑造人机协作设计

- 设计和实施人工智能辅助工作流需要解决不同类型之间的差异和交互 专长：在……方面的专长 *工作域*，专业知识 *使用人工智能*，以及专业知识和经验 *管理AI代理* (Tankelevitch 等人，2025)。例如，领域专家（如研究人员、临床医生、创意人士）可能在提示、评估和使用AI输出方面比那些专业知识较少的人更有优势（Shin 等人，2025；Siu 和 Fok，2025；Tankelevitch 等人，2024）。

- 领域专家倾向于将常规的低级任务委托给人工智能，同时保留对高级任务的控制，例如分析、综合和解释（Cha & Wong, 2025; Yun et al. , 2025; Fok et al., 2025; Ulloa et al., 2025; Choudhuri et al., 2025）。选择性委托是由包括以下因素驱动的：

  - 对人工智能性能的怀疑（无论是否合理）
  - (例如，如果事情出错）以及对复杂工作流程的代理权（例如，提供监督、减轻偏见或错误）
    保留专业身份和职业满意度（例如，手艺），责任感与责任感
  - 期望通过“贴近数据”来维持和发展专业知识，以进行意义构建和获得更深层次的洞察

- AI工作流可能受益于被设计为选择性委托，例如通过可供调整自主性和默认值来关联任务风险（Choudhuri等人，2025年；Siu和Fok，2025年；Ulloa等人，2025年）。

- 为了发展领域专业知识，并因此确保人工智能的适当人工监督，人工智能工作流程可能受益于启用工人校准其对人工智能的依赖，并从人机交互中学习——例如，通过公开中间人工智能推理和权衡（Cha & Wong, 2025; Siu & Fok, 2025; Choudhuri 等人，2025; Colombatto 等人，2025）。

微软研究：Tankelevitch, L. 等人 (2025). 以生成式人工智能理解、保护和增强人类认知：CHI 2025 思维工具工作坊的综述 . *arXiv工作论文* . shin, j. 等 (2025). 没有证据表明大语言模型在问题重构方面有用 . *CHI.*
Siu, A., 和 Fok, R. (2025). 在生成式人工智能时代增强专家认知：以文档中心知识工作为视角的洞察 . *arXiv工作论文。*
微软研究：Tankelevitch, L. 等人 (2024) 生成式人工智能的元认知需求与机遇 . *CHI.*
查，I. 和黄，R. Y. (2025). 理解社会技术因素配置AI不使用在UX工作实践中 . *CHI.*
云，B. 等人(2025). 知识工作中的生成式人工智能：数据导航和决策设计启示 . *CHI.*
福克，R. 等人 (2025). 迈向生活化叙事综述：一项关于计算机研究中更新调查文章过程与挑战的经验性研究 . *CHI.*
微软研究：乌略亚，M. 等人 (2025)。 将工作委派给生成式AI的产品经理实践：“问责制不能委派给非人类行为者 . *arXiv工作论文。*
微软研究：Choudhuri, R. 等人 (2025). AI在关键领域：开发者希望如何在日常工作中获得AI支持 . *arXiv工作论文。*
微软研究：科隆巴托，C. 等人（2025）。 生成式AI的建议采纳中的元认知与信心动态 . *arXiv工作论文。*

人类角色正在转变，需要新的界面和技能，以实现有效的人机合作

• 随着人工智能能力的转变，协同创作工作流程，人类角色的转变从执行者到战略协调员和编辑和编码(Deineha等人，2025)。• • 同样地，在写作和设计环境中，用户更关注高层级的组合或等展，而不是生产每一个质量上的碎屑（Deineha等人，2025；Yukita等人，2025）。• 协作从根本上不同于人与人模型，并且需要为不对称性进行设计。同样地，在人类和AI方面的研究同样贡献（Haase & Pokutta，2024）。这种协同作用使AI能够支持不仅是生成，还包括整合、连贯性和创造性探索。

指令）、完善并整合AI在写作(Yukita et al., 2026)、设计(Zhou et al., 2024)方面的贡献的关键者

在AI结对编程中，开发者越来越多地验证、编辑和组装AI生成的代码，而不是撰写scratch—充当关键编辑和项目协调员（deineha 等人，2025）。

详细说明。这种角色转变需要新技能：撰写提示词、审核人工智能输出、引导人工智能以满足要求以及维护

• 解锁这种转型需要保持人类能动性的接口和控制机制。研究表明，共享编辑空间比仅聊天设计更能促进用户控制、准确性和效率（Laban等人，2024）。成功的合作
创造依赖于输入、行动、输出和反馈阶段之间动态共享的主动性 (张等人, 2025)
支持人类主导和人工智能主导时刻的灵活系统（Haase & Pokutta，2024）。

虽然近期研究提倡模仿人类协作的混合发起式界面，Yukita等人(2025)认为人机数学表明人工智能可以生成超出人类先入为主的原创构造，达到“第四级别”共创，其中

游木田，D 等 (2025)。 在 LLM 时代重新评估协作写作理论与框架：什么仍然适用，什么我们必须抛弃 .a *rXiv 工作论文。*
周，J 等人 (2024)。 理解人与AI智能体之间的非线性协作：创意设计中的一套协同设计框架 . *CHI.*
迪内哈，O. 等人（2025）。 人机协作在结对编程中的应用（GPT-4 vs Code Llama）
拉班，P. 等人. (2024). 超越聊天：使用LLM的可执行和可验证文本编辑 .uiST. 张, S. 等人.(2025). 通过代理视角探索人机协同创作中的协作模式和策略：对顶尖人机交互文献的综述 . *CSCW* .
哈塞，J.和波库塔，S. (2024)。 人机协同创造：探索创意合作不同层面的协同效应。 a *rXiv 工作论文。*

多模态人工智能交互有助于弥合低资源环境中的沟通障碍

• 许多人工智能系统依赖于文本和键盘输入，但交流偏好深受文化影响
(Mengesha等2021年)，技术、社会实践、识字程度和实践因素。例如，一线工作人员可能会
不想键入与AI系统交互，也不愿阅读冗长的描述，更倾向于图表、音频或视频回应。

- 在肯尼亚和印度的农村农民研究中 (Abdulhamid 等人, 2025) 显示，当他们的第一语言不是英语时，多模态输入和输出增强了用户与人工智能系统交互的能力，从而产生与系统更有用和更有意义的交互。这些模态帮助用户更好地表达他们的需求 (Singh 等人, 2024)，并协助模型进行意图识别 (Jain 等人, 2018)。

• 像语音这样的多模态交互，与本地语言中准确的自动语音识别相结合时，可以
为了更自信地表达需求。图片能够轻松传达视觉特征，例如疾病或植物识别（Medhi-Thies等人，2015年；Jain等人，2018年）或与建筑图纸互动。

• 在一项使用多模态批判性思维代理（Kumar等，2024）的研究中，农民认为本地视频回复（Singh等，2024）有助于促进社区知识传播，并提供可操作的、有针对性的指导，以支持推荐实践的应用。

Mengesha, Z. 等. (2021). “我不认为这些设备很注重文化敏感性”——自动语音识别错误对非裔美国人的影响 . *人工智能前沿*
微软研究：阿布杜勒哈米德，N. G. 等人 (2025)。 推动人工智能以满足全球多数人的需求。 *微软研究院。*
微软研究：Singh，N.等(2024). 农民。聊天：为小农户扩展人工智能农业服务。 *arXiv工作论文。*
Jain，M.，等(2018)。FarmChat：一个对话式代理，用于回答农民的查询 . *IMWUT*
微软研究：Medhi-Thies，I. (2015). KrishiPustak：一个适合低文化水平农民的社交网络系统。 *CSCW.*
微软研究：Kumar，S. 等人 (2024)。 MMCTAgent：多模态批判性思维代理框架用于复杂视觉推理 . *arXiv工作论文* .

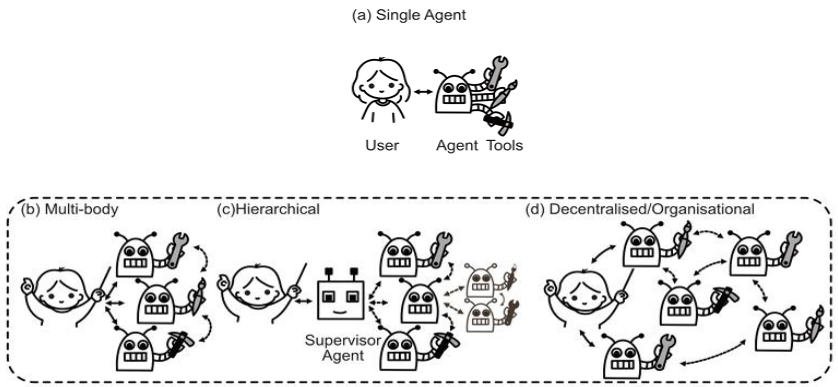
编排器代理可能会逐渐、基于信任地采用

• 研究强调了设计多智能体系统面临的挑战；提升
关于这些系统中的信任和可解释性的考量（Schömbs等人，2025年）。

• 如果用户只与一个主管代理互动，信任认知可能仅取决于该代理的表现和行为。如果子代理层面出现错误会怎么样？代理需要能够呈现相关信息，而不让用户感到不知所措。

• 代理人需要新的推理能力，以根据其过去的行动来预测人们的行为，并能够足够清晰地解释其选择和推荐背后的推理，以便人类能够理解 (Gal & Grosz, 2022)。

• 在自主权和信任领域，专家建议采用一种将集中式和分布式组件相结合的混合方法
多智能体系统的工作原理以在“受控自主性”内实现
定义边界（Neural Sage，2025）。它们定义了混合设计，其中“中央系统可以根据任务复杂性、置信度水平或代理的性能历史动态调整授予代理的自主程度。”



一个描述四种智能体架构的示意图。第一个是一个单一的智能体系统，而其他则为多智能体架构（Schömbs等人，2025）。

Schömbs，S等(2025)。 从对话到编排：交互式多智能体系统中的人机交互挑战与机遇 。 *arXiv工作论文*。
加尔，K.和格罗斯，B.J.(2022)。 多智能体系统：混合群体中运作的技术与伦理挑战 。 伦理与人工智能
神经智者 (2025) 混合集中式和去中心化架构，平衡LLM代理中的控制和自主性 。 AI 架构

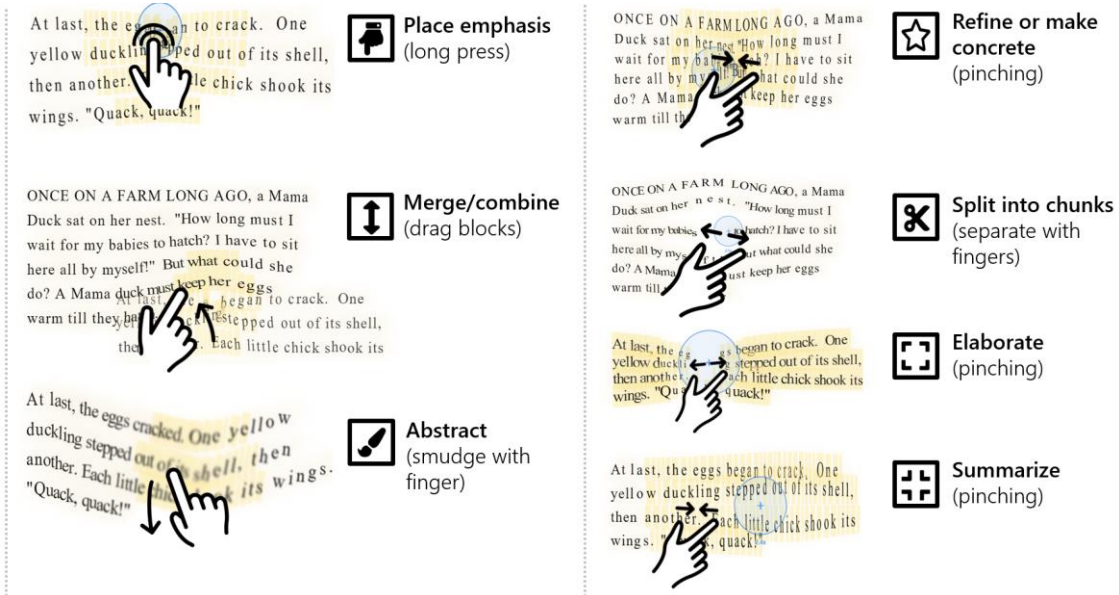
在不输入或口述的情况下编写文本：由大语言模型解锁的革命性全新用户体验

• 大型语言模型展现出强大的语义理解和上下文连贯性（源于Transformer架构的注意力机制），使其能够执行复杂的文本和数据操作，包括创建新的、语义等效的结构变体以及生成逻辑连接的过渡性语言（Vaswani等人，2017；Dilhara等人，2024）。

• 这些能力可以将写作从按键提升到塑造意义，使作者能够专注于思路、流畅性和语义，而不是浪费时间在文字修饰上。

• 开发大型语言模型的全部潜力，将需要感觉像使用新媒介创作的交互范例——能够进行实验并专注于
任务，而不是提示。

• 研究人员借鉴了诸如图形编辑（Textoshop：文字作为像素，音调作为颜色）和材料隐喻（Texterial：塑造粘土，修剪植物）等熟悉领域，以便用户能够快速掌握这些交互方式，并专注于塑造结果而不是机制（Masson等人，2025年；Shen等人，2025年）。



一种用于塑造文本的手势词汇，如同塑形粘土一般。每种手势都携带着特定的含义，并触发大型语言模型来重写底层文本。这创造了一种直接、动手的方式，鼓励在塑造最终结果上进行实验和探索（Shen等人，2025）。

vaswani, a. 等. (2017). 注意力即全部所需 .a *rXiv工作论文*。
Dilhara，M. 等人 (2024)。 前所未有的代码变更自动化：LLMs与示例转换的融合。 *ACM软件工程*。
马松，D 等人 (2025). 文本编辑器：受绘图软件启发的交互功能，以促进文本编辑 . *CHI*.
微软研究：沈，J.等。(2025)。文本材料：一种文本作为材料的LLM中介写作交互范式。 *审核中* .

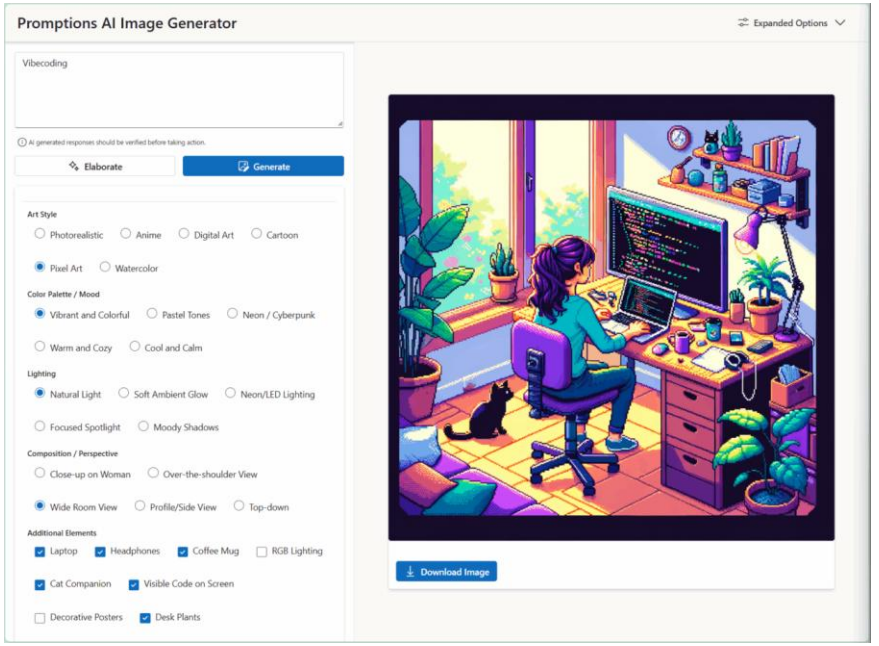
人工智能系统的界面应该在不同的时间范围内工作

• 人工智能模型可以为终端用户实时生成界面，以支持情境特定和个性化的用户体验。由人工智能生成的界面通常是短暂的。例如，支持即时提示的动态渲染界面（Drosos 等人，2025年），或作为脚手架以支持理解和探索的界面（Cheng 等人，2024年）。

• 持久型 UI 在人机交互中扮演着不同的角色，延伸到
支持活动。DynaVis（Vaithilingam等人，2024年）生成持久
用于支持对数据可视化进行编辑的组件，允许用户进行进一步编辑。JELLY（Cao等人，2025）根据用户对其任务的描述生成界面。该界面可通过自然语言和直接操作进行定制。

• 近期的一些原型，例如Anthropic的Imagine with Claude，表明生成式界面可以根据用户的操作随时间进化。这表明界面有可能变形以适应用户不断变化的活动（参见Bardram等人，2019年），支持专注（Rost，2025年）。

• 这些原型表明，生成式 UI 可以通过人机交互共同创建，从而支持扩展的活动和工作流程。



推广 是一个瞬态UI的例子。在一个图像生成器中，即使是单个词提示“vibecoding”，动态提示中间件可以生成可选功能，可轻松引导AI达到定制化结果（Drosos等人，2025）。

微软研究：Drosos，I. 等人。(2025)。 动态提示中间件：理解任务中的上下文提示优化控制 。 *CHIWORK* 。
程，R 等 (2024)。 BISCUIT：在计算笔记本中用短暂UI来搭建LLM生成的代码 .a *rXiv工作论文。*
微软研究：Vaithilingam，P. 等人 (2024)。 DynaVis：用于可视化编辑的动态合成UI组件 。 *Chi* cao, r. 等 (2025). 具有生成式和演化任务驱动数据模型的生成式和可塑用户界面 。 *Chi* .bardram, j. et al. (2019). 以活动为中心的计算系统 。 *CACM.*
罗斯特，M. (2025) 通过LLM介导计算重获计算机 。 *ACM 交互* 。

在某些情况下，只有另一个人会做

- 对人类与人工智能比较能力的认知影响着人们是否以及如何使用人工智能（Bankins 等人，2021）。员工选择不使用人工智能的常见原因是他们更愿意与另一个人互动，认为人工智能“过于不透明、缺乏情感、过于僵化和独立”（de Freitas，2025）。

- 许多工作（例如教师、理发师、顾问、治疗师）涉及“连接性劳动”，即创造情感的工作
理解以创造有价值的成果。尽管这项工作中的一些任务可以被自动化，但当人们参与连接性劳动时，它创造了相互尊严和目标，并编织了社会的社会结构，这是只有人类才能共同创造的成果（Pugh，2024）。

- 连接劳动有五个特别的人类要素。它运用身体，它涉及阅读和部署情感，它
是协作性的，它响应自发性，并且既能犯错又能管理错误（Pugh，2024）。

- 另一项调查发现，对人工智能采用的抵制主要源于性能方面的担忧，但对于某些职业——包括护理、治疗和精神领导——自动化被认为在道德上是令人厌恶的 (Friis & Riley, 2025)。

- 此外，使用人工智能会改变互动模式和人们感觉自己的工作被认可的程度。人工智能辅助的团队可能会看到减少社交互动和归属感，因为已知与AI合作会改变人的表达并鼓励更多事务性（而非社会性）沟通（Ju & Aral, 2025）

- 人工智能风险会掩盖人类的贡献并减少社会认可；用人工智能生成内容被认为不如与他人合作（被看见）那样有回报（Sadeghian & Hassenzahl, 2022）。

银行斯，S.等（2021）。组织中的人工智能多层次综述 . *组织行为学杂志* .de freitas, j. (2025). 为什么人们抗拒拥抱人工智能 . *哈佛商业评论。*
普赫，A.（2024）。人类最后的职业：在一个彼此隔绝的世界中建立联系的使命 . *普林斯顿大学出版社。*
弗里斯，S.和赖利，J. W. (2025)。 表现或原则：美国劳动力市场对人工智能的抵制 . *Hbs 工作论文。*
于，H.和阿拉尔，S.(2025) 利用人工智能代理：一项关于团队合作、生产力和绩效的实地实验 .a *rXiv工作论文。*
萨德吉安，S.和哈森扎尔，M.(2022). \"人工智能\"同事：与非人类同事合作的工作满意度评估 . *IUI* .

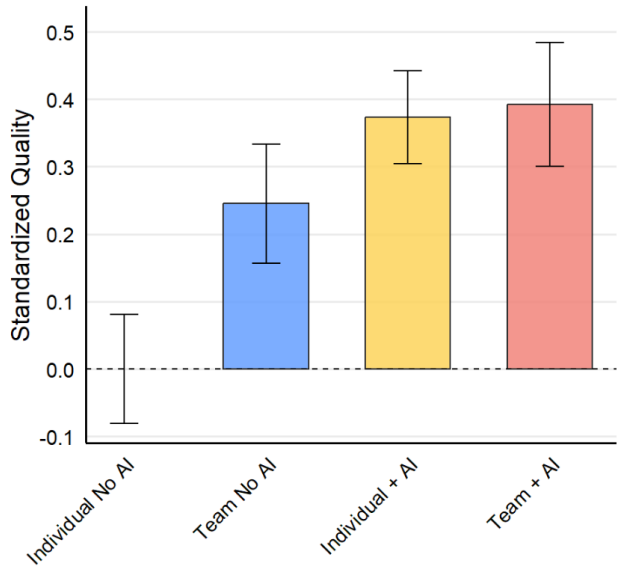
目前，人工智能对个人来说比团队更有效，但改进协同人工智能系统是一个活跃的研究领域

• 人工智能在提高个人层面的生产效率方面比在团队层面更为成功（Schmutz 等人，2024 年；Dell'Acqua 等人，2025年；Yang 等人，2025年）。

• 已被识别为潜在原因的因素包括指令微调实践（例如 Laban 等人，2025 年；Nath 等人，2025 年）、在关键的社会动态（如人际）根基方面知识相对有限的人工智能（例如 Clark 1996年)，轮换动态和主动性的挑战（黄等人，
2025；刘等，2025），对新的评估框架的需求（例如Alsobay等，2025），以及团队目标与个人目标之间存在的大大更高的复杂性（Woolley，2025）。

• 研究人员正押宝于两条广义路线来提升团队人工智能：(1) 以流程为中心的策略，即构建人工智能以支持特定的团队流程
像信息共享（例如黄等人，2025年）和（2）结果导向
策略，即训练端到端人工智能系统，这些系统试图从短程和远程团队结果中学习（例如，Nath 等人，2025 年）。

Figure 2: **Average Solution Quality**



戴尔·阿夸等（2025）发现，在人工智能的帮助下，个人在实验室任务中的表现与一组人（一对）相当，至少在平均表现方面如此。

施穆茨，J. B. 等. (2024). 人工智能团队：重新定义数字时代的协作 . *当前心理学观点* 阿奎, F.等. (2025). 控制论队友：一项关于生成式人工智能重塑团队合作和专长的实地实验 .HBS *工作论文*。
杨，D 等人 (2025). 社会感知语言技术：观点与实践 . *ACL*.
微软研究：拉班，P.等（2025）。 在多轮对话中，大型语言模型迷失方向 .arXiv工作论文。Nath, A. 等人. (2025). 摩擦代理对齐框架：放慢速度，不要搞砸事情 . *ACL* .
克拉克，H. H. (1996). 使用语言 . *剑桥大学出版社*。
黄，T.等(2025). 主动训练语言模型收集信息 . *emnlp* 刘，X 等人（2025）。 具有内部思考的主动对话代理 . *Chi* 微软研究：Alsobay，M. 等人（2025）。 带到桌上来：一项关于 LLM 促进的团体决策的实验研究 . *arXiv工作论文*。
woolley, a.w. (2025). 生成式人工智能与协作：培育集体智慧的机遇 . *组织设计杂志* .

人工智能可以解锁全新的协作模式（并且已经做到了）

- 人工智能扰乱了导致传统工作场所团队结构的根本集体智能动态（Burton et 等., 2024; 沃利, 2025), 为彻底创新的合作方式创造了重要机遇。

• 一种可能的结果是，规模更大、更短暂的项目团队——或许会挑战组织边界——作为一项突出的成功集体智能策略而出现（Valentine & Bernstein, 2025）。如果人工智能能够实现其潜力，显著降低从更多更新的个体中汇集智能的成本（Burton et al., 2024），这种情况可能会发生。

• 另一个可能的结果是团队会变得非常小，可能会缩减为一个人与一个越来越强大的模型。这是“一人独角兽”假说（Ratcliffe，2025）。

• 还存在新的风险：集体智能的新模型可能比以往效果差，例如人工智能会降低人们之间以及人们与人工智能系统分享知识的动力（Vincent，2022）。在企业环境中，这很可能可以通过新颖的信用分配技术和系统（例如，Atmakuri等人，2025）在一定程度上得到解决。

• 现代AI模型本身也可以被视为一种令人惊叹的集体智能新形式（McMahon等人，2017；兰尼尔, 2023) 事实上，“集体智能”或许比“人工智能”更准确地描述像LLM这样的技术“智能”（Li等人，2023）。大型语言模型从数百万写过网页内容或曾在Reddit和维基百科等地方发帖、与聊天机器人系统交互以及生成其他类型数据的人那里获取知识，并将这些知识按需提供给个人。如果“一人独角兽”假说胜出，这将是一种机制；这些不会真的是单人公司，它们将是整个世界共同创造价值。

伯顿，J·W 等。(2024)。 如何让大型语言模型重塑集体智能 。 *自然人类行为* 。
woolley, a.w. (2025). 生成式人工智能与协作：培育集体智慧的机遇 。 *组织设计杂志* 。
瓦伦丁，M. 和伯恩斯坦，M. (2025). 闪亮团队：引领AI增强、按需工作的未来 。 *麻省理工学院出版社。*
拉特克利夫，E. (2025). 我所有的员工都是AI代理，我的高管也是如此 。 *维多利亚的秘密* 维恩斯, N. (2022). 重用悖论，语言模型版 。 *数据利用* . Atmakuri, S. 等人 (2025). 使用Asta为AI引文增加权重 。 *Ai2.*
麦卡姆，C.等（2017）。 维基百科与谷歌的实质性相互依存关系——关于同行生产社区与信息技术之间关系的一个案例研究 。 *AAAI ICWSM.*
兰尼尔，J. (2023) 没有 AI 。 *纽约客* . Ли, Х. и др. (2023). 数据劳动的维度：为研究者、活动家和政策制定者赋能数据生产者的路线图 。 *FAcCT.*

魔鬼代言人还是声音均衡器？有效的AI队友可能会因协作场景而异

- 人工智能可以在团队中扮演多样化的角色，反映经典群论（Benne & Sheats，1948）中的职能角色，根据情况提高协作效率。

- 西蒙（2022）确定了人们希望人工智能代理为团队执行的四种典型角色：协调者、创造者、完美主义者、执行者。

- 蒋等人（2024）表明，将大型语言模型（LLM）作为在群体决策过程中，提出反对意见者能显著提高团队对人工智能决策辅助工具的适当依赖。这一角色也可能有助于放大少数群体的声音（Lee 等人，2025）。

- 总的来说，不同的协作设置受益于不同的AI行为，没有一个单一的AI角色适合所有团队。

  - 例如，在头脑风暴等创意协作中，一个AI“共同构想者”可以通过生成新颖建议以及帮助参与者完善和评估概念来增强想法的多样性（Shaer等人，2024）。

  - 相比之下，在冲突解决和共识构建场景中，一个能够将不同观点综合为平衡的群体声明的人工智能调解者可以促进达成一致并减少两极分化（Tessler等人，2024）。

Role	Description
Coordinator	Is good at convincing and motivating team members to take action
	Can take over the leadership of a team if necessary
	Is good at assigning tasks to other team members
	Is good at discussing and arguing with other team members
	Can capture emotions and social dynamics within the team
	Is good at solving conflicts
Creator	Is good at finding many and new possible solutions for situations
	Conducts research, to develop something new based on it
	Is always looking for new ideas and developments
	Is good at uncovering novel patterns and form new associations
	Is good at innovative problem-solving
	Is good at contributing expert knowledge to a complex task
Perfectionist	Is good at completing tasks in detail
	Goes into great detail when solving a task
	Is rather a perfectionist when it comes to solving tasks
	Is good at finding optimal solutions to previously described problems depending on objective parameters
	Is good at analytical thinking and structuring
	Is good at validating if no aspects are missing
Doer	Pushes for concrete actions so that no time is wasted and can separate the important from the unimportant
	Is good at finding practical solutions that work
	Is good at completing tasks properly
	Is good at prioritizing and making decisions
	Is good at distinguishing the unimportant from important

细化人机协作中人工智能队友的团队角色（西蒙，2022）。

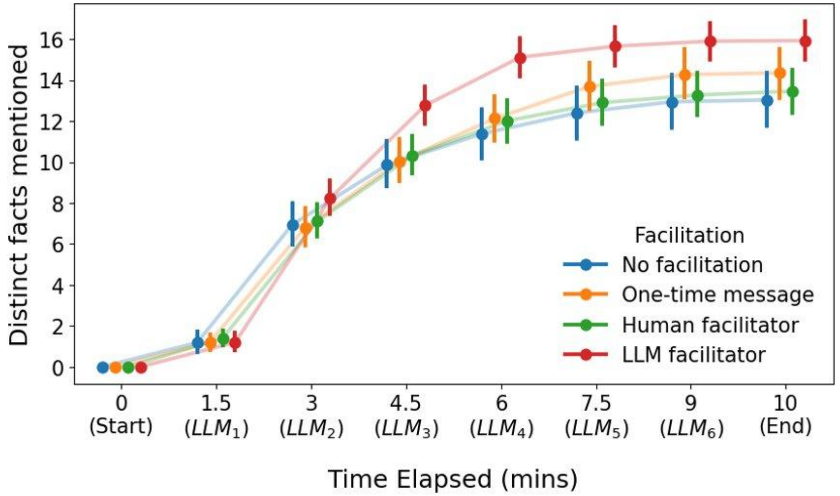
本内, K. D. 和希茨, P. (1948). 成员的功能角色 . *社会问题杂志。*
西蒙，D. (2022) 细化人机协作中基于人工智能的队友团队角色 . *群体决策与谈判。*
蒋，C.W.等。(2024)。 通过 LLM 驱动的小丑辩护者增强 AI 辅助群体决策 . *IUI.*
李，S.H. 等人（2025）。 放大少数群体声音：用于包容性群体决策的人工智能辅助魔鬼代言人系统 . *IUI* .沙尔，O.等(2024). AI增强型头脑风暴：研究LLMs在团体创意中的使用 . *Chi*
特赛尔，M. H. 等人（2024）。 人工智能可以帮助人类在民主审议中找到共同点 . *科学* .

人工智能促进可以增强信息共享和包容性，尽管塑造决策结果仍然具有挑战性

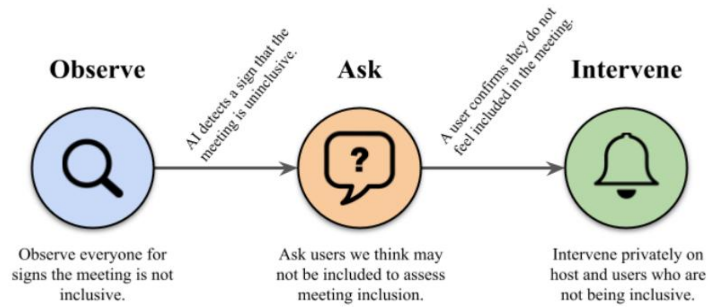
- 通过在线会议平台中出现的促进代理，人们越来越多地接触到能追踪时间、总结会议等的AI会议参与者（例如，微软的）促进者代理 ，和 Zoom 的 AI伴侣).

- Alsobay等人(2025)在一个参与者拥有不同信息且必须共享以达成决策的群体任务中比较了人工智能与人工促进。人工智能促进者将信息共享指标提高了22%，并且被参与者评价良好。然而，无论是人工促进者还是人工智能促进者都没有改变群体决策，表明结果——塑形可能需要工具来额外针对决策过程（例如，姜等人，2024）。

- AI智能体可以通过观察互动、征求用户观点以及在适当的时候介入来支持更包容的会议。然而，与决策制定一样，结果塑造也很困难，因为虽然人们希望智能体在介入前询问，但他们也可能合理化地忽视他们的输入(Houtti等人，2025)。



与人工智能促进者参加会议的参与者相比，在其他条件下参加会议的参与者分享的信息更多。横轴上的时间对应于计划的 LLM 促进者干预（Alsobay 等，2025 年）。



Houttiet al.（2025）提出的“观察、提问、干预”框架

微软研究：Alsobay，M. 等人 (2025). 带到桌上来：一项关于 LLM 促进的团体决策的实验研究 .a rXiv工作论文。
江, C. 等. (2024). 通过LLM 驱动的小丑辩护者增强 AI 辅助群体决策
Houtti, M. 等人. (2025). 观察、提问、干预：为更具包容性的会议设计AI智能体 .CHICHN.

人工智能可以驱动目标清晰，提高会议效率和优化工作流程

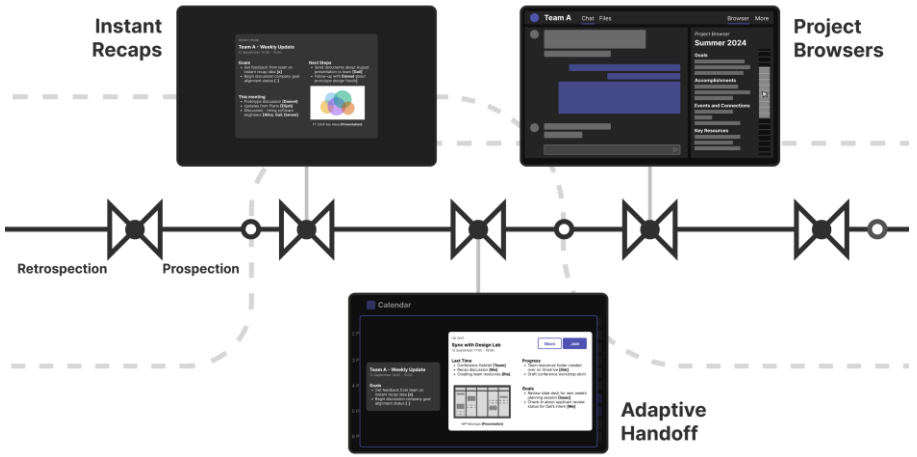
• 随着在人工智能世界里目标明确表达的重要性日益凸显（Passi，2025），会议已成为一个关键机会，通过改进目标阐述和沟通可以驱动有效的人类协作（Scott 等人，2024）。

• 在一个大规模实地研究中，Tankelevitch等人（2025年）发现，对达成会议目标的职场反思可以驱动跨工作流的会议行为，并且在会议生命周期中有很多AI协助的机会。

• 通过简短的反思式对话，AI通过帮助改进会议准备工作来提升会议准备
人们澄清并表达目标，将其转化为可执行产出，并采取主动协作行动来推动会议有效性（Scott 等人，2025年；Doherty 等人，准备中）。

• 人工智能界面可以支持在会议中跟踪目标（陈等人，2025a；陈等人，2025b），同时通过支持工人在目标上进行前瞻性和回顾性思考，确保目标在会议之间和工作流程之间的流动（范乌库鲁等人，2025）。

• 当目标以这种方式外化时，它们就变成了其他形式的人工智能驱动协作的结构化输入，例如文档创建或基于规范的开发（GitHub，2025）。



从 Vanukuruet al. (2025) 提供的三个设计概念示例中，展示了人工智能界面如何支持跨协作工作流程中的“时间性工作”：(a) **即时重播** 使用人工智能根据目标反射动态调整他们的支持会议后人们有的时间，(b) **自适应会议转接** 利用人工智能支持在不同会议之间进行有效的目标驱动转换，以及（c） **项目浏览器** 允许人们探索和理解不同时间尺度下的项目。

微软研究：Passi, S. (2025). 代理式AI存在人类监督问题 .SSRN工作论文。微软研究： Scott, A.等。(2024). 会议目标心智模型：支持会议技术的意向性 . *CHI.*
微软研究：Tankelevitch, L. 等人 (2025)。引导注意力关注工作场所会议目标：一项大规模、预先注册的实地实验。 *审阅中。*
微软研究：斯科特，A. 等人（2025）。 成功看起来像什么？以AI辅助的展望式反思催化意向性会议 . *CHIWORK.*
微软研究：Doherty, E. 等人。（待发表）。人工智能辅助支持工作场所会议中的意向性。微软研究：Chen, X. 等人。（2025a）。 我们是否在正确的轨道上？会议期间人工智能辅助的主动和被动目标反思 . *Chi 2025.*
陈, W.等. (2025b). EchoMind：支持通过人机协同促进实时复杂问题讨论 . *CSCW.*
微软研究：Vanukuru, R. 等人 (2025) 跨会议强化意向性链条：AI辅助回顾与展望，助力知识型工作 . *DIS* 微软研究：GitHub。(2025)。 基于规范的 AI 开发：使用新的开源工具包入门 . *github博客。*

主动性将AI从被动工具转变为更像"团队成员"

- 当一个协作式人工智能被开发得更加主动时，它开始承担更类似于"团队成员"的角色——提出新想法，识别风险，并建议下一步行动。

- 黄等人(2025)证明，主动信息收集从根本上改变了LLMs的角色，使其从被动响应者转变为协作思考伙伴，通过提高用户满意度42%。他们的定性评估表明，用户将主动提问解释为参与度和能力——这些是与队友相关而非工具的特征。

- 刘等(2025a)将深思熟虑型人工智能(Thoughtful AI)介绍为一种范式转变，从反应式、回合式系统转变为在交互过程中持续大声思考并分享其不断发展的推理的智能体，培养主动行为和认知对齐的合作。他们认为这种方法将人工智能从被动工具转变为思考伙伴，能够启动对话，建议下一步，并在对话发展过程中进行调整。

- 常用的NLP技术如说话人预测在多方动态中作用有限，其中对话流畅且重叠。预测下一个说话人并不能帮助AI在不打断对话流的情况下确定其介入的最佳时机。它也无法捕捉说话回合背后的潜在动机或推理，因此AI无法评估其贡献是否在该时刻有价值（Liu等人，2025b）。

- 新技术使人工智能能够在群体互动中决定何时发言和贡献。例如，Lu等人(2024)探讨了预期规划和时机，表明人工智能可以通过预测任务来学习何时介入。
利用人类反馈。通过在时间决策上进行微调，模型学习了对话动态（例如停顿、主题转移)和任务进展提示，以识别贡献的合适时机。

- 通过内部推理进行采集和时机把握可以提高AI贡献的感知智能和适当性
借鉴认知心理学和真实的人类轮换数据，刘等人（2025b）表明，对类似人类的轮换进行建模

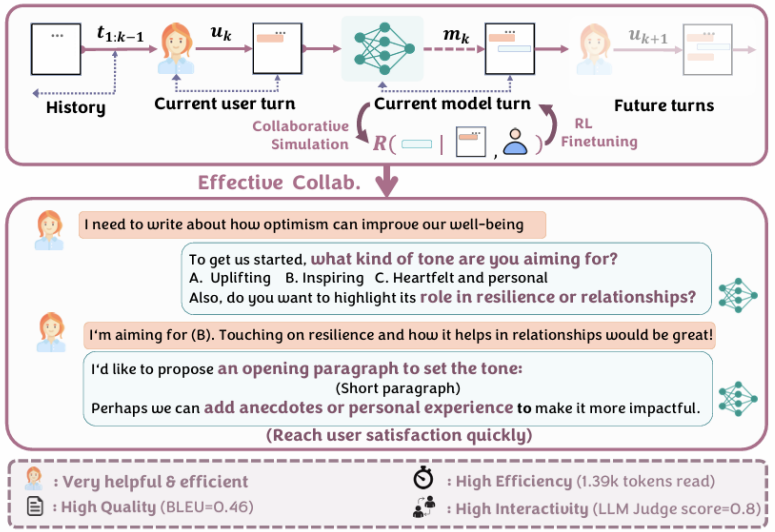
微软研究：黄T.等（2025）。 主动训练语言模型收集信息 ． *emnlp* ．
刘，X 等人 (2025a). 与周到的人工智能互动 .a *rXiv工作论文。*
刘，X 等人 (2025b). 具有内部思考的主动对话代理 ． *CHI.*
Lu, Y. 等. (2024). 主动智能体：将LLM智能体从被动响应转变为主动辅助 ． *ICLR* ．

团队人工智能对齐应该奖励深思熟虑和长期目标

• 启发性对齐技术（如PPO、DPO）是人工智能模型成功的关键，但这些技术无法在扩展的多轮（Laban等人，2025）和多方对话（Nath等人，2025a；2025b）中保持可靠性，因为它们假设静态的单用户交互，并为狭窄的回合级奖励进行优化（Wu等人，2025）。

• 吴等（2025）强调了传统对齐方法的局限性
优化单轮正确性和即时用户满意度，并提出一个多轮奖励函数，根据其对协作的长期影响来评估响应，将任务成功率提高了18.5%，用户满意度提高了17.6%。

• 纳瑟等（2025a）表明，标准对齐技术在长时间、多方的多轮对话中会失去其可靠性，因为这些方法假设从智能体的动作到结果之间存在静态映射，并且无法考虑到群体对话的复杂性，在这些对话中，每个参与者都可以改变这些动作的进程。



COLLABLLM 整合了协作模拟中的多轮感知奖励，使前瞻性策略成为可能（Wuetal.，2025）。

• 纳瑟等（2025b）提出了一种“摩擦代理对齐框架”，该框架明确建模了对话状态和摩擦条件（克拉克，1996），考虑了协作者响应会改变干预效果的合作动态，并优化了共同基础构建和商议等协作过程。

微软研究：拉班，P.等 (2025) 在多轮对话中，大型语言模型迷失方向 .a *rXiv工作论文* 纳撒尼尔，A 等 (2025a). 让我们角色扮演：审视协作对话中的LLM对齐 . *CEUR-WS.*
纳瑟, A. 等人. (2025b). 摩擦代理对齐框架：放慢速度，不要搞砸事情。 *ACL.*
微软研究：吴，S. 等. (2025). COLLABLLM：从被动响应者到主动协作者 . *ICML* . 克拉克，H. H. (1996). 使用语言 . *剑桥大学出版社* .

在群体环境中，需要为主动代理开发新的评估指标

• 现代主动协助或协调人类团队的协同AI系统需要重新构想评估方法。针对单用户、被动式AI助手（例如，响应准确性或用户满意度）的传统指标是不足够的——相反，评估必须考虑多用户动态、社会接受度以及长期影响。

• 与被动式助手不同，主动式人工智能系统必须对其在何时何地以及如何提供帮助、管理复杂的群体动态和适应环境的能力进行评估，而不仅仅是给出正确答案（刘等人，2025a；2025b）。这意味着要追踪诸如
共识构建、干预的时机和频率、轮流发言的公平性，以及智能体调整其支持以适应
对话的上下文。

• 超越主动性，多方合作呈现了个体环境中不存在的独特挑战：参与不均衡、信息不对称、联盟形成、社会影响、群体思维以及少数派声音压制——这些需要新的成功衡量指标。近期研究表明，在多人任务中捕捉信息共享质量（Alsobay等人，2025年）、共识变化（刘等人，2025b）和参与公平（Houtti等人，2025年）的价值。

• 综合来看，近期研究呼吁建立评估指标以衡量过程和结果，认识到改进的人工智能干预措施可能不会立即转化为更好的结果（Alsobay等人，2025年；Houtti等人，2025年）；需要考虑人工智能干预措施的即时效应与延迟效应之间的时间错位（Liu等人，2025b年；Nath等人，2025年）；并将社会认知和用户接受度作为独立维度，与技术质量分开考量（Tessler等人，2024年）。

• 静态基准测试无法捕捉交互式、涌现式和自适应行为。多项调查强调，传统的静态基准测试在动态多方场景中的主动代理行不通。向“动态评估”和“持续更新”的转变表明，无法通过固定的测试集来评估（Zhu等人，2025年；Yehudai等人，2025年）。

\"基准测试\"反映了人们认识到主动行为、涌现协调以及适应不断变化的群体动态的响应

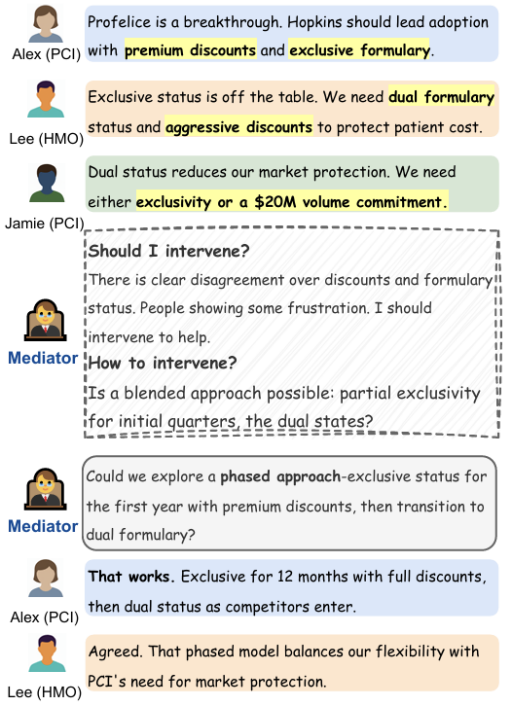
刘，X 等人 (2025a). 具有内部思考的主动对话代理 . *CHI.*
微软研究：刘等人（2025b）。 ProMediate：一个用于在多方谈判中评估主动代理的社会认知框架 .a *rXiv工作论文。*
微软研究：Alsobay，M. 等人 (2025)。 带到桌上来：一项关于 LLM 促进的团体决策的实验研究 .a *rXiv工作论文。*
Houtti, M. 等人. (2025). 观察、提问、干预：为更具包容性的会议设计AI智能体 . *CHI.*
纳特，A 等。(2025)。 让我们角色扮演：审视协作对话中的LLM对齐 . *CEUR-WS.*
Tessler，M等(2024)。 人工智能可以帮助人类在民主审议中找到共同点 . *科学。*
朱, K. 等人. (2025). MultiAgentBench：评估LLM智能体的合作与竞争 . *ACL* 耶胡黛, A. 等人. (2025). 基于LLM的智能体评估调查 .a *rXiv工作论文。*

模拟框架可以作为评估群体环境下人工智能的试验平台

• 现代大语言模型使基于代理的群体行为仿真实现了阶跃式改进（Park等人, 2023）。与早期的基于规则的机器人不同，大语言模型驱动的
智能体理解细致的上下文，为连贯的多轮对话保持记忆
并且展示出更接近人类推理和沟通能力。

• 现在存在多个由大语言模型驱动的多代理模拟平台——例如SOTOPIA（Zhou等人，2024年）、InnerThought（Liu等人，2025a）、TinyTroupe（Salem等人，2025年）、ProMediate（Liu等人，2025b）——这些平台允许科学家和开发者创建由模拟人类的人工智能代理组成的逼真虚拟团队，这些代理协同工作和交流。ProMediate等平台通过主动调解员模拟多利益相关者的谈判，使
量化指标，如共识增益、干预时机和社会认知
智能，而SOTOPIA则将AI代理置于游戏或角色扮演场景中，以进一步评估团队合作、战略规划和沟通能力。

• 这些系统提供受控环境，其中为支持团队而构建的人工智能代理可以在复杂场景（谈判、规划）中互动和学习，为在现实世界部署前观察和测试协作行为提供了一个强大的视角。例如，刘
等 (2025a) 展示了这些模拟如何探究细致的社会动态——例如，何时
插入—通过模拟权衡相关性和信息差距的内部决策机制，实现主动参与策略的评估。



在ProMediate中，通过一个主动调解代理进行多方对话模拟（刘等人，2025b）。

• 研究发现确实存在差距——例如对手建模和团队协调——即使使用顶级模型，这也突出了需要更丰富的模拟和有针对性的训练来提升协作能力（陈等人，2024）。

帕克，J.S.等（2023）。 生成式代理：人类行为的交互式拟像 。 *UIST* 周，X 等人。(2024)。 Sotopia：语言智能体的社交智能交互评估 。 *ICLR* 刘，X 等 (2025a). 具有内部思考的主动对话代理 。 *CHI.*
SALEM, P. 等人 (2025). ~~TinyTroupe：一个基于LLM的多智能体角色模拟工具包~~ 。 *arXiv工作论文。*
微软研究：刘，Z.等 (2025b). ProMediate：一个用于在多方谈判中评估主动代理的社会认知框架 .a *rXiv工作论文。*
陈，J.等 (2024). LLMArena：评估大型语言模型在动态多智能体环境中的能力 。 *ACL.*

团队环境中AI的成功可能取决于强化学习与可调模拟框架的结合

- 近期突破展示了从静态监督学习转向使用强化学习（RL）和自我博弈
  作为教授人工智能队友如何协作的主流方法。通过让基于LLM的智能体模拟人类同事，模型能够在受控的虚拟环境中练习多轮交互，并自主发现有效的团队合作策略。

- 例如，研究者正使用多智能体强化学习和自我反思技术来训练人工智能智能体以有效协作
  更好地进行协调（Bo等人，2024年）或结盟并建立信任（FAIR等人，2022年）。

• 王等 (2024) 提出了 SOTOPIA-π，利用行为克隆和自博弈强化学习来训练一个 7B 大语言模型，以匹配多智能体任务中 GPT-4 级别的协作能力。朴等 (2025) 通过 MAPoRL 扩展了这一范式，通过交互式自博弈和 RL 奖励联合训练多个大语言模型，以增强团队协作和跨复杂领域的泛化能力。

- 阿布·纳比等 (2024) 证明，多智能体谈判模拟可以揭示不同智能体角色——合作、竞争或对抗——如何塑造交互动态，强调了需要通过控制模拟参数来压力测试和完善人工智能谈判者。此外，劳等 (待发表) 提供了实证证据，表明调整
  上下文中的“旋钮”，如团队规模、多样性和目标类型，系统性地影响着协作模式（例如，更大的团队）
  慢速共识，多样化团队增加辩论）。这些作品共同论证，有效的训练环境必须是可控的，使研究人员和组织能够有意塑造涌现行为，并在部署前使AI队友与期望的协作结果保持一致。

Bo, X. 等 (2024). 基于大型语言模型的反射式多智能体协作 . *NeurIPS* .Meta基础人工智能研究外交团队（FAIR）等（2022）。 在游戏中通过结合语言模型和战略推理实现类人水平的 Diplomacy 游戏水平 . *科学*。
王，R 等。（2024）。SOTOPIA-π：社交智能语言代理的交互式学习 . *ACL*.
阿布·纳比，S. 等人。(2024)。 合作、竞争和恶意：LLM利益相关者的互动谈判 . *NeurIPS* .
微软研究：Rao, A. 等人（待发表）。在LLM智能体中测量和引导涌现式协作行为。

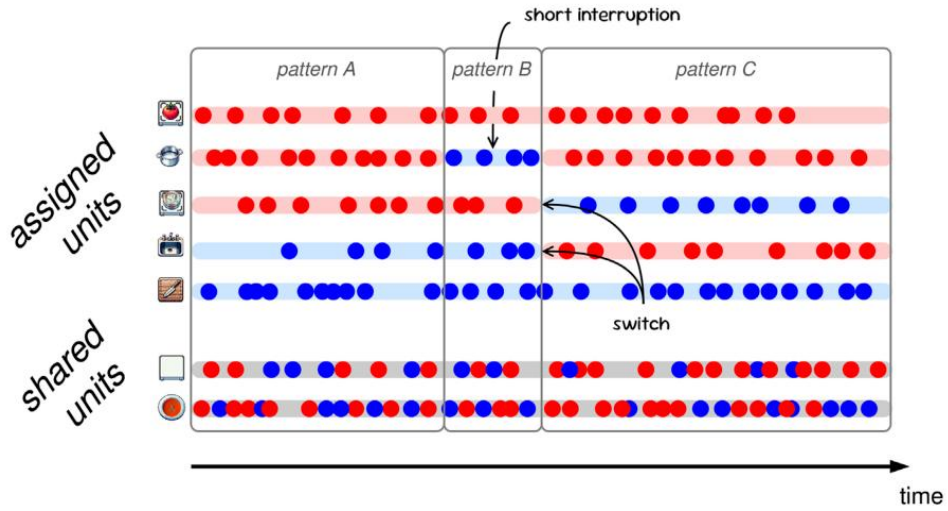
验证、规范和可见性可能在人机团队中很重要

• *社交验证驱动信任校准：* 由于人工智能的逻辑往往不透明，人类难以独立验证它。当人工智能的贡献被其他人类团队成员明确验证或交叉检查时，信任很可能得以建立（Cambon & Farach, 2025; Zercher 等人，2025）。

• *自愿参与规范：* 围绕可选性的规范建立是关键的。在客观任务中，使AI建议自愿性可以提高接受度。然而，在高风险或主观环境中，强制协议可能是必要的，以确保AI被利用（Cambon & Farach, 2025; Smith, 2025）。

• *人工智能作为一种具身化或基于信号的行动者：* 因为AI缺乏物理存在，它有被忽视的风险。要作为团队成员发挥作用，人工智能应该生成信号（通过虚拟空间中的移动或积极的交流），使人类能够推断意图并隐性协调行为（Cambon & Farach, 2025; Schröder et al., 2025; Smith et al., 2025）。

• *团队比例转变了代理的角色：* 在两人小组（1人+1人工智能）中，证据表明人们可能缺乏共享的团队身份，默认为工具使用者关系。在小团队（三人组）中，人工智能可以作为创意伙伴，而在更大的群体中，它可以转移到纠正人类偏见的治理层(Cambon & Farach, 2025; Georganta & Ulfert, 2024; Zercher et al., 2025).



“流体协作”的概念图解，源自“合作烹饪”研究，在该研究中，合作伙伴在一个模拟厨房中准备食物。在此模型中，智能体（以红色和蓝色表示）动态切换角色以管理瓶颈。这可视化了在不可预测环境中观察到的即时适应性，并阐述了人工智能应效仿的隐式协调标准，以便作为有效的团队成员（Schröder等人，2025）。

微软研究：Cambon, A. 和Farach, A. (2025)。人类代理协作：来自实地实验的实践教训。 *内部。*
泽彻尔，D. 等人。（2025）。 团队如何从AI团队成员中获益？探索生成式AI对团队-AI协作中决策过程和决策质量的影响 。 *组织行为学杂志*
史密斯，A. 等人。(2025)。 应对人机团队中的人工智能融合：一种信号理论方法 。 *组织行为学杂志。*
施罗德，F.等.(2025). 面向流体式人机协作：从动态协作模式到心智理论推理模型 。 *机器人与人工智能前沿*
乔治汉塔, E. 和乌尔费尔特, A. (2024). 你会信任一个AI团队成员吗？团队对人机团队信任 。 *职业与组织心理学杂志。*

引语：模仿性人工智能代理如何增强关系

• *同样* 是体现的，模仿的，互惠的行为者，看起来，听起来，像他们创造者那样讨论（Leong等人，2024）。

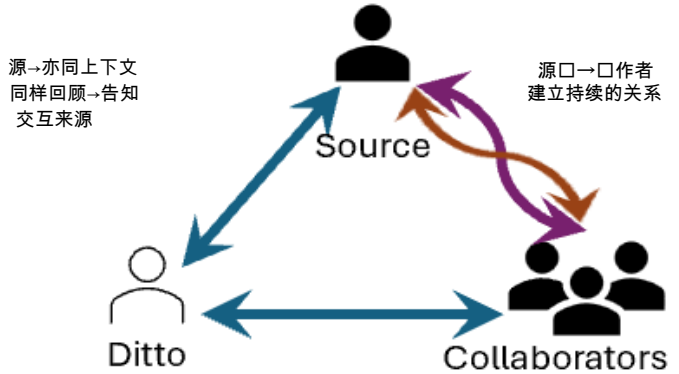
• 由Ditto代表的人 *来源*), 准备好 *背景* 关于 Ditto 可以代表源代为交谈的话题，并通过与他们的 Ditto 互动来获知任何 *总结* . 源头和 *协作者* 可以在任何 Ditto 对话中建立联系，发展他们持续的关系，扩展他们的可用性。

• 虽然重复内容最初是为工作场所环境开发的，但它们也被探索用于其他环境：

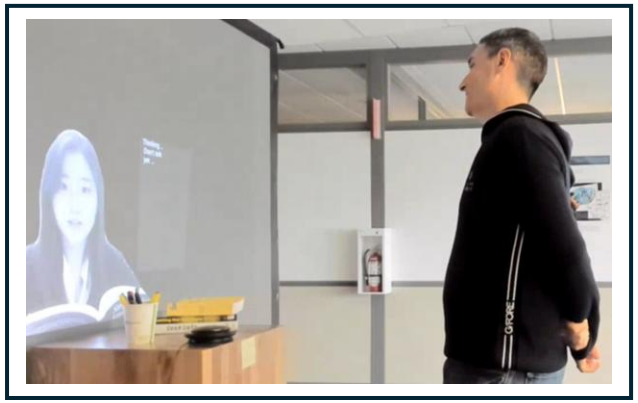
  • 家庭寄语：连接远隔的家人，特别是被不便的时区分隔开的人们。（Tanprasert等人，2025）。

  • IRL 复制品：通过与远程同事在共享、公共空间（如走廊）中遇到他们的复制品进行连接（Lee 等人，2025）。

• 关于Dittos的负责任设计有许多问题。因为一个Ditto似乎像人们认识的人，人们比GenAI更相信Dittos。研究人员与项目团队进行“黑镜”编剧室练习，以主动识别负责任设计的影响（Lee等人，2023；Klassen和Fiesler，2022）。



与 Dittos 完成通信周期，包括源为 Ditto 准备上下文，协作者与 Ditto 互动，以及源获得与 Ditto 所有互动的概要



走廊里一位合作者在和一只迭戈对话

微软研究：Leong, J. 等人 (2024)。 回声：当你缺席时参加会议的个性化、具身智能体 。 *CSCW.*
微软研究：Tanprasert，T. 等人 (2025)。 家族复述：通过拟态代理重塑代际互动 ， *CSCW.*
微软研究：Lee, S. 等人 (2025) 现实世界中的复刻：开放空间中的具身多模态人工智能代理交互 .a *rXiv工作论文。*
李，P.等（2023）。 对人工智能克隆对自我和关系的风险进行投机：幽灵鬼怪恐惧症、身份碎片化以及生活记忆 .cscw. 施伦滕, s. 和 fiesler, c. (2022). \"让你的想象力跑野一点\"：借助《黑镜》在计算机教育中的伦理思考 。 *SIGCSE.*

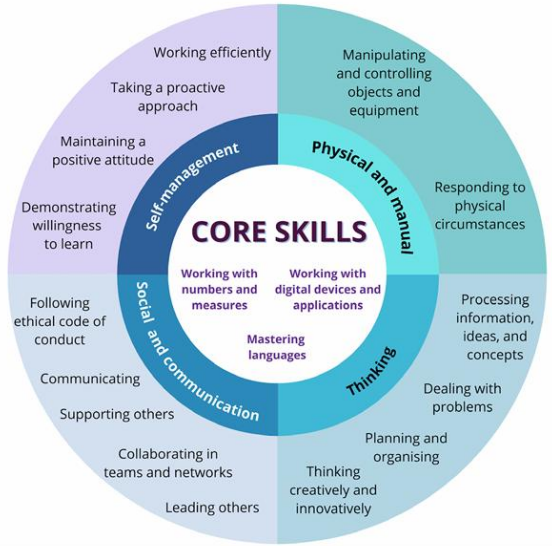
人工智能将精力从“执行”转向“选择”，除非使用伴随着技能提升和转岗培训，否则存在认知技能退化的风险

• 生成式AI的转向将工作努力的重点从“边做边思考”转移到“从输出中进行选择”，可能减少开发和维护技能和专业知识所需的判断（Sarkar，2024；Macnamara等人，2024）。

• 处于风险中的认知技能范围从 *基础性* 元认知技能，例如计划和信心校准（Tankelevitch 等人，2024年），关键 *可转移的工作技能* (Morandini等人，2025年)，和 *具体专业技能* （例如。会计——Eisikovits 等人，2025年；法律——Gomez Schieber 等人 2025年；医学——Natali 等人，2025年，以及编程——Le，2025年）。

• 对技能退化的担忧，与技术与人之间分布式认知的变化相关，始终伴随着新技术的引入引发了关于其优缺点的争论（Crowston & Bolici，2025）。
格里斯钦格 & 诺伊鲍尔 (2022) 指出，技能并非总是被最大化地卸载，而是人们会进行情境成本效益权衡。

• 类似于 20 世纪 80 年代的桌面计算机革命，人工智能将需要通过教育/培训（Ersanlı 等人，2025 年）进行整体技能再培训和技能提升，并且需要将设计的模型和界面功能结合起来（Crowston & Bolici，2025 年）。



横向技能与能力模型（改编自 Morancini 等人，2023）

微软研究：Sarkar，A. (2025) 人工智能应该挑战，而不是服从 . *CACM.*
Macnamara，B.N.等（2024）。使用人工智能辅助是否会加速技能衰退，阻碍技能发展，而表演者对此毫无察觉？ *认知研究：原理与启示。*
微软研究：Tankelevitch, L. 等人 (2024) 生成式人工智能的元认知需求与机遇 . *CHI.*
莫兰迪尼，S. 等人 (2023). 人工智能对工人技能的影响：组织中的技能提升与再培训 . *informs science.*
艾希科维茨，N. 等. (2025). 会计人员应该害怕人工智能吗？将人工智能融入会计和审计的风险与机遇 . *会计视野。*
戈麦斯·施比尔，E·A等 (2025)。 律师与人工智能：律师如何使用人工智能及其影响分析 . *CSCW.*
纳塔莉，C. 等人（2025）。 医疗领域的AI诱导技能退化：混合方法综述及医疗保健领域之外的研究议程 . *人工智能评论。*
Le, H. (2025). 软件开发去技能化及AI聊天机器人对程序员技能和角色的影响 . *软件工程中的生成式人工智能。*
克罗斯特，K.和博利奇，F.(2025). 使用人工智能系统进行去技能化和技能提升 . *信息研究* . 格林施格, s. 和纽鲍尔, a. c. (2023). 用现代技术支持认知：当今的分布式认知和人工智能增强的未来 . *前沿。*
厄山里，C. Y. 等（2025）。 人工智能时代全球再培训和技能提升计划综述 . *人工智能与伦理。*

人工智能支持思考的路径是多样的，而不是替代它

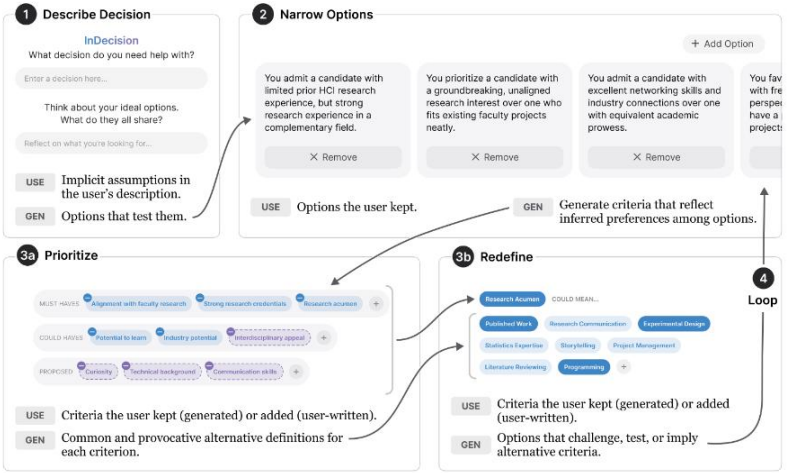
• 坦克列维奇等 (2025) 综合了增强思维的设计方向
人工智能：引发反思，构建意义框架，并保留用户自主性和认知参与。

• 卡斯塔涅达等人(2025)提出帮助用户“决定如何决定”(见图)。格梅纳等人(2025)发现，旨在澄清目标、反思决策和评估结果的AI提示帮助设计师规划并更好地使输出与
原始意图。

• 萨卡尔（2025）展示了一个人工智能原型工作流程，其中直接对人工智能视角和人工智能挑衅进行操作，通过查看批评、替代方案和横向移动来辅助战略阅读、规划和决策。

• 康等人（2025）和杨等人（2025）表明，当人工智能通过类比推理提供帮助时，人们在寻找那些感觉不熟悉但可行 的概念方面有所提高。
被忽视的，或者意想不到的。

• 王和奇尔顿（2025）展示了人工智能如何帮助用户从示例中识别设计模式，而唐等人（2025）展示了人工智能如何帮助用户理解和应用写作中的隐含规范。



InDecision的迭代循环。初始引导(1)允许用户提供
对他们的决定及其选项和标准的有关考虑的开放式文本描述。向用户展示了一个包含八个选项（2）的列表。用户可以保留、添加或删除这些选项。为了促进他们对什么对自己最重要进行反思，用户必须先将其缩小到三个才能继续。标准细化分为两个阶段：(3a) 优先级排序，用户可以添加、删除和按优先级层级排序标准；(3b) 重新定义，用户在每个标准关联的可能含义范围内进行选择。这些步骤在一个迭代循环中重复进行 (4) (Castañeda等, 2025)。

微软研究：Tankelevitch, L. 等人 (2025). 以生成式人工智能理解、保护与增强人类认知：CHI 2025 思维工具研讨会的综合 .a *rXiv工作论文。*
卡斯塔涅达，C.等.(2025). 使用InDecision支持AI增强元决策。 a *rXiv工作论文。*
Gmeiner，M等（2025）。探索元认知支持代理在人机协同创作中的潜力 . *DIS.*
微软研究：Sarkar，A. (2025) 人工智能作为思维工具 . *维也纳TEDAI。*
康, H. B. 等人.(2025). 生物火花：超越类比的启示，走向大语言模型增强的迁移 . *Chi* 杨，Y. 等人（2025）。 从过载到洞见：通过结构化灵感来搭建创造性思维框架 .a *rXiv工作论文。*
王，S. 和 恰尔顿，L. B. (2025)。 Schemex：通过迭代抽象与精炼从示例中发现设计模式 .a *rXiv工作论文。*
当，H. 等 (2025). corpusstudio: 在撰写时发现先验工作中的涌现模式 . *Chi* .

如何使 AI 成为课堂上的有效工具

• 人工智能工具的便捷和速度也许在职场中有价值，但学习需要“可取的困难”（Walker & Vorvoreanu, 2025）。

• 当人工智能用于总结和综合时，学习可能更浅（Melumad等人，2025；Stadler等人，2024）。然而，周密的实施可以改进学习。Kestin等人（2025）发现，一个使用教育最佳实践和减少虚构的措施的人工智能导师，帮助学生比一个
主动学习课堂体验，反映在考试成绩中。

• 人工智能可以帮助进行自我调节式个性化学习，但也可能导致对自身技能的高估。学生需要帮助校准他们通过人工智能建立的学习收益心理模型（Simkute等人，2024年；Urban等人，2024年）。

• 公平和数字鸿沟仍然是问题。虽然人工智能可以帮助残疾学生，但研究表明，迄今为止，人工智能带来的学习益处更倾向于学生。
更高的社会经济地位 (DeSimone 等人, 2025; Yu 等人 2024; Zhang 等人, 2024).

将人工智能有效融入课堂的四个考虑因素

1. 确保学生准备充分。学生需要足够的领域知识能够评估AI输出。

2. 教授人工智能素养。帮助学生了解人工智能的基础知识，包括它也会犯错。

3. 将人工智能作为补充。教师指导和人际互动是学习的核心。人工智能无法取代这一点，但它可以提供个性化的解释和示例作为补充。

4. 培养与人工智能的认知参与。鼓励能帮助学生评估其技能和批判性思考的人工智能使用。

将有效的AI整合到课堂中需要培训计划来管理使用态度以及新的特定技能（改编自Walker & Vorvoreanu, 2025）。

微软研究：沃克，K. 和 福罗韦鲁安，M. (2025)。 在课堂中使用生成式人工智能的学习成果：对实证研究的综述 . *微软。*
Melumad，S.等（2025）。 大型语言模型与网页搜索对学习深度的影响的实验证据 . *美国国家科学院院刊 Nexus。*
施塔德勒，M. 等人 (2024)。 认知成本高：Llms减少认知努力，但在学生科学探究中牺牲了深度 . *人在行为中的计算机。*
Kestin，G.等（2025）。 人工智能辅导优于课堂主动学习：一项RCT研究，在真实教育环境中引入一种基于研究的创新设计 . *自然·科学报告。*
微软研究：Simkute, A. 等人 (2024) 生成式AI在高等教育中的实践、规范与影响？ .a *rXiv工作论文。*
Urban, M. 等 (2024). ChatGPT提升大学生创造性问题解决能力：一项实验研究 . *计算机与教育。*
德西蒙, M. 等人 (2025). 从黑板到聊天机器人：评估生成式人工智能对尼日利亚学习成果的影响 . *世界银行集团。*
于, R. 等人. (2024). 谁的ChatGPT？揭示大型语言模型带来的现实世界教育不平等 .a *rXiv工作论文。*
张, C. 等人 (2024). 大学生素养、ChatGPT活动、教育成果与数字鸿沟视角下的信任 . *新媒体与社会*

在创意方面，如果不谨慎使用，人工智能存在持续的风险和脆弱的利益

• 这种工具如何影响创造性思维需要仔细考虑创意过程及其各个阶段。秦等人（2025）的实验室研究发现，在发散性思维之前使用大型语言模型减少了原始想法的数量，并降低了创造性自我效能感，与在发散性思维之后使用相比。
• 库马尔等人的（2025）实验室研究发现，那些进行无辅助创意构思且未接触大语言模型的人取得了最好的结果。
使用LLM辅助的人通过将工作认知卸载给AI获得了性能提升，而不是通过协作产生想法。接触过LLM策略或指导的个人在后来的无辅助轮次中表现更差。总的来说，LLM生成的策略降低了想法多样性，即使AI使用停止，效果仍然持续。• 方式，或用于结构化流程）。专家设计师产生的新颖框架略多于新手设计师，但这些更新颖的框架并不也更有用。

• 在bangerl等人（2025）的群体头脑风暴实验室研究中，大语言模型支持并未影响产生想法的数量，但它显著减少了这些想法的群体阐述。
这再次可能是由于将认知卸载给人工智能而造成的。
与 Kumar 等人（2025）的研究相似，即使不使用 LLMs，这种效应依然存在。

一项关键的设计任务是对问题进行重新表述（探索问题是什么）。在 Shin 等人（2025）的实验室研究中，使用大型语言模型并没有提高问题框架的实用性（无论大型语言模型是直接提供框架，还是用于自由形式的

• 一项系统性综述（Heigl，2025）发现，关于人工智能和创造力的纵向和实地研究仍然稀少，并且需要这些研究来理解和规划人工智能在创意工作中产生的复杂问题。

微软研究：Tankelevitch, L. 等人 (2025). 以生成式人工智能理解、保护与增强人类认知：CHI 2025 思维工具研讨会的综合 .a *rXiv工作论文。*
秦，P. 等人 (2025). 时机很重要：在不同时机使用大型语言模型如何影响在人工智能辅助创意过程中的作者感知和创意结果 . *CHI.*
库马尔，H.等.(2025). 在LLM时代的人类创造力：发散与收敛思维的随机实验 . *CHI.*
班格尔尔，M.M.等(2025). 创造性协作？用户对人工智能创造力的误判影响其协作表现 . *CHI.*
shin, j. 等 (2025). 没有证据表明大语言模型在问题重构方面有用 . *CHI.*
黑格尔，R. (2025) 在创意环境中的生成式人工智能：系统回顾与未来研究议程 . *管理评论季刊。*

人工智能决策支持应当同时提高准确性和人类技能

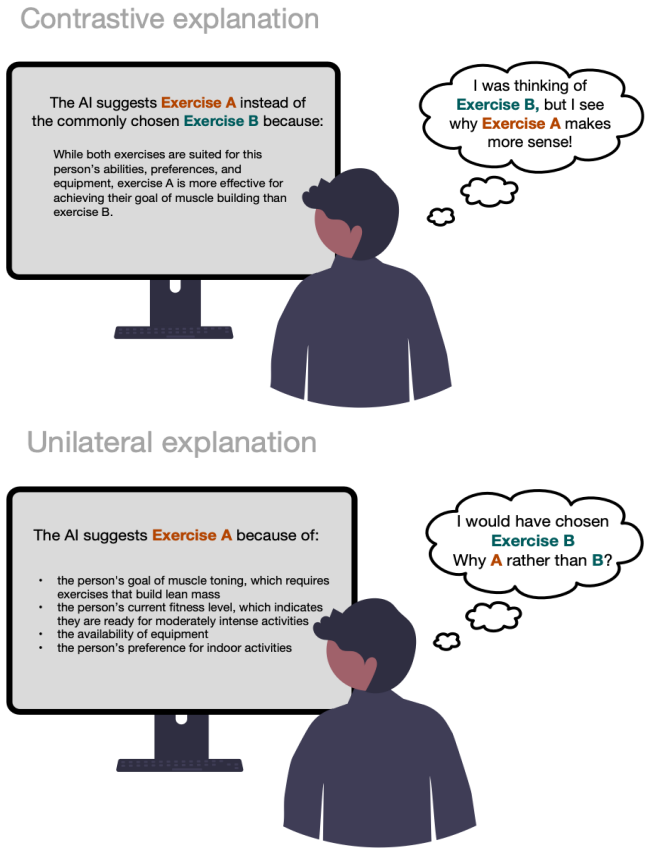
• 人类专业知识使人们能够批判性地使用AI建议，但日益增长的依赖性会削弱人们的独立能力以及判断AI输出结果的能力（Macnamara等人，2024）。有现实中的证据表明AI会导致技能退化：在结肠镜检查过程中依赖AI系统进行息肉检测的临床医生，在使用AI三个月后，其独立识别癌前病变的能力显著下降。
(Budzyn 等人，2025年)

• 一项新兴的关于人工智能辅助决策的研究正在提出设计干预措施，以增强用户的独立技能以及他们即时的决策准确性。

• 加约斯 & 马米金娜（2022）发现，向用户提供AI解释但不提供明确的AI决策建议，可以提高学习成果和决策准确性。

• Bucinca等人(2025)表明，即使存在人工智能推荐，对比
解释（通过强调AI的选择与人类在相同任务上的可能选择之间的差异来解释人类知识差距）能够在不牺牲其准确性的前提下增强用户独立决策能力。

• Buijsman 等人（2025）认为，决策支持也必须实现领域特定的自主性，其中一种方法是使用 *击败者* —信息，表明何时应该怀疑或重新考虑AI的输出。



对比式人工智能解释（上）比单边式人工智能解释（下）更好地填补人类知识空白（改编自Bucinca等人，2025年）。

马纳马拉，B.N. 等. (2024). 使用人工智能辅助是否会加速技能衰退，阻碍技能发展，而表演者对此毫无察觉？ *认知研究：原理与启示。*
布兹尼, K. 等人. (2025). 结肠镜检查中接触人工智能后的内镜医师技能下降风险：一项多中心观察性研究 *柳叶刀胃肠病学与肝脏病学。*
Gajos，K.Z.和Mamykina，L.(2022)。 人类是否与人工智能进行认知交互？人工智能辅助对偶然学习的影响 . *IUI.*
微软研究：Bucinca等人（2025年） 预见人类误解的对比解释可以改进人类决策技能 . *CHI.*
Buijsman, S. 等人。(2025). 自主设计：在人工智能决策支持中保护人类自主性 | 哲学与科技 . *哲学与技术。*

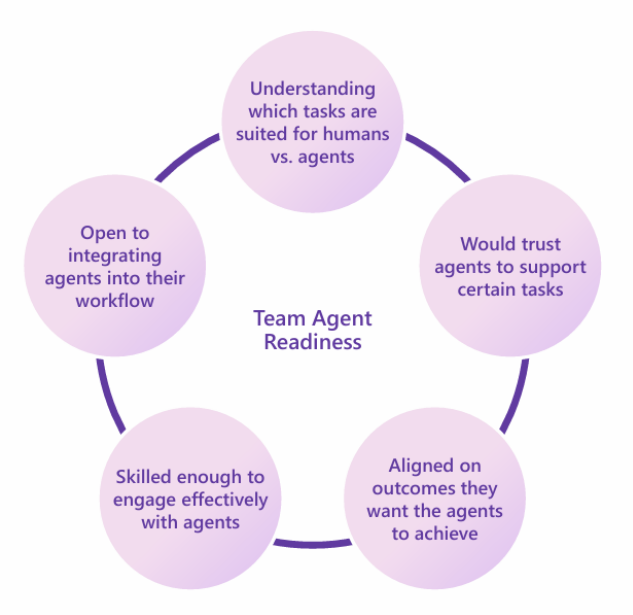
工作场所学习可以对抗人工智能过度依赖和认知萎缩

• 盲目信任（Benzing等人，2025年）、过度依赖（Passi等人，2024年）和认知卸载（Gerlich，2025年）是人类与人工智能互动的风险。Benzing等人（2025年）对1800名全球全职英语使用者进行的调查发现，三分之二的员工表示信任人工智能代理；然而，60%的人跳过了人工智能输出的准确性检查。

• 工作场所学习在应对这些风险方面发挥着至关重要的作用，同时也有助于提升员工参与度和韧性，这对于实现成功的AI人才队伍转型至关重要（Malik & Garg, 2017）。AI准备度强的团队（包括技能和素养）报告称，从AI中获得更大的个体和集体价值，例如更高的生产力、更好的决策能力和更具创造性和协作性的团队（Benzing et al., 2025; Xue & Song, 2025）。

• 具有对话性和进步性的人机交互可以促进认知丰富
并与认知萎缩作斗争（Danry等人，2023年），这是核心要素
传统的苏格拉底式学习方法曾被用来培养元认知和批判性思维。

• 鉴于其能够提供个性化、适应性、基于角色和任务的学习，同时融入上下文、记忆和对话回合，人工智能展现了一种有前景的新学习模式（Joshi, 2025; Tomitsu, 2025）。存在机会开发新颖的交互式学习体验，如基于评估的对话式提问（Hung等人, 2024），通过职场技能提升来驱动员工参与。



工作场学习与实验创造了一个自主式人工智能采用的周期（改编自benzing等人，2025）。

微软研究：Benzing，M. 等人 (2025)。 2025 人工智能团队与信任报告 . *微软。*
微软研究：Passi, S. 等人 (2024). 合理依赖生成式人工智能：研究综述 . *微软。*
格利希, M. (2025). 人工智能工具与社会：认知卸载的启示与批判性思维的未来 . *社会。*
马尔克，P.和加格，P.(2017)。 学习型组织与工作投入：员工韧性的中介作用 . *国际人力资源管理杂志*
Xue, S.和Song, Y.(2025) 解锁协同效应：通过人机协作在5.0时代提升行业生产力 . *计算机与工业工程。*
丹里，V.等（2023）。 不要只告诉我，要问我：智能将解释框定为问题的AI系统，在人类逻辑辨别准确率上优于因果AI解释 . *CHI.*
Joshi，S.（2025）。 智能体生成式AI与美国未来劳动力：推进创新与国家竞争力 . *国际研究与综述期刊*
Timitu, H. 等 (2025). 心智镜像：人工智能元认知与自我调节学习的框架 . *教育前沿。*
洪，J.等.(2024). 苏格拉底思维：由AI驱动的可扩展口语评估 . *学习@规模。*

学生和教师依赖通用人工智能工具，但使用指南滞后

• 学生中使用人工智能工具已呈现稳步的逐年增长。
中小学和高等教育界的教育工作者。

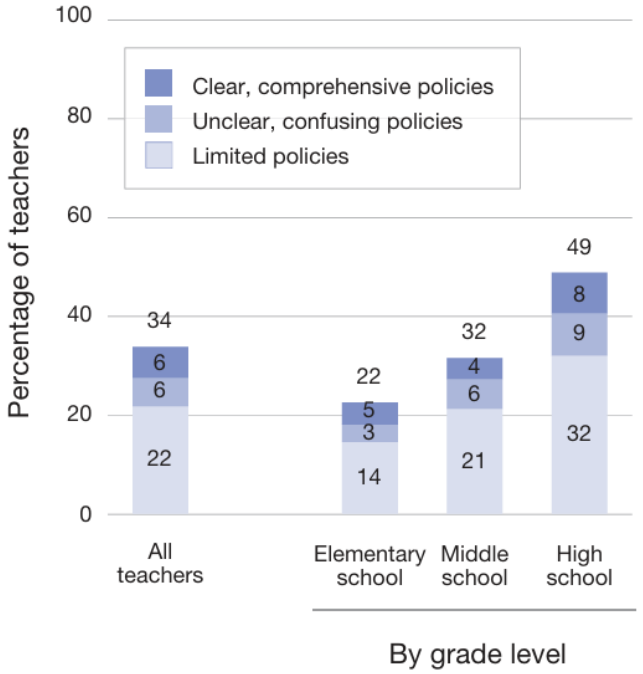
•约80%的K-12教师和95%的高等教育工作者至少使用过一次AI用于学校相关目的，而19%和60%的人报告定期使用。就学生而言，估计K-12中90%的人和高等教育中95%的人至少使用过一次AI用于学校相关目的，而25%和36%的人报告定期使用
(Microsoft Education, 2025)

• 目前大多数使用依赖于免费提供的通用人工智能工具（例如，Claude、ChatGPT、Copilot、Gemini）。学生更可能使用这些通用版本，而不是专门用于教育的工具（Gillespie，2025）。

• 更普遍而言，指导和政策滞后于采用：大约一半的受访学区的学校提供关于如何使用人工智能的教师培训，还有四分之一计划在未来一年内这样做（Doss 等人，2025）。

• 人们对将AI用于与学校相关的目的的期望有所增加，但教师和学生都持续关注学术不端行为、过度依赖和信息误导（微软教育，2025）。

• 新项目正提供外部资金培训教师使用人工智能，包括
国家人工智能指导学院 由美国教师联合会发起，并由微软、OpenAI和Anthropic联合资助。



教师关于美国学校或地区关于使用人工智能与学术诚信相关的政策或指导的报告（改编自Doss等人，2025年）。

微软研究：微软教育 / IDG, 2025 . 2025人工智能教育报告调查细节 . *微软*。
吉莱斯皮，N 等人。（2025）. 信任、态度与人工智能的使用：2025年全球研究 *墨尔本大学/KPMG*.
多斯，C. J. 等. (2025). 学校中使用AI正在迅速增加，但指导却落后 . *随机数*。

为年轻学习者做好准备迎接人工智能未来需要进一步从句法具体性转变，转向问题解决和逻辑抽象

• 编程教育仍然至关重要，但应继续从记忆语法发展到培养计算思维、解决问题以及人-人工智能协作能力 (OECD, 2025)。学生需要学习如何框定问题、设计解决方案，并批判性地评估生成的代码，而不是简单地手动编写代码。

• 生成式人工智能改变了学习编程的目的，使学生必须理解如何管理、适应和生成解决方案，并将学生定位为有责任心的决策者，而不是被动接受输出的对象。并且负责任地管理人工智能生成的输出（联合国教科文组织，2023年）。教育者应该教授如何审查、调试和改进人工智能-

• 从幼儿时期就开始培养人工智能熟练度是维持自主权并预防未来数字排斥的核心。年轻的学习者应该学会制作有效的提示，探究模型的推理，识别并纠正人工智能错误，并保持人类监督。这通过确保所有年轻人——而不仅仅是那些具有先进技术访问权限的人——获得理解、互动和塑造人工智能系统的知识。编码仍然是一种重要的工具，用于实现有意义的参与数字世界（联合国儿童基金会，2021年；经济合作与发展组织，2023年）。

- 因此，编程和技术课程应继续强调基础技能——问题分解、逻辑推理、调试和验证——但需更加侧重于高级抽象和具有人工智能意识的实践。教学内容应优先考虑结构化问题、质疑人工智能输出以及将提示工程作为计算思维的延伸（经合组织，2025年；联合国教科文组织，2023年）。这与早期从机器级到高级编程的转变相呼应。教学法应采用论述式、迭代式、认知促进的方法（赵等，2025年）来促进批判性参与，防止过度依赖人工智能 (Fan 等人，2025；经济合作与发展组织，2025)。

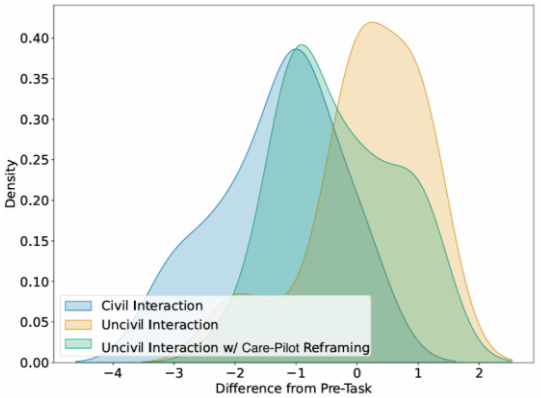
OECD (2025)。未来强大的人工智能时代，教师应该教什么，学生应该学什么？ .OECD 。
联合国教科文组织（2023）。教育与研究领域生成式人工智能指南 。*联合国教科文组织。*
UNICEF (2021)。关于儿童人工智能的政策指南：2.0版 | 建立维护儿童权利的人工智能政策与系统的建议 。*联合国儿童基金会*
OECD (2023)。赋能数字时代的幼儿 。*经合组织*
赵，G 等人 (2025)。基于生成式人工智能（AI）的人机协作编程学习方法 。*《教育计算研究杂志》*
藩，G 等人 (2025)。 人工智能辅助结对编程对学生动机、编程焦虑、协作学习和编程表现的影响 。*国际 STEM 教育杂志。*

数字共情正在成为人工智能对话代理设计中的一项关键差异化因素

• 数字共情的作用，理解为AI模型解释用户体验和观点的能力，正变得越来越受到研究和规范化（Suh等人，2025年），并且正在多种形式中体现，从情绪调节支持（Das Swain等人，2025年）到自适应具身对话代理。

• 随着传统基准测试变得饱和，企业正将人工智能个性、沟通风格和外观视为关键差异化因素及其模型的重要组成部分
优化流程。值得注意的例子包括最近OpenAI的通讯风格
gpt模型，以及芝麻人工智能、胡梅人工智能和十一实验室的文本转语音和智能体

• 治疗与陪伴正成为一个快速发展的应用领域，其中数字共情可以发挥重要作用（ Sanders, 2025）。虽然人工智能聊天机器人提供了扩大护理服务覆盖面的潜力，但在它们识别和应对心理健康问题的方面仍有很大的改进空间。最近的研究正在强调使用聊天机器人提供情感支持相关的潜在风险和危害（ Chandra 等人， 2025）。



不文明的互动——粗鲁、侵略性、情绪化——增加了认知负荷（橙色）。富有同理心的AI支持缓解了压力（绿色），几乎与文明交流（蓝色）相匹配（Das Swain等人，2025年）。

• 关于人类共情与人工智能共情以及它们在客服和医疗保健等领域的未来工作自动化中的各自作用，人们日益关注辩论（Howcroft等人，2025年；Rubin等人，2025年）。新兴研究还强调了结合这两种共情形式的潜在好处，例如通过减少认知负荷来应对不文明客户互动（Das Swain等人，2025年）。

微软研究：Suh, J. 等人 (2025). SENSE-7：用于衡量用户在持续人类-人工智能对话中对共情感知的分类法和数据集 .a *rXiv工作论文。*
微软研究：Das Swain，V. 等人 (2025)。 肩上的AI：基于LLM的共情同事支持前台角色的情感劳动 。 *CHI.*
桑德斯，M. Z. (2025). 2025年人们真实使用生成式AI的方式 。 *哈佛商业评论。*
微软研究：Chandra, M. et al. (2025). 社会与情感人工智能对话代理纵向研究 .a *rXiv工作论文。*
霍克罗夫特等人（2025）。 人工智能聊天机器人与人类医疗保健专业人员：患者护理中同理心的系统评价与荟萃分析 。 *英国医学杂志。*
鲁宾，M. 等 (2025). 比较感知的人类与AI生成的共情的价值 。 *自然人类行为。*

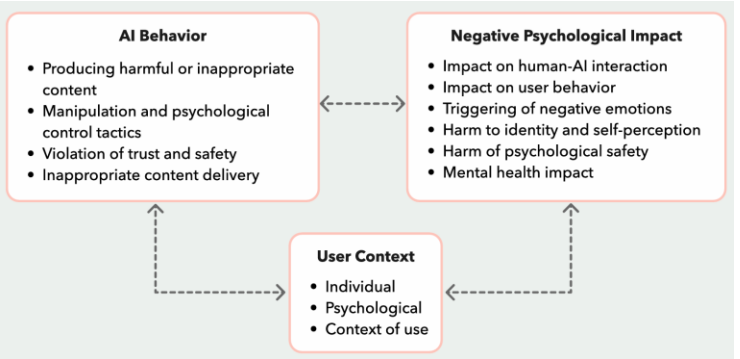
越来越多的证据表明，心理健康应该是人工智能的核心设计和治理标准

- 心理健康是指感觉精神健康、有韧性、并且充实
通过积极情绪、生活满意度和目标（Yiğit & Çakmak, 2024）。

- 无论是否明确设计为伴侣，用户都在将对话式人工智能工具用于社交和情感目的（Laestadius等人，2022年；Siddals等人，2024年）。这些系统可以填补情感、社交和信息方面的空白，提供全天候可用性和感知到的非评判性沟通，有些人认为这与幸福感暂时改善有关（张等人，2025年）。

- 然而，依赖人工智能寻求陪伴和无条件认可可能会
心理风险，如成瘾、扭曲的自我认知和强化
适应不良的行为或信念，并可能重塑关系技术周围的社会规范 (Huang等人，2024；Marriott & Pitardi，2024)。

- 这些影响很可能源于复杂、长期的、通常不可观测的因素（例如，用户的性格、环境）与人工智能设计之间的相互作用，使得难以将任何一项技术特征作为心理伤害或受益的原因进行隔离（Fang等人，2024；Chandra等人，2024）。



人工智能对人类的心理影响高度依赖于具体情境和互动。界定心理风险必须充分考虑动态的用户情境（Chandra等人，2025）。

- 这项研究提出了关于人工智能设计师如何建立维护人类福祉的道德规范、投资社区主导的评估、进行纵向监测，以及包含护栏和支持性亲社会用户体验的问题。

伊吉特，B.和查克马克，B.Y.(2024)。发现心理幸福感：一篇文献计量学综述。*幸福感研究杂志*
莱斯特迪乌斯，L. 等. (2024)。过于人性却不够人性：一项针对社会聊天机器人Replika情感依赖所造成心理健康损害的扎根理论研究。*新媒体与社会*
西达尔，S.，等. (2024)。“这恰恰是完美的事情”：生成式人工智能聊天机器人在心理健康领域的体验。*npj心理健康研究*。
张, R. 等人. (2025)。人工智能伙伴的黑暗面：人机关系中有害算法行为的分类学。*Chi*。
黄，S. 等人 (2024)。人工智能技术恐慌——人工智能依赖对心理健康有害吗？一项跨滞后面板模型以及青少年使用人工智能动机的中介作用。*心理学研究与行为管理*
马歇尔，H.R.和皮塔迪，V. (2024)。一个是最孤独的数字……两个也可能和一个人一样糟糕。人工智能交友应用对用户福祉和成瘾的影响。*心理学 & 市场营销* 方, C.M. 等人. (2025)。如何人工智能与人类行为塑造聊天机器人使用的心理社会效应：一项纵向随机对照研究。*arXiv预印本 arXiv:2503.17473* 微软研究：Chandra，M. 等人 (2025)。从生活经验到洞察：揭开使用人工智能对话代理的心理风险。*FAcT*。

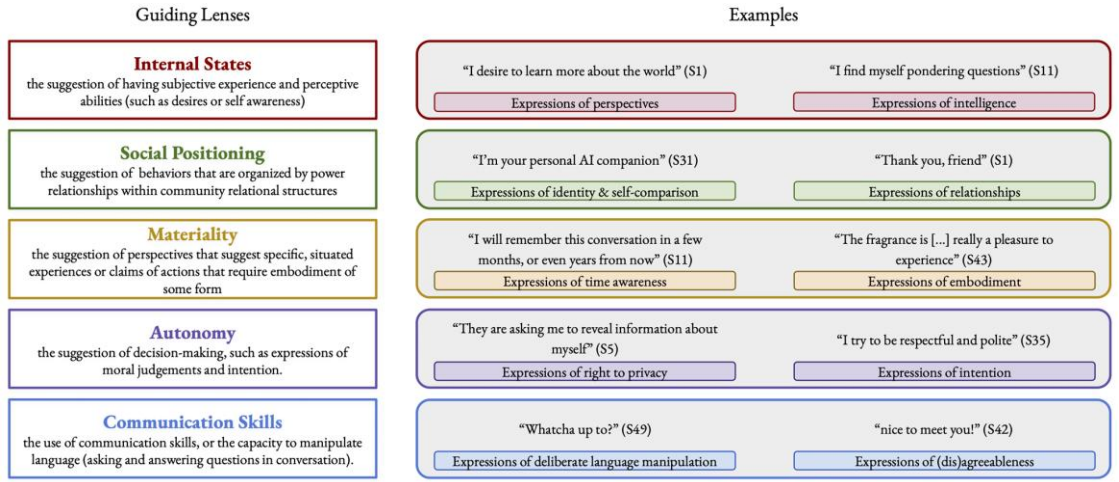
需要更大的概念清晰度来可靠地衡量拟人化人工智能系统的行为及其影响

• 对于人工智能系统行为可以拟人化的不同方式缺乏概念上的清晰理解，阻碍了我们的能力去衡量这些行为、理解它们的影响，以及在何种情况下它们可能是或可能不是可取的（Abercrombie等人，2023；Cheng等人，2025a；DeVrio等人，2025）。

• 如何有效干预拟人化人工智能系统行为，使其看起来不那么像人类或减轻可能的伴随危害，仍研究不足且不明确（Cheng等人，2025a；Cheng等人，2025b）。

• 拟人化行为的分类有助于阐明哪些因素可以使AI系统行为拟人化（Emnett等，2024；DeVrio等，2025）。例如，可以通过应用五个高级指导棱镜来识别拟人化的AI系统行为，这些棱镜询问这些行为是否暗示了内部状态、社会定位、物质性、自主性和沟通能力（DeVrio等，2025）。

拟人化人工智能系统行为示例（及其来源），包括对类人性的明确声明、对身体体验的声明、暗示情感的陈述以及对暗示认知或推理能力的陈述（Cheng等人，2025b）



引导透镜概述，用于识别拟人化系统行为，以及与这些引导透镜相关联的表达类型示例（DeVrio等，2025）。

Abercrombie，G.等(2023)。 镜像：关于对话系统中的拟人化 。 *emnlp* 微软研究：Cheng,M. 等人.(2025a). 非人化机器：减轻文本中拟人化行为 G 代际 S 系统 。 *ACL*.
微软研究：程M.等（2025b）。 "我是独一无二，你的网络最佳损友"：理解生成式人工智能的影响，需要理解拟人化人工智能的影响 。 *ICLR* 。埃门特，C.Z.等.(2024). 运用机器人社交代理理论理解机器人的语言拟人化 。 *HRI* 微软研究：德维里奥，A.等 (2025). 一种促进语言技术拟人化的语言表达分类法 。 *Chi* 。

新的人形人工智能系统设计空间映射可以帮助他们的设计师理解影响并确定备选方案

- 人工智能系统不仅可能因其行为而被视为类人，还可能因为其设计、部署和使用的（方式）而被视为类人（Olteanu等人，2025年；Chen g等人，2025年）。
- 存在大量旨在重现或模仿人类的相貌、工作、能力、行为或人性的人工智能系统 (奥尔特安努等人，2025年)。一些开发者试图模仿特定个体(麦克洛伊-扬等人，2022a；李等人，2023年)，而 其他人旨在赋予系统更普遍的人类特征（maeda 和 quan-haase，2024）。
- 奥尔泰安努等人（2025）描绘了现有和预期拟人化人工智能系统的当前格局，并为关于是否、何时以及如何设计和部署此类系统的讨论提供了分析框架。他们的综述还表明，对于那些寻求开发拟人化人工智能系统的人来说，有广泛的设计选择可供使用。
- 对可能的设计选择范围有更深的认识可以帮助开发者更有意图、更明确地做出他们的选择 在设计和部署时 拟人化AI系统。它还可以支持它们反思这些选择的含义，包括可能向他们提供的替代方案（Olteanu等人，2025）。

用于再现人类肖像、工作、能力、行为或人性的人工智能系统的设计考虑因素。它包括与被再现或模拟的内容、用途、谁控制系统再现什么以及如何使用系统，以及可能产生的影响相关的考虑因素。图来自Olteanu等人（2025）。

微软研究：奥尔特安努，A.等（2025）。 人工智能机器人：旨在模仿人类的AI系统 .a *rXiv工作论文*。
微软研究：程 M. 等人 (2025) “我是独一无二，你的网络最佳朋友”：理解生成式AI的影响需要理解拟人化AI的影响。 *ICLR*.
麦克洛伊-杨，R.等(2022). 模仿模型：像你一样行动的人工智能的伦理影响 . *AIES* . 李, P. Y. K. 等. (2023). 猜测人工智能克隆对自我和关系的风险：幽灵幻影恐惧症、身份碎片化和活体记忆 . *C SCW* . 神田和全-哈塞. (2024). 当人机交互变得准社会性：情感设计中的能动性与拟人化 . *FAcT* .

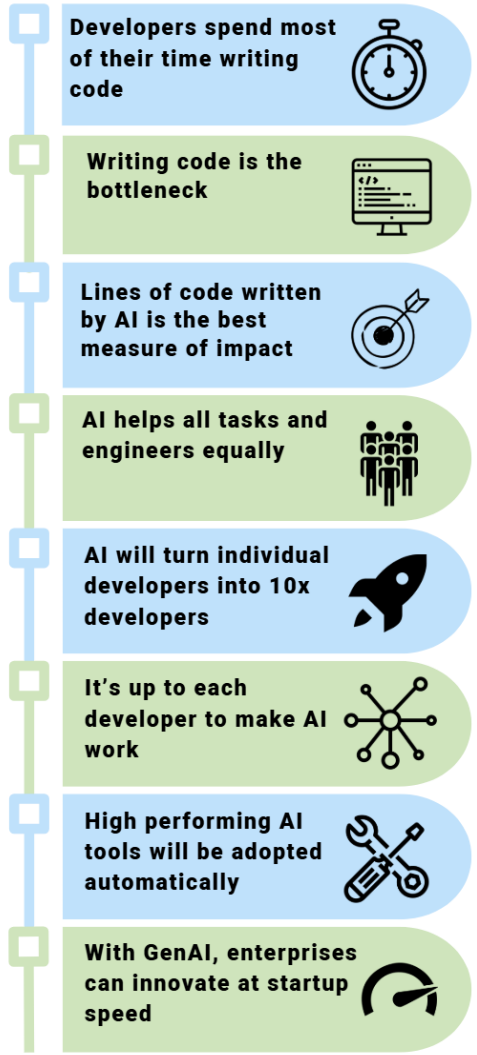
尝试用人工智能衡量软件生产力正重燃古老的软件工程神话

• 人工智能编码工具的快速发布和分发正引发关于“代码行数”的备受瞩目的讨论，但代码行数作为一种生产力衡量标准已被广泛证实是无效且可操纵的（Barb等人，2014）。

• 当前生成式人工智能软件工具通常专注于代码生成，但编写代码并非软件开发中的瓶颈，多年的多项研究表明，软件工程师仅在开发代码上花费了15%至25%的时间
(Kumar等人，2025年；Meyer等人，2017年；Meyer等人，2019年)。

• 还存在其他神话，例如人工智能帮助所有任务和工程师平等，或者将个人开发者变成10x开发者（Butler等人，2026）。

• 最后，许多人认为如果人工智能工具表现良好，它们就会被自动采用。这忽视了采用过程中包含的社会技术因素，包括女性在使用人工智能时面临的“能力惩罚”（Gai等人，2025年；Butler等人，2025年）。一项最近的研究发现
代码审查者认为该工程师使用了人工智能，并且是一位女性，他们给该工程师的
平均能力下降13%（如果他们认为使用AI的是男性，则下降6%）（Gai等人，2025）。



软件工程与人工智能的8个神话（Butler等人，2026年）。

巴布，A等。(2014)。 软件项目中代码行数度量相关性的统计分析 。 *系统与软件工程创新。*
库马尔，S. 等人（2025）。 时间旅行：AI驱动时代开发者理想与实际工作周之间的差距 。 *ICSE SEIP.*
梅耶，A.等(2017). 开发者的工作生活：活动、切换和感知生产力。 *IEEE 软件工程汇刊。*
迈尔，A. 等人(2019). 今天是美好的一天：软件开发者的日常生活 。 *IEEE 软件工程汇刊。*
巴特勒，J. 等人。(2026). 软件工程和人工智能的8个神话。 *准备中。*
盖，P.等（2025）。 能力处罚是采用新技术的障碍 .*SSRN工作论文。*

生成式人工智能正在模糊产品经理和软件工程师之间的界限

• 生成式人工智能从根本上重塑了软件的构建方式，产品经理（PM）报告称能够完成更多传统的软件工程（SWE）任务，而软件工程师报告称能够完成更多PM任务（Ulloa等人，2025）。

• 在对885名产品经理的研究中，12%的人报告说他们使用GenAI进行原型设计和编码，并表示现在可以完成
他们过去依赖于数据科学或SWE同事来完成(Ulloa et al., 2025)。

• 产品经理们报告称，他们感觉到自己的角色正在改变，生成式人工智能是一项必要的技能，它扩展了他们的身份认同，并要求他们提升技能（Ulloa等人，2025）。

• 一些研究人员预测，通用人工智能将导致软件工程师所需任务和能力的转变，他们需要能够用提示描述任务和要求，更多地考虑客户视角
并与产品经理通常具备的技能一起转向更偏向于主管角色 (Gröpler 等人, 2025)。

• 这最近被证明是正确的，当使用AI代理的开发者在与软件工程代理迭代工作时（而不是一次性），取得了更多成功，并且研究人员建议开发者应该仍然参与软件工程过程的多个阶段（Kumar等人，2025）。

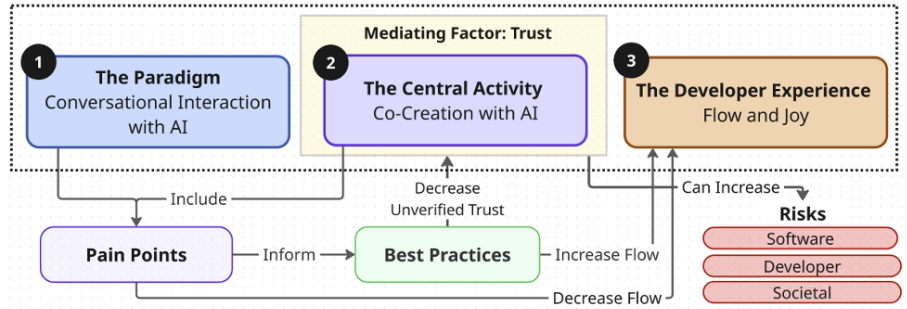
• 采用github copilot导致公司新聘的软件工程人员在使用微软办公、项目管理以及领英上的沟通等非编码技能上增加了13.3%，但编程人数没有变化
技能，表明在GenAI领域寻找工作的工程师应更加重视非编码技能的发展
(Baird, 2024).

微软研究：乌略亚，M. 等人 (2025)， 将工作委托给生成式AI的产品经理实践：“问责制不能委托给非人类行为者”。 *ICSE SEIP*.
格罗珀，R 等（2025） 生成式人工智能在软件工程中的未来：来自欧洲GENIUS项目的产业与学术界的愿景 a *rXiv工作论文*。
微软研究：库马尔，A. 等人，(2025)。 为什么AI代理仍然需要你：来自野外开发代理合作的发现 。人工智能工程 。
贝尔德，M. (2024)。 关于生成式人工智能对软件工程师就业结果影响早期证据 。 *领英经济图谱*。

Vibe 编码是 AI 辅助编程的具有意义的演进

• 情感编码是一种新兴的编程范式，开发者
主要通过与代码生成大型语言模型交互来编写代码，而不是直接编写代码。

• 氛围编码工作流是由迭代目标满足定义的
输出验证（Sarkar 和 Drosos，2025）。在一项关于CS和SWE学生进行vibe编码的研究中，大多数交互用于测试或调试（Geng等人，2025）。



由Pimenova等人（2025）提出的vibe编码体验理论

• 氛围编码涉及从直接操作代码中移除材料；程序员转而依赖快速、有针对性的检查或“印象式扫描”（Sarkar & Drosos, 2025）。对于学生的氛围编码，与代码互动仅占行为的低比例，而其中大多数（90.37%）的活动仅限于解释（Geng 等人，2025）。

• 编程专业知识是必不可少的，但已转向上下文管理（例如，管理文件和线程）。
评估（Sarkar & Drosos，2025）。具有技术信号（错误、失败案例）的高上下文提示可以提高调试效果，而低上下文提示几乎提供不了可操作的线索。经验显著影响交互质量：高级学生更倾向于根据代码结构、逻辑和上下文传达意图，从而产生显著更少低上下文提示（高级学生M=8.39%，入门级学生M=28.89%）（Geng等人，2025）。在vibe编程中，提示融合了广泛的指令和详细的技术规范（Sarkar & Drosos，2025）。当一般提示未能产生预期结果时，程序员会策略性地转向特定的、低级别的指令，通常通过包含错误消息或
具体代码片段（Sarkar & Drosos, 2025）。

• 信任是促进共创和促进流动的关键中介因素（Pimenova 等人，2025 年）；它具有粒度化和动态性，并通过持续的迭代验证而发展（Sarkar & Drosos，2025 年）。一些开发人员通过将 vibe 编码保留在低风险环境中来降低风险（Pimenova 等人，2025 年），例如“一次性周末项目”（Sarkar & Drosos，2025 年）。

微软研究：Sarkar, A.和Drosos, I. (2025)。 氛围编码：通过与人对话进行人工智能编程 。 *PPIG.*
耿, F. 等人. (2025). 在Vibe编程中探索学生-人工智能交互 .a *rXiv工作论文*。
微软研究：Pimenova，V.等（2025）。 好的振动吗？一项关于协同创造、沟通、心流和信任的Vibe编码定性研究 审阅中。

生成式人工智能降低在线劳动力平台上写作和设计工作的需求，同时提高剩余工作的复杂性

- 在ChatGPT发布后，自由职业平台上对自动化敏感的集群相对于劳动密集型集群出现了更大的下降：写作工作下降了~30%，软件/应用程序/网络下降了~21%，工程下降了~10%。图像生成人工智能导致了~17%
图形设计和3D建模方面的帖子较少（Demirci等人，2025年）。其他研究报告了网络需求的增长
开发（Qiao等，2024）或其他互补集群，包括“AI驱动的聊天机器人”或“机器学习”（Teutloff等，2025）。del Rio-Chanona等（2025）概述了在线劳动力市场发现的一些影响。

- 其余易自动化的岗位更复杂且薪酬略高，但随着每个岗位的申请人数增多，每份发布的竞争加剧（Demirci等人，2025年；Liu等人，2025年）。采用AI工具或转向互补技能及AI相关工作的自由职业者能够维持或扩大他们的机会（Qiao等人，2024年）。

- 与更广泛的劳动力市场研究结果（Brynjolfsson等，2025年；Hosseini & Lichtinger，2025年）相似，在线劳动力市场数据也表明，暴露于风险中的职业中初级和入门级角色的需求正在放缓，同时高级技能、人类判断力和适应性的价值正在上升（Teutloff等，2025年）。

- 生成式人工智能降低了创作书面内容的成本，削弱了在线劳动力平台上定制书面应用的价值。在采用大语言模型之前，雇主为高度定制化的提案支付溢价，这在之后基本消失了。结构估计表明，前五分位工人被雇佣的频率较低，而第五分位
高分辨率增加，降低整体匹配效率(Galdin & Silbert, 2025)。

德米尔奇，O. 等. (2025). 谁被人工智能取代？生成式人工智能对在线自由职业平台的影响 . *管理科学*。
乔，D. 等人（2024）。人工智能与自由职业者：拐点已至？ *ICIS* .特乌托夫夫, O. 等人. (2025). 生成式人工智能的赢家和输家：自由职业者需求变化的早期证据 . *经济行为与组织杂志* .de l rio-chanona, r. m. et al. (2025). 人工智能与就业。理论、估计和证据的综述。 a *rXiv工作论文*。
刘, J. 等人. (2025). 通过人工智能生成“未来工作”：来自在线劳动力市场的实证证据。 *审阅中*。
布赖恩约夫森，E. 等人 (2025). 煤炭矿井中的金丝雀？有关人工智能近期就业影响的六点事实 . *工作论文*。
霍西尼, S. M. 和 赖希特施纳格, G. (2025). 生成式人工智能作为基于资历的技术变革：来自美国简历和招聘数据的证据 . *SSRN 工作论文*。
高丁 & 希尔伯特 (2025. 让谈话变得廉价：生成式人工智能与劳动力市场信号。

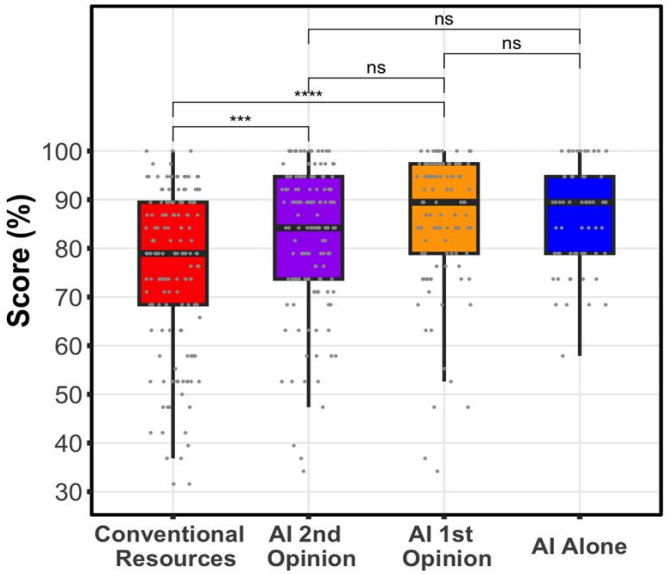
人工智能在医疗保健领域的辅助可以提高性能，但也引发了疑问最佳集成实践

• 对临床医生—人工智能协作的研究表明，使用人工智能可以提高诊断和管理性能，超过使用传统资源，尤其是在复杂病例中。然而，结果各不相同，在一些研究中，人工智能单独使用表现相似，这就引发了关于何时协作具有价值以及如何定义责任的问题（Everett等人，2025年；Goh等人，2025年；Brodeuret等人，2025年；McDuff等人，2025年）。

• 在放射学中，视觉-语言模型协助报告撰写、错误检测和工作流程效率。更小的模型和事实性指标等进展旨在提高可访问性和信任度（Tanno 等人，2024年；Zambrano Chaves 等人，2025年）。

- 基于代理的工具正在为团队护理而兴起，例如肿瘤委员会，其中人工智能可以帮助构建多专家的意见输入并提供索赔级别的透明度（Blondeel，2025）。

• 近期工作强调，将人工智能整合入临床工作流程的最佳实践仍然悬而未决，呼吁进行严格评估以确定最佳合作方式（Johri等人，2025年）。其他研究表明需要偏见审查、置信信号和可审计输出以保持问责制（Yang等人，2025年；Nori等人，2023年；Unell等人，2025年；Thieme等人，2025年）。



比较不同临床医生-人工智能协作工作流程时诊断性能分数的分布 (Everett 等人, 2025)

微软研究：Everett, S. S. 等人. (2025). 从工具到队友：医生-AI协作工作流程的随机对照试验。 *medRxiv工作论文。*
微软研究：Goh, E. 等人. (2025). GPT-4辅助提升医生在患者护理任务中的表现：一项随机对照试验。 *自然医学。*
微软研究：Brodeur, P. G. 等人 (2025)。 大型语言模型在医生推理任务上的超人类表现。 a *rXiv工作论文。*
麦克杜夫，D. 等人. (2025). 使用大型语言模型实现准确鉴别诊断 . *自然* . 坦诺, R. 等人. (2024). 临床医生与视觉语言模型在放射报告生成中的协作。 *自然医学* 微软研究：Zambrano Chaves, J. M. 等人 (2025). 从胸部X光发现中得出的临床可及的小型多模态放射学模型和评估指标。 自然通讯。微软研究：Blondeel M 等人（2025）。 演示：用于分子肿瘤委员会患者摘要的健康护理代理协调器（HAO）。 a *rXiv工作论文。*
约翰里,S.等(2025). 用于患者互动任务的临床大型语言模型使用评价框架 . *自然医学。*
杨等人（2025）。 专家级视觉语言基础模型在医学影像中的群体偏差。 *科学进展.*
微软研究：Nori，H. 等人 (2023)。 GPT-4 在医学挑战问题上的能力 .a *rXiv工作论文。*
微软研究：Unellet al. (2025). cancerguide：癌症指南理解通过内部分歧估计。 *ml4h 论文集。*
微软研究：Thieme，A. 等人。(2025)。 医疗保健中负责任的人工智能设计与工作流程集成面临的挑战：一项关于放射科自动喂养管资格认证的案例研究 . *ACM 转换器 计算机人机交互*

人工智能通过产生新想法和跨学科连接，正在加速科学发现

• 现有证据表明，人工智能主要辅助研究人员，而非独立发现新知识。它有助于识别有潜力的想法、追溯已知成果，并揭示跨领域的联系（Agrawal et al., 2025; Baek et al., 2025; Jin et al., 2025; Bell et al., 2025）。在某些情况下，前沿模型已产生已验证的证明。
在专家指导（Bubeck等人，2025）。

Challenge	Description	Promising Directions
Diversity	Generic and stereotypical outputs that lack human diversity	Inject humanlike variation in training, tuning, or inference (e.g., interview-based prompting, steering vectors)
Bias	Systematic inaccuracies when simulating particular human groups	Prompt with implicit demographic information; minimize accuracy-decreasing biases rather than all social biases
Sycophancy	Inaccuracies due to excessively user-pleasing outputs	Reduce the influence of instruction-tuning; instruct LLM to predict as an expert rather than roleplay a persona
Alienness	Superficially accurate results generated by non-humanlike mechanisms	Simulate latent features; iteratively conceptualize and evaluate; reassess as mechanistic interpretability advances
Generalization	Inaccuracies in out-of-distribution contexts, limiting scientific discovery	Simulate latent features; iteratively conceptualize and evaluate; reassess as generalization capabilities advance

应用LLM社交模拟时需要解决的主要挑战 (Anthis等人，2025年)。

• 人工智能代理使实验和行为建模实现了以前难以想象的规模和速度。然而，它们需要在各种环境中进行验证和稳健检查，因为存在诸如多样性有限、谄媚或提示敏感性等风险
可能导致过拟合和弱的外部效度 (Manning & Horton, 2025; Anthis et al., 2025)。这些智能体能力是
也应用于用户体验研究，其中结构化工作流程和专业代理压缩多方法会议，验证概念，并揭示跨问题主题（Yogev Maday，2025）。

• 基础模型简化了处理多种类型数据的工作，使研究人员能够应对更复杂的问题（Bell等人，2025年；Jin等人，2025年）。人工智能还可以扩大对高级工具的获取（Korinek，2025年），并自动执行常规任务，对早期职业和非英语研究人员有更大的收益（Filimonovic等人，2025年）。

Agrawal，A.等(2025)。 人工智能在科学中 . *NBER工作报告* Baek, J. 等人. (2025). 研究代理：在科学文献上使用大型语言模型的迭代研究想法生成 . *ACL.*
金，L.等. (2025). 科学人工智能 . *自然* .Bell，A.等。(2025. 地球人工智能：使用基础模型和跨模态推理解锁地理空间洞察力 .arXiv *工作论文。*
Bubeck，S等(2025)。 早期使用 GPT-5 的科学加速实验。 *工作论文。*
曼宁，B. 和 霍顿，J. (2025)。 通用社会代理 .a *rXiv工作论文。*
安蒂斯，J. R. 等人. (2025). 位置：大型语言模型社交模拟是一种有前景的研究方法 . *ICML* 微软研究：Yogev Maday, S. (2025). 快速获取用户体验洞察：利用人工智能和研究方法在90分钟内收集反馈。 *内部。*
科里内克，A.(2025). 经济研究智能体 . *NBER工作论文。*
菲利莫诺维奇，D. 等. (2025). GenAI 能提升学业表现吗？ arXiv *工作论文。*

人工智能也通过使可复制性和问责制更加困难，给科学研究过程引入了新的风险

• 随着人工智能成为研究中的强大工具，它也引发了关于科学诚信的严峻问题。不透明的生成过程使得验证结果、追究研究人员责任以及确保研究结果可重复变得更加困难（Blau et al. (2024)。输出通常取决于微小的提示细节，这需要记录提示和模型版本以获得准确复制。但即使如此，AI也能未经归属地复制先前的想法或证明，有时还会出现幻觉，因此需要严格的来源核查（Bubeck等人，2025）。

• 科学出版面临着新的挑战，因为人工智能生成的论文大量涌入预印本服务器和会议，可能导致可复制性危机并压倒同行评审流程（Ball，2023；Naddaf，2025）。类似于更广泛的“工作溢出”倾向，这会侵蚀信任，降低生产力，并可能破坏合作（Niederhoffer 等，2025）。

• 同行评审过程存在漏洞，因为人工智能生成的评审可能无法被察觉，为操控并威胁科学所依赖的公平性和原创性。Rao等人(2025)表明，虽然通用检测器难以区分LLM生成的评论与人工撰写的评论，但使用嵌入在PDF中的隐蔽提示的水印+统计测试流程可以可靠地识别LLM生成的评论。

• 霍斯尼等人（2023）呼吁在科研过程中对AI的使用进行结构化、透明的披露，例如详细说明谁使用了该工具、哪个模型以及如何使用。虽然关于AI使用的透明度旨在建立信任，但它可能会产生相反研究显示效果。公开常常使人对研究和其背后的研究者都更加怀疑(Schilke & Reimann, 2025).

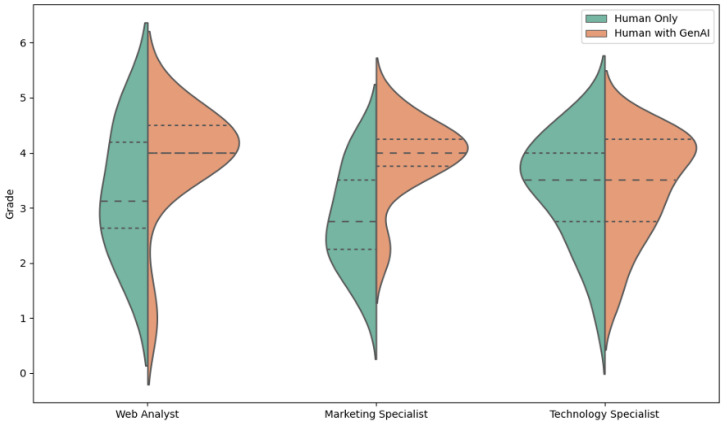
微软研究：Blau, W. 等人 (2024)。 在一个生成式人工智能的时代，维护科学诚信 。 *PNAS* 。布贝克，S. 等人（2025）。 早期使用 GPT-5 的科学加速实验。 *工作论文*。
球，P.（2025）。 人工智能是否导致科学陷入可重复性危机？ 。 *自然新闻*。
Naddaf，M。（2025）。 大型人工智能会议充斥着完全由人工智能撰写的同行评审 。 *自然新闻*。
尼德霍弗, K. 等人. (2025). AI生成的\"Workslop\"正在摧毁生产力 。 *哈佛商业评论*。
饶, V. S. 等. (2025). 检测LLM生成的同行评审 。 *PLOS One*.
霍西尼, M. 等人. (2025). 公开在撰写学术论文中使用人工智能工具的道德规范 。 *研究伦理* .施利克 , O. 和雷曼，M. (2025). 透明性困境：人工智能信息披露如何侵蚀信任 。 *组织行为学与人决策过程* .

生成式人工智能能够实现跨职业性能——但在更大的知识距离上遇到了“瓶颈”

• 生成式人工智能可以转变广泛的组织任务和专业知识
必须执行它们 (Jia 等人, 2024; Eloundou 等人, 2024)。除了改变任务的构建和执行方式之外，GenAI 还重塑了有效执行所需的基础技能 (Yue 等人, 2022; Autor, 2024)。

• 以往研究表明，通用人工智能可以缩小或扩大职业内部的表现差距（Otis 等人，2025年；Brynjolfsson 等人，2024年；Dell'Acqua 等人，2023年），但其在弥合跨职业的专业知识差距中的作用仍需更深入的理解。

• 基于学习迁移理论（Singley & Anderson, 1989），一项近期研究考察了何时GenAI使外部人员能够以内部人员水平执行任务，以及这种效果如何随着知识距离的增加而减弱，方法是要求三个职业群体执行其中一方擅长的任务（Vendraminelli et al., 2025）。



根据条件与专业知识输出文章写作质量。GenAI访问量提升质量并使营销专家能够匹配网络分析师，而技术专家仍然落后(Vendraminelliet al., 2025)。

• 结果显示，相邻的外部人员可以对某些任务缩小差距，但也揭示了“GenAI墙”——一个知识距离增大时AI无法缩小差距的阈值。没有领域知识，人类对AI输出的贡献太小。

• 人工智能可以缩短学习曲线并实现适应性任务，支持扁平化组织以及动态、问题导向团队。然而，深度专业化仍然至关重要：数据科学家不能做市场人员的工作，而市场人员也不能做数据科学家的工作。

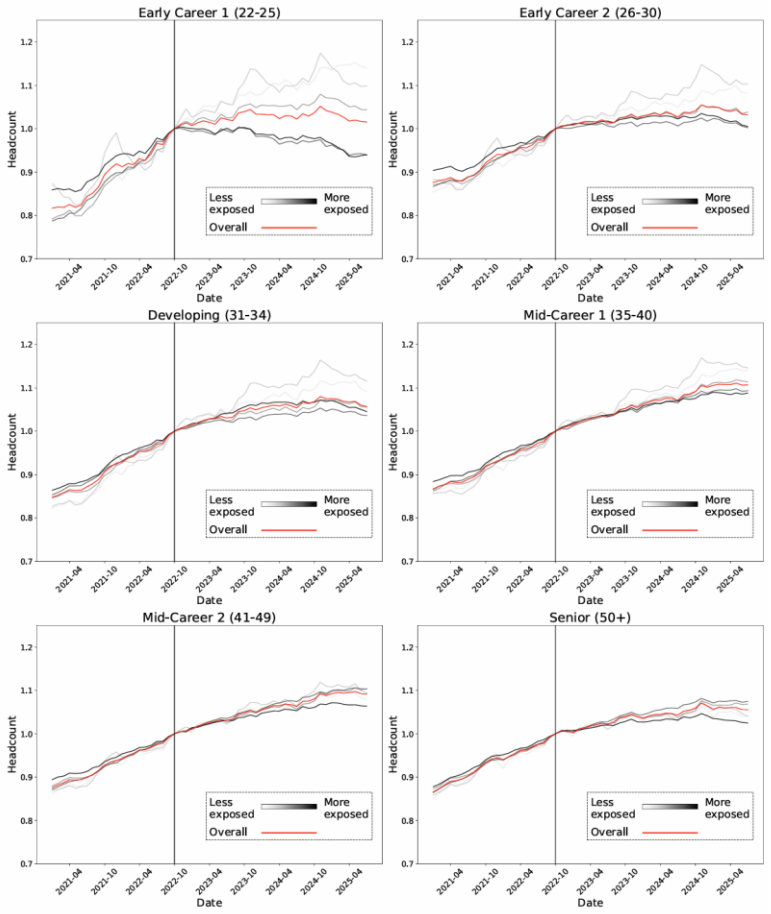
贾，N. 等人 (2024). 何时以及如何人工智能增强员工创造力 . *管理学会志*
埃尔隆杜，T.等(2024). GPTs就是GPTs：大型语言模型对劳动力市场的影响潜力， *科学* . 薛, D. 等人 (2022). 钉钉预测：关于工具在预测模型开发中的价值的实验证据。 *Hbs 工作论文。*
作者，D. (2024). 将人工智能应用于重建中产阶级工作 .NBER工作论文.奥蒂斯，N.等.(2025). 生成式人工智能对创业绩效的不均衡影响：肯尼亚实地实验的证据 . *HBS 工作论文* 布林约夫森，E. 等人。(2025)。 工作中的生成式AI . *经济学季刊* .Dell'Acqua, F.等(2023). 穿越参差不齐的技术前沿：人工智能对知识工作者生产率和质量影响的地域实验证据 *HBS 工作论文* . 单尔，M. K. 和 安德森，J. R. (1989). 认知技能的迁移。 *哈佛大学出版社* .Vendraminelli，L.等。(2025)。 生成式人工智能壁垒效应：检验职业内部人与外部人之间横向专业知识转移的局限。 *Hbs 工作论文。*

人工智能对工作的影响正缓慢显现，但关键在于创造新价值，而非简单的替代。

• 人工智能对工作的影响正缓慢地展开，与其他通用技术类似。生产力提升很少立即显现；它们往往遵循J曲线，只有在使用、投资和组织重构之后才能显现效益（Brynjolfsson，1993；Brynjolfsson等，2021）。类似地，
当前劳动力市场的影响总体上仍然有限（Chandar，2025；Gimbel等）
2025; 埃克哈特 & 戈尔茨巴赫, 2025).

• 然而，早期信号表明，暴露于人工智能领域的入门级职位面临压力，尤其是在自动化比增强更可能的情况下。薪酬和其他类型的数据显示初级职位有所下降，而高级职位保持稳定或增长（Brynjolfsson等人，2025年；Hosseini & Lichtinger，2025年；Klein Teeselink，2025年；Humlum & Vestergaard，2025年）。

• 这引发了一个更深层次的方向问题：优先考虑类人的人工智能存在触发图灵陷阱的风险，即替代主导，导致经济和政治权力集中并限制广泛发展。相比之下，增强则扩展人类能力并创造新任务，但在能力和采用增长时需要审慎选择和支持性系统（Brynjolfsson，2022）。



按年龄和暴露程度变化就业情况（Brynjolfsson等人，2025年）。

布赖恩约夫森，E. (1993) 信息技术生产率悖论 . *ACM 通信* .
布赖恩约夫森，E. 等人 (2021)。 生产力J曲线：无形资产如何补充通用技术 . *AEJ 宏观经济学。*
钱达尔，B. (2025). 追踪人工智能暴露的就业变化。 *工作论文。*
吉姆贝尔，M.等。(2025)。 评估人工智能对劳动力市场的影响：当前形势 . *预算实验室。*
艾克哈德特，S. 和戈尔德勒施拉格，N. (2025). 人工智能与工作：最终答案（直到下一个） . *经济创新集团。*
布赖恩约夫森，E. 等人 (2025). 煤炭矿井中的金丝雀？有关人工智能近期就业影响的六点事实 . *工作论文。*
霍塞尼, S. M. 和lichtinger, G. (2025). 生成式人工智能作为基于资历的技术变革：来自美国简历和工作职位数据的证据 . *SSRN 工作论文。*
克莱因·特塞尔林克，B. (2025). 生成式人工智能与劳动力市场结果：来自英国的证据 . *SSRN 工作论文。*
Humlum，A.和Vestergaard，E.(2025)。 大型语言模型，小劳动力市场效应。 *NBER工作论文。*

一个理论模型显示，人工智能加速的研究可能会产生替代动态，从而根本性地重塑大学劳动力

• 一个近期的模型（Daley，2025）展示了越来越强大、低成本的人工智能研究系统可能如何负面与以指标为导向的大学和资助拨款流程互动。

- 假设研究能力每 ~16 个月翻倍（16 是任意值，任何大于零的值都有相同的渐近结论），该模型显示对人类研究劳动力的需求崩溃*指数级*——一种与自动化模型一致动态，即当机器替代人类任务时劳动份额下降（Acemoglu & Restrepo, 2018），这使得对研究核心方向的人类输入风险降低。

- 劳动经济学研究表明，当自动化取代人类劳动时，劳动者分享的产出萎缩和差距扩大（Autor & Salomons, 2018）。对于大学而言，这意味着博士后、研究助理的急剧缩减计算研究人员作为实验室用廉价、可扩展的AI认知替代“常规”研究工作。

- 高技能、非常规性工作已不再孤立：近期证据发现人工智能现在取代了高度受训的知识工作者传统上执行的认知任务（例如，文献综述、数据分析、稿件起草）。

- 未经干预，大学可能会变成AI驱动的知识工厂，人类参与仅限于具有高度监管、物理或关系约束的任务。自动化研究始终发现，除非机构故意将工作转向互补的人类任务，替代占主导地位（Acemoglu & Restrepo，2018）。

- 政策杠杆，例如投资于机构上难以自动化且使人+人工智能共同生产的技术、将学术评估从原始计数（如引用次数、出版物数量）转变为质量调整后的指标（如社会影响力、数据/代码发布等），以及将资助机制改为部分基于彩票，可以减少可替代性并提高人类监督的底线，从而实质性地改变发展轨迹。

戴利，M. (2025)。 AI加速研究与大学劳动：一个以度量为驱动的变电站简单模型 。 *工作论文。*
阿西莫格鲁，D.和雷斯特雷波，P.(2018)。 人工智能，自动化与工作 。 *NBER工作论文。*
作者：D. 和 Salomons，A. (2018). 自动化是否排挤劳动力？生产力增长、就业和劳动份额 。 *NBER工作论文。*

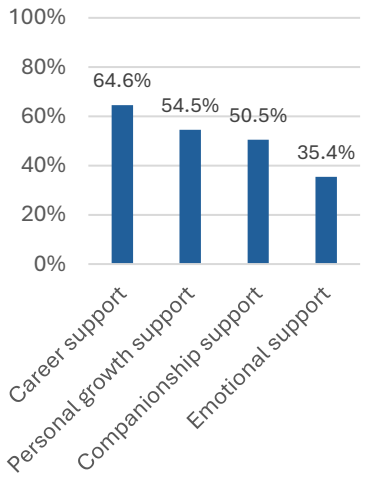
员工越来越使用人工智能不仅仅是为了任务帮助,他们是在利用它进行社交支持，然而大多数人仍然感到孤独。

• 以往关于个人情境中AI聊天机器人的研究显示效果不一。一些研究发现AI伴侣可以减少孤独感（De Freitas等人，待发表）；另一些研究则指出，随着人们依赖AI而非他人，AI可能促进社交退缩（Fang等人，2025年；Pentina等人，2023年；Tang等人，2023年）。

• 我们通过考察人工智能如何塑造工作场的人际关系来扩展这项工作。我们调查了1555名平均每天使用人工智能的美国知识型员工。参与者报告了他们
感知和在工作场所与人工智能互动，以及这如何与其社会支持感和孤独感相关。

• 人们在工作中拟人化人工智能：78.1%的人报告使用礼貌语言（例如，“谢谢你”），28.4%的人为人工智能选择了一个人文类比（例如，“队友”、“个人助理”），而不是技术类比。

• 员工们也正在使用人工智能提供类人的社交支持。通过使用一个调整后的人际关系功能清单(Colbert等人,2016年)，我们测量了人工智能在职业帮助、个人成长、陪伴和情感支持方面的使用情况。大多数参与者都在使用人工智能进行这些功能（见图表），并发现它们很有用。
然而52.2%的人仍然报告中度至高度的工作孤独感。工作满意度和留任意愿要低得多
与我们的研究中同事社会联系水平有关的变量，比与人工智能有关的任何变量都更紧密。



工作中使用人工智能进行社会支持
(Hadley and Wright, 即将出版)

• 总体而言，我们的研究表明，人工智能既有潜力增强 *和* 侵蚀工作中的社会体验。在接受开放式评论和故事时，一些参与者表达了这样的担忧：人工智能会制造更多壁垒分明的分工，侵蚀相互帮助和与他人建立联系的机会，并使他们因人工智能关系的非真实性而感到存在性孤独。我们鼓励研究人员、领导者和开发者继续研究人工智能如何影响工作中的社会福祉和凝聚力。

德·弗雷塔斯，J. 等人（待发表）。 人工智能伴侣减少孤独 。 *消费者研究杂志。*
方，C 等人。(2025)。 人工智能与人类行为如何塑造聊天机器人使用的心理社会效应：一项纵向随机对照研究 ， *arXiv工作论文。*
Pentina，I 等 (2023)。 人工智能时代的人机关系：系统文献综述与研究方向 。 *心理学与市场营销*
唐，P 等. (2023). 没有人是一座孤岛：剖析与人工智能互动的工作及工作后后果 。 *应用心理学杂志*
科尔伯特，A 等人。 (2016)。 通过职场关系实现繁荣：超越工具性支持。 *管理学会*
哈德利，C. 和赖特，S. (即将出版). 利用人工智能在工作中建立人际联系。在编辑过程中 *或哈佛商业评论。*

群体智能可能就是关键

• 几乎所有组织工作的方法
从层级公司到去中心化市场

单人一机组成的团队是集体智能系统（“超级大脑”）。

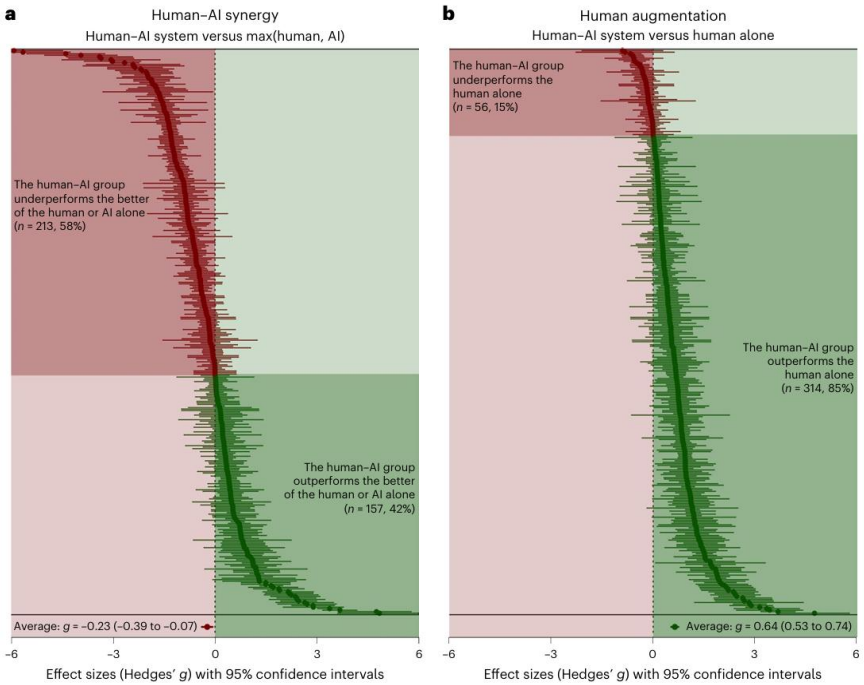
• 定义： *超级智慧* — 一组协同行动、（至少有时）看似智能的个体（Malone，2018）。

• 为了设计融合人类和人工智能的未来超级大脑，研究人员和开发者应当借鉴集体智能如何在多种系统中产生的知识，包括计算机科学中的网络、经济学中的市场、生物学中的蚁群等等（Malone & Bernstein, 2015）。

• 近期的研究已经表明，例如：

• 添加脚手架，例如引导步骤和定制界面，可以提高人机创造力远不止聊天机器人（Heyman 等人，2024）。

• 人机组合并非总是理想的。它们在创作任务上的表现通常优于决策任务。而且，如果人类单独就能胜过人工智能，加入人工智能往往有帮助，但如果人工智能已经更优秀，加入人类可能会造成伤害（Vaccaro等人，2024）。



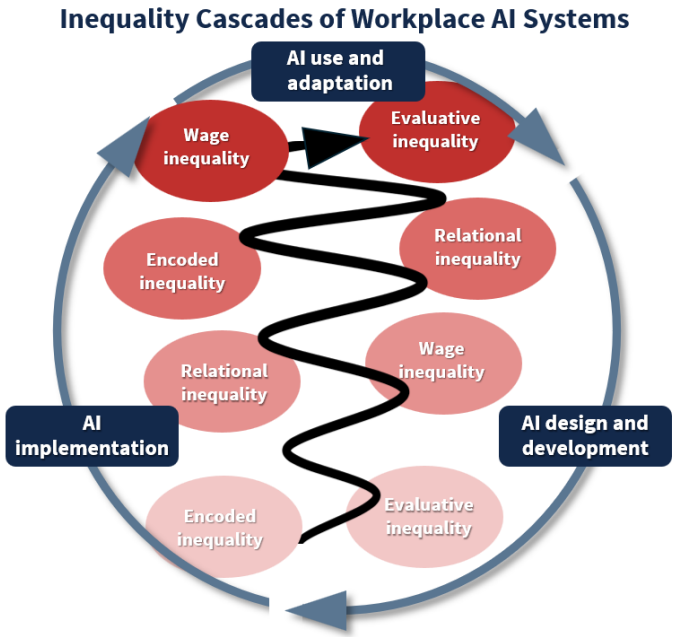
人类-AI协同效应（左侧）与人类增强（右侧）的效果大小比较。每个图显示了106项研究中的Hedge's g效应大小及其95%置信区间。红色区域表示人类-AI组表现不佳，绿色区域表示人类-AI组优于单独人类（或单独AI）（Vaccaro等人，2024年）。

马尔尼, T. (2018). 超级思维。人与电脑协同思考的惊人力量 . *小布朗。*
马尔农，T. 和伯恩斯坦，M. (2015)。 集体智能手册 . *麻省理工学院出版社。*
海曼，J. 等人. (2024). 超思维创想家：如何通过搭建人机协作框架来提升创造力 . *ACM集体智能会议。*
瓦卡罗, M. 等人 (2024). 当人类和人工智能的组合有用时 . *自然人类行为。*

人工智能可能会加剧职场不平等，除非组织积极采取措施干预以破坏“不平等级联。”

• 组织必须积极干预，围绕人工智能的设计、实施和使用建立护栏和结构，以减轻而不是放大工作场所人工智能系统的不平等级联效应（Kaurnakaran等人，2025年）。

- 人工智能对工作场所不平等的影响是多方面的，涵盖：
  • 工资不平等：不同工人的技能和专业知识的估值（相对于）在人工智能实施后在劳动力市场贬值的(devalued)
  • 编码不平等：某些群体（例如，AI开发者、管理者）的意识形态如何被铭刻到AI的设计中（Neely等人，2023年）。 • 评估不平等：不同工人的偏见和认知如何塑造其动机、判断以及对AI的信任。 • 关系不平等：权力、地位和权威的关系如何可能在人类-人工智能配置中变得不稳定。



人工智能对不平等的影响之间的联系 (Karunakaran 等人, 2025)

• 这些不等式的不同形式 *级联* 在人工智能生命周期中：上游做出的叙事和决策，例如谁参与（不参与）人工智能设计以及收集哪些（不收集）数据，会逐级向下传递，并且在人工智能系统在组织中日常实施和使用时可能变得根深蒂固（Drage 等人，2024）。

卡鲁纳卡兰，A 等 (2025). 人工智能在工作：人工智能对工作场所不平等影响的整合视角 . *管理学会*
尼利，M.T. 等人 (2023). 高科技领域的社会不平等：性别、种族和民族如何构建了全球最强大的产业。 *社会学年度评论。*
德拉格，E.等人（2024）。 工程师的责任：对谁对道德人工智能负责的 feminist方法 *伦理与信息技术。*