

生成式人工智能的影响：个体、组织与社会

陈 钊 张 洲 钟岳霖*

摘要：生成式人工智能(AI)在各领域快速发展并展现广阔前景,其社会影响备受关注。本文基于实证研究与跨学科框架,系统评估其在个体、组织与社会层面的多重影响。个体层面,生成式 AI 显著提升常规任务效率,但在复杂任务中存在局限,且会因为使用者 AI 素养差异而出现分化。对生成式 AI 的长期使用可能削弱批判性思维并加剧认知依赖。组织层面,该技术有助于团队协作与企业效率提升,但可能抑制创意多样性。社会层面,生成式 AI 重塑劳动力市场,既能缩小技能差距也会加剧 AI 素养差异导致的不平等,同时在社会偏见、意识形态与虚假信息治理中发挥双重作用。综合生成式 AI 的多重影响,我们认为,为了让 AI 更好地服务于人类价值的实现,个人应提升 AI 素养,兼顾效率提升与独立思考,政策层面则应关注认知主导与社会公平。

关键词：生成式人工智能;大语言模型;多维影响

DOI: 10.13821/j.cnki.ceq.2026.03.01

一、引 言

自 2022 年 OpenAI 推出 ChatGPT 以来,基于大语言模型(Large Language Model, LLM)构建的对话式人工智能(Artificial Intelligence,以下简称 AI)系统迅速引发全球关注,成为生成式人工智能(Generative AI,以下简称生成式 AI)最广为人知的应用代表。ChatGPT 展示了大语言模型在自然语言理解与内容生成方面的突破性进展,也因此成为用户增长速度最快的消费级应用之一(Hu, 2023)。

简洁易用的交互式产品也推动了生成式 AI 在各行业的快速普及。根据麦

* 陈钊,复旦大学中国社会主义市场经济研究中心、复旦大学经济科学智能研究中心;张洲、钟岳霖,复旦大学中国社会主义市场经济研究中心。通信作者及地址:钟岳霖,上海市杨浦区国权路 579 号,200433;电话:13978454088;E-mail:24110680030@m.fudan.edu.cn。本文受复旦大学当代中国经济与社会工作室 2025 年 AI 专题春季系列论文研讨会的启发而作,感谢刘志阔教授共同参与组织,感谢张宇、易管华、施希文、尚庆宇、鲁菁菁、陈诺、马淑华、马欣榕、付晶晶、王子方、侯佳宜先后进行论文报告。感谢教育部哲学社会科学重大专项(2023JZDZ019)的资助。

肯锡全球 AI 调查,2024 年全球使用 AI 技术的机构比例由 50% 左右跃升至 78%;其中,生成式 AI 的使用率更是在一年内由 33% 跃升至 71%,成为推动 AI 应用扩张的重要力量(图 1)。在投资层面,2024 年全球生成式 AI 私人投资达到 339 亿美元,较 2022 年增长超过 8.5 倍(Maslej et al., 2025),这一强劲增长势头凸显了生成式 AI 在技术与产业层面的战略价值。2025 年 8 月,国务院发布《关于深入实施“人工智能+”行动的意见》,进一步强调人工智能在重塑生产生活方式、推动生产力跃迁和生产关系深层变革中的潜在重要作用。

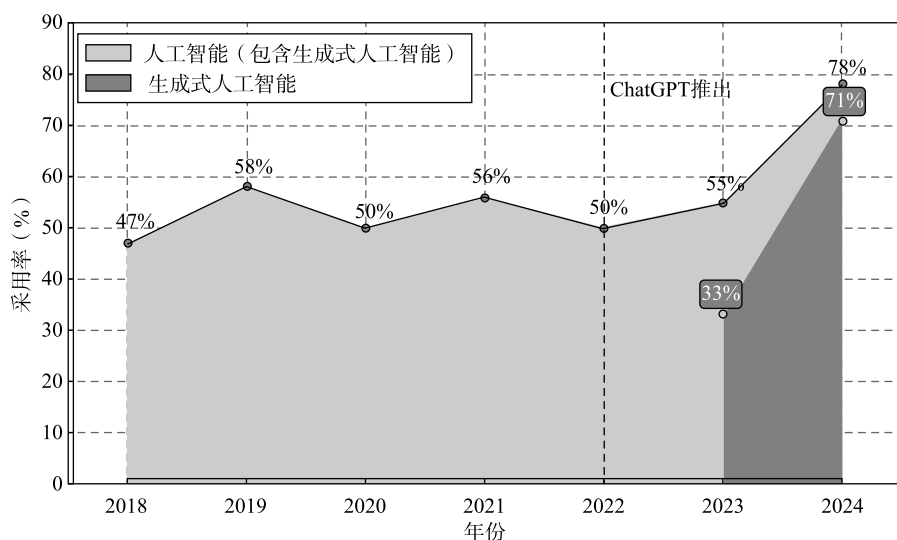


图 1 全球机构 AI 技术采用率变化趋势

注:采用率指至少有一个业务职能部门使用 AI 的公司或组织机构的比例。

数据来源:麦肯锡全球 AI 调查, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>, 访问时间:2025 年 7 月 30 日。

技术进步对经济的影响一直是经济学研究的重要主题,学界长期以来围绕其对生产率提升、就业结构调整、收入分配格局以及长期经济增长的作用展开了广泛的讨论。生成式 AI 作为一种突破式的技术进步正被广为应用,相应地,深入了解生成式 AI 对经济社会的潜在影响,对于理论与实践而言都显得尤为重要。

事实上,生成式 AI 的迅速发展与广泛应用,已经引起经济学界的关注。基于相关的早期文献,已经出现了三类不同的综述性研究。Xie and Avila(2025)基于生成式 AI 的技术特征与影响机制,对其潜在的社会影响进行了理论层面的分析与探讨。Capraro et al.(2024)、Mo and Ouyang(2025)和杨扬等(2025)聚焦于生成式 AI 影响的实证证据,重点探讨其在劳动力市场、教育、医疗、金融及行为科学等领域的具体影响。Korinek(2023)以及胡诗云和易君健(2025)则更多地介绍了生成式 AI 的应用方式,尤其是在经济学研究中的实际应用。

本文同样关注生成式 AI 影响的实证研究,并特别注重跨学科的综合视角。具体来说,为了尽可能地涵盖不同学科对生成式 AI 影响的最新研究,我们将遵循问题导向的思路,对多领域的最新实证成果进行较为系统的整合。以经济学理论为基础,我们将构建涵盖个体、组织与社会三个层次的分析框架,试图更清晰地揭示生成式 AI 在各层面的作用机制及其内在联系,从而帮助读者从更为全面的视角理解生成式 AI 的深远影响。表 1 对现有综述文献及本文工作进行了一个高度概括的比较。

表 1 生成式 AI 影响的文献综述比较

文献	综述类型	覆盖层面	主题侧重
Xie and Avila(2025)	理论分析	社会	国际竞争,社会不平等
Capraro et al.(2024)	实证证据	社会	工作、教育、医疗领域的社会经济不平等
Mo and Ouyang(2025)	实证证据	个体-组织-社会	金融领域:研究工具、金融行为、劳动力市场
杨扬等(2025)	实证证据	个体	行为科学领域
Korinek(2023) ^①	应用方法	个体	经济学研究工具指南
胡诗云和易君健(2025)	应用方法	个体	经济学研究全流程辅助
本文	实证证据	个体-组织-社会	跨学科全面总结生成式 AI 影响

二、生成式 AI 简介^②

(一) 生成式 AI 的历史演进

生成式 AI 是指通过对大规模训练数据进行学习与建模,能够在文本、图像、音频和视频等多种形式上生成连贯、新颖内容的 AI 技术。生成式 AI 与专注于分析、分类或预测的判别式 AI 存在显著差异。作为此领域的前沿进展,以 GPT 系列为代表的大语言模型极大地提升了内容生成的质量与复杂度,并推动了其在各领域的快速应用。

从广义上讲,生成模型的思想可以追溯至 20 世纪初的马尔可夫链(Markov, 1913)。该模型证明了可以通过概率来建模数据中事件间的顺序依赖关系,为

^① 安东·科里内克(Anton Korinek)在 2024 年和 2025 年先后发布了多次更新版本,包括“LLMs Level Up—Better, Faster, Cheaper”“LLMs Learn to Collaborate and Reason”和“AI Agents for Economic Research”。这些文章进一步讨论了生成式 AI 在推理能力、协作工作空间以及自动 AI 代理辅助经济学研究等应用的最新进展。详情可见 <https://www.genaiforecon.org/>,访问时间:2026 年 1 月 23 日。

^② 这一部分的写作中,我们参考了 Cao et al.(2025)和 Zhao et al.(2023)撰写的关于生成式 AI 和大语言模型的综述。

统计生成模型的出现建立了理论基础。20世纪80年代末至90年代,基于概率方法的统计生成模型,如N-gram语言模型(Shannon, 1948; Jelinek, 1990)和隐马尔可夫模型(Rabiner, 1989),逐渐成为主流。它们依托马尔可夫假设构建词序列预测模型,从而能够生成符合统计规律的文本或语音。然而,尽管能够在一定程度上模拟人类语言行为,这些模型本质上只是统计工具,与现代基于神经网络的人工智能仍存在较大差异。

进入21世纪10年代,随着深度学习^①的崛起,生成式AI取得了飞跃式的发展,并在不同领域呈现出多样化的创新。在自然语言处理(Natural Language Processing, NLP)领域,长短期记忆(Long Short-Term Memory, LSTM)模型(Hochreiter and Schmidhuber, 1997)作为一种特殊的循环神经网络(Recurrent Neural Networks, RNNs),凭借其“门控机制”(gating mechanism)对长序列信息的有效筛选与保留功能,在Graves(2012)等研究的推动下,被广泛应用于文本生成任务。而在计算机视觉领域,2014年提出的生成对抗网络(Generative Adversarial Networks, GANs)(Goodfellow et al., 2014)是一个重要成果,其核心思想是通过两个神经网络——生成器(generator)和判别器(discriminator)进行对抗性训练,生成高度逼真、真假难辨的图像。这一无监督学习方法在图像生成与创作领域得到了广泛应用。

2017年,Transformer框架(Vaswani et al., 2017)的问世,是生成式AI发展史上的一个决定性里程碑。不同于逐一按顺序处理词元的循环神经网络,该框架引入的“自注意力机制”(Self-Attention Mechanism)允许模型在处理序列数据时,同时关注并权衡上下文中所有词元(token)的重要性。该机制不仅显著提升了对长距离上下文的理解和记忆能力,还可以通过并行化处理大幅加快训练速度,天然适用于大规模数据集的训练。这种突破性架构不仅为文本、图像等领域的生成任务奠定了坚实基础,还通过统一处理不同模态的序列数据,为后续CLIP(Contrastive Language-Image Pre-training)(Radford et al., 2021)等跨模态对齐模型^②提供了架构基础。

基于Transformer架构,预训练语言模型(Pre-trained Language Model, PLM)应运而生,其中最广为人知的代表是BERT(Bidirectional Encoder Representations from Transformers)和GPT(Generative Pre-trained Transformer)。预训练语言模型确立了“预训练—微调”(Pre-train, Fine-tune)的核心学习范式:首先在海量无标签文本数据上进行大规模预训练,学习通用的语言知识,

① 深度学习(deep learning)是一种利用多层神经网络模拟人脑信息处理方式,从而实现特征提取与模式识别的计算方法。Hinton et al.(2006)提出的方法实现了多层神经网络的高效训练,推动了深度学习算法的广泛应用。

② 跨模态对齐是通过各种技术手段,实现不同模态数据(如图像、文本、音频等)在特征、语义或表示层面上的匹配与对应。

然后在特定任务上进行微调,以适应广泛的自然语言理解或生成需求。根据其架构侧重和预训练目标的不同,预训练语言模型主要分为两类:一类是以BERT为代表、基于编码器(encoder)的模型,这类模型通过双向注意力机制在语义理解任务上表现卓越;另一类是以GPT为代表、基于解码器(decoder)的模型,这类模型通过单向的自回归方式预测后续文本,在文本生成任务上能力出众。

在此背景下,GPT-1的推出成为生成式AI发展又一个关键节点。它首次成功实践了“生成式预训练—任务微调”的范式,证明了一个通用模型在不改动其核心架构的情况下,仅通过下游任务微调就能胜任多样化的文本生成任务,输出逻辑连贯且与上下文高度相关的高质量内容。这些基于Transformer的早期成果在多个任务上取得突破性表现,为后续模型在参数规模、训练数据和功能表现上的指数级提升,以及大语言模型时代的到来奠定了坚实基础。

(二) GPT系列的发展与大语言模型特征

1. 大语言模型的概念

作为当前最具影响力的生成式AI模型之一,GPT系列模型的参数规模与训练数据随其发展而不断扩张。在预训练语言模型被提出后,研究人员发现扩大语言模型的规模(包括参数数量和训练数据量)能够显著提升其性能(Kaplan et al., 2020)。遵循“规模定律”(Scaling Law),OpenAI持续扩大GPT模型的参数规模,相继推出了GPT-1(参数量1.17亿)、GPT-2(参数量15亿),并于2020年发布了参数量高达1750亿的GPT-3。伴随参数增长的,是训练数据集的急剧扩张:从早期包含约7000本书的BookCorpus,到包含超800万个网页的WebText,再到GPT-3所使用的一个大小约570GB,融合了网页、维基百科和书籍等来源的庞大语料库,这使模型得以掌握海量的世界知识(world knowledge)。

GPT-3最引人注目的突破在于,当模型规模达到一定阈值后,它在解决复杂任务时表现出了小型模型所不具备的、无法通过简单外推预测的全新能力。自然语言处理学界将这类规模庞大的预训练语言模型定义为“大语言模型”,并将其展现出的新能力称为“涌现能力”(emergent abilities)(Wei et al., 2022)。这种大语言模型的特有能力使其在各领域的任务中发挥出独特效果,也引发了广泛关注。图2直观地梳理了生成式AI、预训练语言模型与大语言模型等关键概念的分类与从属关系。

2. 大语言模型的能力特点

大语言模型能够胜任多样化任务的强大泛化能力,主要归因于其在规模扩展过程中涌现出的以下三方面关键能力。

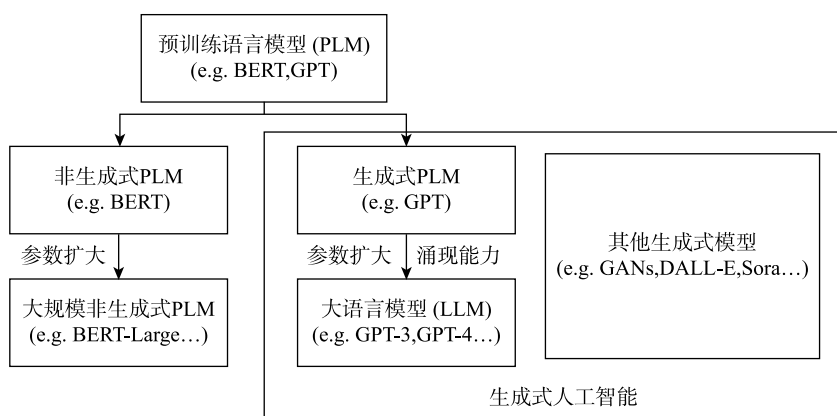


图2 生成式 AI 主要概念关系示意

注:在多领域应用中,以 BERT 为代表的编码器模型由于具有较大的参数规模,在语义理解等任务中表现出色,也被称为大语言模型。但依照更为通常的说法,本文所指的大语言模型是图中具有涌现能力的生成式预训练语言模型。

上下文学习 (in-context learning) 是大语言模型最典型的涌现能力之一, GPT-3 (Brown et al., 2020) 首次在超大规模模型上对其进行了系统验证与展示。上下文学习指的是大语言模型无需进行任何模型参数更新,仅通过提示词 (prompt) 中提供的少数几个示例就能理解并执行一个全新的任务,大语言模型也因此被称为少样本学习者 (few-shot learner)。

指令遵循 (instruction following) 是大语言模型经过训练后获得的另一项重要能力。为了进一步增强大语言模型的实用性,研究者们开发了更高级的微调技术,使其能够在没有示例的情况下,遵循从未见过的指令。例如,谷歌团队提出的 FLAN (Wei et al., 2021) 通过在大量以自然语言描述的多任务指令上进行微调 (即“指令微调”, instruction tuning), 显著提升了模型在未见过的任务上的泛化能力,使其成为高效的零样本学习者 (zero-shot learner)。而 OpenAI 团队则在 GPT-3 的基础上,通过“基于人类反馈的强化学习” (Reinforcement Learning from Human Feedback, RLHF) 技术来对齐模型,开发出 InstructGPT (Ouyang et al., 2022), 使模型的输出更符合人类的偏好和价值观,减少有害内容的生成。

逐步推理 (step-by-step reasoning) 也是大语言模型重要的涌现能力。研究发现,对于足够大的语言模型,采用“思维链” (Chain-of-Thought, CoT) 提示策略,即在提示中展示包含中间推理步骤的范例,可以显著提升模型解决数学题、逻辑推理等复杂任务的准确性 (Wei et al., 2022)。

更新的 GPT-4 系列模型则进一步增强了多模态处理和复杂任务解决能力。与此同时,生成式 AI 的前沿进展也体现在其他模态上,例如以 Stable Diffusion、Midjourney 和 DALL-E 为代表的文本生成图像 (文生图) 模型,以及以

Veo 和 Sora 为代表的文本生成视频(文生视频)模型。这些同样基于预训练范式的模型与大语言模型并行发展,各自凭借独特的生成能力和易于使用的特性,共同推动了生成式 AI 技术的广泛普及和应用。

3. 大语言模型的局限性

尽管大语言模型在能力上取得了巨大进展,目前仍然存在一些明显的局限性。在实际应用中,一个引起广泛关注的局限性就是幻觉(hallucination)问题。在自然语言处理领域,幻觉指模型生成的内容毫无意义,或与提供的源内容不符(Filippova, 2020; Maynez et al., 2020),其大致可分为内在幻觉(生成的信息与现有来源相冲突)和外在幻觉(信息无法通过现有来源验证)(Ji et al., 2023)。幻觉在现有的大语言模型中广泛存在,其来源于形成大语言模型能力的训练、推理等各个环节,且无法通过现有的微调与对齐策略完全解决(L. Zhou et al., 2024; Huang et al., 2025)。大语言模型因其概率性质而不具备对知识的辨别和控制能力,这种不可靠性为其应用带来风险,也可能加深人们对其偏见和担忧。

(三) 生成式 AI 的应用

生成式 AI 技术突破的价值,在广泛的应用中得到突出体现。这些应用正在重塑人类任务的形态,并成为推动个体、组织乃至社会层面深刻变革的重要起点。总体而言,生成式 AI 的应用可梳理为三个递进层次:从对非结构化信息的理解与结构化,到基于需求的内容生成与创作,再到具备自主性与交互性的智能体行为。

1. 内容理解与结构化

生成式 AI 的一个核心基础能力,是能够处理海量的非结构化信息,理解并提炼出清晰的逻辑与结构。以大语言模型为代表的生成式 AI,依托其强大的上下文学习能力和海量语料训练,在知识抽取与语义归纳方面展现出卓越性能。大语言模型能够在没有显式标注的情况下,从文本、语音、图像等杂乱信息中识别出关键实体,建立语义关系,并进行抽象与总结。面对一部长篇小说,大语言模型能够迅速识别主要人物与事件,并在理解其叙事脉络的基础上抽象出人物关系与时间线;在处理冗长的商业或学术文本时,它能够通过语义分析提炼核心内容,从而显著降低信息处理成本。例如,H. Fang et al.(2025)就利用大语言模型,从 300 万份政府文本中提取了中国产业政策的目标、行业、基调、手段等各维度的丰富信息,其规模和深度在以往的信息处理手段下是难以想象的。这一能力能够帮助个体减少阅读与筛选所耗费的时间,降低信息处理的认知负荷,从而提升学习和决策效率;也能使组织在知识管理与沟通中降低成本,从而加快决策与协调的进程。

2. 信息生成与创作

除了可以完成信息的理解与结构化之外,生成式 AI 更进一步展现出主动创造的能力,即生成全新的、有价值的增量内容。它不仅仅是被动地组织信息,而是能够按照用户的目标与需求,生产出涵盖文本、图像、音频、代码在内的多模态内容。在日常办公和专业场景中,生成式 AI 可以根据指令生成标准化的文档、合同、邮件甚至测试代码,从而大幅度提升工作效率;在创意产业与营销领域,它又能够模仿不同的风格与语境,创作出艺术作品、广告文案或社交媒体内容,成为一种“创意伙伴”。例如,被广泛使用的 Cursor 等 AI 代码编辑器,让开发者无需手动编写精确的代码就能直接用日常语言下达指令,让生成式 AI 自动完成代码的生成、修改和纠错。于是,个体借助 AI 工具就能够突破自身的技能限制,完成以往需要专业训练才能完成的任务,推动工作方式由操作执行转向统筹与验证。

3. 交互智能体与模拟

随着技术深入发展,生成式 AI 的应用正从静态内容生成向动态任务执行演进,催生了“交互式智能体”(agent)这一新兴范式。智能体通常具备任务分解、工具调用、记忆管理与自我反思(self-reflection)等能力,能够在复杂情境中进行自主规划、执行多步任务,并通过角色模拟与人机协作来履行多样化职能。在商业场景中,它可以作为自动化代理独立完成客户沟通、订单处理乃至市场调研;在教育与医疗等公共服务领域,它能够模拟教师、教练或心理顾问,提供因人而异的指导与情感支持。例如,在营销领域,生成式 AI 智能体可以独立完成市场洞察、素材生成、效果分析、投放优化的全流程,极大提高了宣传推广的效率。这一能力使得个体能够以前所未有的低成本获得高度个性化的服务体验。

三、生成式 AI 在个体层面的影响

前文提到生成式 AI 的具体功能和应用场景,本部分将分析这些应用在个体层面的影响。我们将从效率影响和情感支持两方面展开讨论。就效率影响而言,我们将分析生成式 AI 对不同类型日常任务的作用以及适用边界,对个体技能获取和认知能力的影响,以及该影响是否存在明显的个体差异。就情感支持而言,我们将探讨生成式 AI 对心理健康和情感支持的作用,包括其中可能产生的情感依赖和风险。本部分最后总结了生成式 AI 在个体层面影响的核心文献,以便于读者比较不同研究的应用场景和实验设计(见表 2)。

(一) 效率和质量影响

关于生成式 AI 对个体工作效率和质量的影响,现有文献并未得出一致结

论。这主要是因为不同文献往往聚焦于不同性质的特定任务或工作岗位。为了更全面理解生成式 AI 的影响,有必要对不同类型的任务进行区分。可以借鉴的是,在对计算机替代劳动力的研究中,Autor(2003)提出了一种任务分类方法,并被后续研究广泛引用。在他的理解中,任务可以根据“体力-认知”和“常规-非常规”组合成四种类型。生成式 AI 理论上主要影响常规认知任务和非常规认知任务。前者重复性强且遵循固定流程,如行政助理、客服人员等从事的核心任务;而后者需要个体具备认知、分析、决策能力,如程序员、分析师、作家与设计师等从事的核心任务。

现有文献一致表明生成式 AI 提高了常规认知任务的效率和质量。这类任务的共同特征是流程相对固定、评价标准明确,恰好契合生成式 AI 在信息处理和格式化输出方面的核心优势。例如日常办公时,使用生成式 AI 工具(如 Microsoft 365 Copilot)的知识工作者在电子邮件管理上花费的时间显著减少,在文档撰写上也表现出更快的完成速度,但对多人会议时间的影响不大(Dillion et al., 2025)。在阅读图文影像等信息(information consumption)时, Kim et al.(2024)发现生成式 AI 生成的视频摘要能显著提升用户参与度和理解效率,尤其在长视频和实用型视频中效果更为显著。这说明生成式 AI 加速了个人的信息处理效率。

客服对话中,生成式 AI 能帮助客服人员更有效率地响应客户咨询(Ni et al., 2024)。其中,Brynjolfsson et al.(2025)基于准自然实验检验了生成式 AI 的引入对客服人员的影响。他们的研究发现借助生成式 AI 助手能提高客服人员的应答速度,同时提升客户服务质量。这种改进源于生成式 AI 能够快速调用大量语料库,将个性化的沟通需求转化为规范化的表达模板。这些证据表明生成式 AI 通过标准化处理和模板化输出,不仅提升了常规认知任务的工作效率,还极大保证了输出质量的一致和可靠。

生成式 AI 在非常规任务中的表现差异较大。综合接下来的文献研究,我们可以概括为它更擅长定义清晰、目标明确、不涉及复杂判断的任务,如专业资格考试、视觉和文本创作、教育辅导等。然而,当任务涉及多维度判断、高度不确定性或创造性思考时,例如对于学术科研、临床决策、高风险投资等任务,生成式 AI 的局限性就会暴露无遗。

生成式 AI 在简单非常规认知任务上有着突出的表现和增强效果,这在不同领域中都有例证。这也意味着在这些应用场景中,生成式 AI 能极大提升个人效率。例如专业考试和资格认证领域,Eulerich et al. (2024)发现经过提示词调试的 ChatGPT-4 能够通过所有会计资格考试,其整体均分(85.1%)可以达到与多数人类会计师相当的专业水平。同样法学考试中,Choi and Schwarcz (2024)发现 ChatGPT-4 辅助显著提高了学生的多选题成绩(提升了 29 个百分

位点)。经过提示词调试,ChatGPT-4 甚至能够独立在本科与高级法学课程中取得显著高于中位数的成绩。值得注意的是这种效果在客观题上表现突出,但在需要主观论述的题目上帮助有限^①。这些证据表明了生成式 AI 凭借强大的知识储备,尤其擅长在既定规则下进行高效的检索和推理。

生成式 AI 在技术开发领域的应用同样效果显著,因为这种模式化的代码生成任务恰好契合了生成式 AI 的核心能力。Peng et al.(2023)发现程序员使用 AI 工具后网页开发速度提升超过 55%。而 Cui et al.(2026)基于 4 867 名开发者的大规模随机对照实验研究显示,借助 GitHub Copilot 的开发者完成任务量平均提升 26.08%,且经验较少的开发者收益更为显著。这意味着在这些应用场景中因个人能力而产生的差异可能被显著缩小。

生成式 AI 在创意和分析任务中的表现更为复杂但总体积极,现有文献从日常创意、文本和视觉创作等领域进行了探索。在日常创意领域中, Lee and Chung(2024)通过五个日常创意实验验证了生成式 AI 的优势。在选择礼物、制作玩具、改造闲置物品和设计餐桌等创意场景中,ChatGPT 辅助相比传统搜索而言显著提升了方案的创造性。Urban et al.(2024)的实验进一步证实,ChatGPT 在创意生成任务中显著提升了完成任务的信心、解题质量和原创性。

在专业写作领域,Noy and Zhang(2023)提供了实验证据。写作任务包括新闻稿、简短报告、分析计划、电子邮件,人们独立完成这些任务平均费时 26 分钟。他们的实验表明,ChatGPT 能通过减少耗时和提高产出质量来显著提升专业写作的效率,同时缩小工作者之间的表现差距。在视觉创作中,Zhou and Lee(2024)对 400 万幅 AI 画图作品进行了分析。研究发现以文生图的 AI 应用将人均创作效率提高了 25%,点赞率提升了 50%。虽然这些作品的平均内容新颖性有所下降,但最高评分作品新颖性却得以增加。

此外,人机协作策略影响着生成式 AI 辅助的表现(Boussioux et al., 2024)。作者在众包平台招募了 125 名参与者进行创意实验。研究发现虽然纯人工方案在新颖性方面表现更佳,但人与生成式 AI 的协作在可行性和商业价值上更具优势。特别是在人类通过多轮互动引导生成式 AI 产生差异化的解决方案时,人机协作的表现显著优于生成式 AI 独立提供的结果。

以上文献说明生成式 AI 利用突出的概念整合和分析能力,使创意表达和内容呈现更加清晰生动。同时,恰当的人机协作策略设计能够进一步放大这种优势。然而,当非常规认知任务的复杂度进一步提升时,生成式 AI 开始显露其能力的不足。这些任务往往体量庞大、环环相扣,需要深度整合多种信息源和专业知识。正是在这种复杂性面前,越来越多文献发现了生成式 AI 的局限性。

^① 后续研究发现生成式 AI 辅助在法律实务上有着类似的影响(Choi et al., 2024)。该实地实验发现,在标准化程度高的合同起草、起诉状撰写任务上,生成式 AI 辅助显著提升完成质量。但在需要法律推理和客户经验判断的条款编撰、法律意见书撰写等复杂任务中,其对质量提升帮助有限。

例如对于学术写作与研究任务,生成式 AI 的参与降低了学术任务的原创性表现。Niloy et al.(2024)基于随机对照实验探究生成式 AI 对论文写作的影响,研究发现使用 ChatGPT 辅助的实验组论文的总创造力显著下降。虽然论文的内容表现和细节方面略有提升,但内容准确性和原创性出现显著削弱。更直接的证据来自 Si et al.(2025)的研究,他们让 43 位专家随机将人类学者或生成式 AI 提出的科研想法付诸实施并最终形成科研论文。结果表明,虽然生成式 AI 事前的研究想法更新颖,但其最终的科研产出在创新性等诸多指标上的表现都显著逊于人类。

复杂任务中生成式 AI 的局限在医疗决策中也有很好的例证。生成式 AI 在医疗领域不同类型的非常规认知任务上,呈现出清晰的能力范围。在标准化诊断任务中,生成式 AI 表现出色(Pais et al., 2024; Wang et al., 2025),人机协作甚至能超越单独工作的医生(Zöller et al., 2025)。然而,一旦涉及复杂临床决策,问题便开始显现。Hager et al.(2024)让多个大语言模型诊断 2 400 个真实临床病例,发现其对信息顺序、措辞变化高度敏感,诊断准确率始终低于医生。更令人担忧的是过度信任问题——患者往往无法识别生成式 AI 回复的错误,反而因其表面的完整性和可信度给予更高评价(Shekar et al., 2025)。这是因为标准化实验提供的病例信息都比较完整,有助于生成式 AI 发挥中立客观的优势准确诊断。但是在真实临床环境中直接可得的信息可能并不完整,这就需要医生根据专业经验观察并主动询问,这些依赖直觉的决策限制了生成式 AI 的作用。

对于其他不确定性强的高风险决策,也有文献印证了这类任务上生成式 AI 的局限。Spatharioti et al.(2025)发现,虽然生成式 AI 搜索能让用户解题更快且满意度更高,但在模型出错时用户也会更依赖错误信息。此时,需要通过高亮标注等方式提示不确定性,缓解用户依赖。在金融领域中,生成式 AI 辅助也存在着一道清晰的界限。如生成式 AI 可以有效普及“超额收益”和“市场风险”等基础概念,引导投资者购买市场组合基金以规避非系统性风险(Lu, 2025)。当生成式 AI 需要辅助个性化理财咨询时,仍然存在高风险情境中过度依赖历史数据、用户无法识别低质建议、难以调和用户冲突的需求等弊端。这些问题可能让决策者暴露在盲目风险中(Takayanagi et al., 2025)。

总体而言,生成式 AI 在不同类型认知任务上有着不一样的影响。在标准化、模块化任务中游刃有余,但面对需要全局思维和不确定性管理的任务时,其局限性便暴露出来。在复杂任务上,生成式 AI 的幻觉问题容易放大。一些细微的偏差可能在多轮对话中积累,并最终影响复杂决策的质量,而人机之间合理的分工有助于缓解彼此短板。这与简单任务中生成式 AI 的优异表现形成鲜明对比,说明生成式 AI 的能力边界并不平滑,呈现出“锯齿状边界”(Dell'Acqua et al., 2026)。

(二) 认知债务积累

尽管生成式 AI 在完成简单任务时发挥了较强作用,但这可能只是短期效果。许多人在享受生成式 AI 带来的便利时,不知不觉地失去了独立解决问题的能力。这种现象被称为认知债务(cognitive debt),指的是个体过度依赖外部智能系统时所积累的认知成本(Kosmyna et al., 2025)。长期积累的认知成本会导致人的批判性思维能力下降、对任务的掌控感降低。

认知债务最直接的表现是一旦离开生成式 AI,个体就很难独立完成任务。Wu et al.(2025)开展了数千人的实验研究,发现生成式 AI 虽然能提升即时任务表现,但这种提升仅限于短期影响。当参与者从 ChatGPT 辅助转向独立工作后,提升效果完全消失。这种情况在教育场景更令人担忧。Bastani et al.(2025)在高中生数学课上分别让实验组和对照组使用 ChatGPT-4 和课堂教材练习,而测试时所有人都闭卷独立完成。研究发现虽然 ChatGPT-4 提升了实验组的练习表现,但这种效果未能在测试成绩上得以体现。也就是说,学习者对工具产生依赖,规避了真正的思考过程,这意味着不当使用生成式 AI 可能损害长期学习。

认知债务更深层的表现在于虽然生成式 AI 能让人类更轻松地完成工作,但这种轻松的代价是大脑丧失了进行深度思考与判断的机会。由此可能产生的认知能力的退化得到了神经科学证据的支持。Kosmyna et al.(2025)基于脑电图实验和论文写作任务,发现使用生成式 AI 的组别大脑专注度和活跃度最低。即使写作满意度和质量提高,他们对作品的归属感也明显降低,在独立工作时记忆力和推理能力显著下降。相反,独立写作组的大脑思维强度最为活跃,并且其活跃程度在之后使用生成式 AI 时会进一步提高。

其他文献也从认知表现角度验证了这一发现。Stadler et al.(2024)比较了 91 名大学生使用 ChatGPT 和 Google 搜索引擎的学习效果,结果发现 ChatGPT 虽然显著降低了认知负荷(cognitive load),但学生最终报告的推理与论证质量都显著降低。这说明学习轻松并不意味着学习效果好。生成式 AI 在信息获取阶段减轻了认知负担,但同时也削弱深度学习所需要的高强度认知投入。Chen et al.(2025)在插画任务中观察到了类似现象。生成式 AI 会将质量曲线压平(quality-effort curve):使用者初期获益明显,但继续投入的回报骤降。结果导致大量插画师为节省时间主动降低成稿标准,这就削弱了工作过程中精益求精的动力,阻碍了更高层次的创新。

最令人担忧的是,完成任务的轻松感给个体一种无所不能的膨胀感,这可能导致个人过度高估自己的能力范围。这种对自己认知能力的错误估计被称为元认知偏差(metacognitive bias),它削弱了个体根据任务性质合理调整策略的能力。在教育领域中,较受关注的几篇文献探讨了这一现象及其危害。

Urban et al.(2024)发现使用 ChatGPT 的受试者虽然客观成绩提高,任务感知难度降低,但是他们的自我评分虚高于实际表现。这种外部工具带来的错觉让个体误认为是自身能力的提升。学生无需深度思考即可获得高分,这不仅降低了学习积极性,还削弱了深度调控自身认知的能力(Fan et al., 2025)。在一项生成式 AI 辅助编程教学的研究中也发现了这种危害,元认知困难的初学者会不加理解地进行代码粘贴,以及盲目依赖生成式 AI 的建议导致学习进度受阻(Prather et al., 2023)。

所幸的是,并非所有情形都会积累认知债务。Essel et al.(2024)基于 126 名加纳大学生的随机对照实验发现,课内教育适度使用生成式 AI 可以促进知识理解。在统计研究方法课教学中,实验组学生使用 ChatGPT 而对照组学生使用传统数据库和搜索引擎。研究发现实验组在批判性、创造性和反思性方面得分显著高于对照组。生成式 AI 帮助学生将抽样技术应用于现实场景,并以更吸引人的学习方式互动。这一发现提醒我们使用方式和程度是关键因素。在专业指导下的适度使用能够促进认知增强,而无节制的依赖则会积累认知债务,最终可能削弱人类独立思考和自我调节的核心能力。

(三) 个体表现的异质性

前述章节阐述了生成式 AI 在提升效率的同时也引发了认知债务的风险。然而,这些影响在不同个体之间并不相同。结合前文对简单和复杂任务的分类与讨论,生成式 AI 的个体异质性影响可以概括为以下特征:在简单任务上,生成式 AI 缩小了个体之间的差距;而在复杂任务上,生成式 AI 却加大了个体间的差距。对于后者而言,个体的专业经验和 AI 协作素养成为新的分化因素,进一步产生了新的不平等。

生成式 AI 在简单任务上和人类形成很好的互补,使得低技能水平人群也能显著提升工作效率,而高技能人群获益相对较小。这在不同领域的文献中均得到了验证(Brynjolfsson et al., 2025; Ni et al., 2024; Choi and Schwarcz, 2024; Noy and Zhang, 2023; Lu, 2025; Chen et al., 2025)。其中,Dell'Acqua et al.(2026)招募了 758 名咨询公司顾问开展实验,以检验 GPT-4 辅助真实商业任务的效果。结果显示在 GPT-4 可以胜任的任务中,它的增强效果呈现出明显的个体间差异:低绩效者增益 43%,高绩效者增益 17%。但在复杂任务(作者根据 GPT-4 能力极限而设定)中则没有增益效果。这说明在简单任务上,生成式 AI 可以缩小个体间的差距。在广告文案任务中,Chen and Chan (2024)发现新人使用生成式 AI 评估文案可显著提升点击率,并减小与经验丰富者的差距,但让其作为经验丰富职员的代笔反而拉低表现。

这些发现共同支持了一个重要的机制:生成式 AI 技术可以适配简单任务,弥补低技能者在从业经验和分析能力方面的不足;同时生成式 AI 在简单任务

中对高技能者增益较小,削弱了从业经验和能力的边际回报。这弱化了简单任务中不同个体技能差距的影响。

然而,当任务变得复杂时,生成式 AI 不一定能很好地适配,其表现高度依赖个体的协作能力和工作经验。此时个体间差距不仅没有缩小,反而进一步扩大。已有文献阐述了专业经验和能力对表现分化的影响。例如,He and Xia (2025)在线上医疗问诊平台的实验发现即使所有医生都获得患者提供的 Deep-Seek 预诊断信息,从业时间久的医生更容易利用其中信息提高就诊质量。这种差异源于经验丰富的医生具备更强的信息甄别能力和临床判断基础,能够有效整合生成式 AI 建议和自身专业知识。Sun et al.(2025)在某技术咨询公司开展实验,进一步验证了个体的元认知能力在人机协作中的关键作用。研究发现生成式 AI 显著提升了创造力,但擅长信息整合和分析能力的个体的增强效果较对照组放大近一倍。这些个体在方法有效性认知以及调用信息解决复杂任务方面有着更高的自评得分,能够更有效地在技术咨询任务上发挥人机协作的优势。

随着生成式 AI 的普及,人们在 AI 素养^①方面的差距正在取代传统的技能差距。这种新的不平等在编程学习中表现得尤为明显。Prather et al.(2024)观察了 21 位编程初学者使用生成式 AI 工具的学习过程,发现生成式 AI 工具能够帮助几乎所有人完成编程任务,但他们的学习效果却有很大差异。学习困难的初学者往往依赖生成式 AI 提供的错误代码,对这些错误缺乏足够的判断力,导致他们高估自己的能力;相对而言,学习能力较强的初学者能够有效运用这些工具加速编程进度,并选择性忽略不准确的提示。类似地,Jing et al.(2024)在 Python 绘图作业中也发现了学习者使用 ChatGPT 的有效性和 AI 素养有关。熟知生成式 AI 能力范围的学习者能更高效地解决问题,而缺乏相关素养的学习者则面临更多障碍。该研究指出,教学培养重点应该从具体问题转向学会使用生成式 AI 工具。所幸,AI 素养具有可塑性。通过适当的训练,人们可以学会更好地与 ChatGPT 协作(Li et al., 2024)。

总的来说,生成式 AI 能在适配的简单任务上提供技术支持,缩小低技能者和高技能者之间的差距。当从业者的经验和能力的回报因此降低时,人们不再有动力学习基础任务的技能,技能差距的减小进一步使低端工作的竞争变得更加激烈(Lysyakov and Viswanathan, 2023)。而在生成式 AI 不适配的复杂任务上,个体之间的表现被进一步拉大。此时拥有更高批判性思维能力、熟悉工具优缺点的素养正成为一种新的优势,其提高了能熟练运用生成式 AI 的从业者的高端工作的回报。这重新定义了人力资本的关键构成,也预示着对未来劳动力市场的影响,本文将在社会层面劳动力市场这一部分对此进行讨论。

^① 本文中的 AI 素养(AI literacy)定义是个体批判性评估 AI 技术、与 AI 有效沟通与协作,并能在日常生活与工作等情境使用 AI 工具的能力集合(Long and Magerko, 2020)。该定义所涵盖的素养维度,同样适用于评估运用生成式 AI 所需的核心能力。

（四）情感支持和交互

除了效率和认知影响之外,生成式 AI 也正在重塑个体的情感体验和社会交互模式。在当前心理健康需求激增而专业支持不足的背景下,生成式 AI 正在填补这一空白,但同时也可能带来情感依赖的风险,包括对真实人际关系的替代和负面的心理影响。我们将以生成式 AI 的情感支持能力为起点,进而分析其场景适应性,最后探讨过度使用的潜在风险。

凭借大规模预训练获得的丰富语料和先进对话能力,生成式 AI 在情感支持和心理健康干预方面展现出接近专业水准的表现。Yin et al.(2024)的研究发现,当参与者倾诉个人困境与情绪时,ChatGPT 的回应比人类更能让他们感受到积极信号。这是因为 ChatGPT 倾向于提供更多的情感支持而非纯粹的实用性解决方案,这让参与者感到被充分倾听和理解。更为重要的是,严格的临床实验证实了生成式 AI 的治疗有效性。Heinz et al.(2025)进行的为期 4 周的随机对照实验发现,基于生成式 AI 的聊天机器人 Therabot 可以显著减轻重度抑郁、广泛性焦虑和进食障碍高风险人群的症状,首次证明生成式 AI 能在临床水平缓解心理健康问题。

如果生成式 AI 的这些能力能够大规模应用,将有助于填补心理问题激增与专业人士短缺的供需矛盾。已经有文献开始探讨这种规模化应用的可能性。借助生成式 AI 的辅助,非专业的志愿者也有可能执行一些专业心理干预(Fu et al., 2023)。此外,大部分普通用户也可以利用生成式 AI 进行自主干预,改善自己的认知能力,从而更有效地摆脱负面想法(Sharma et al., 2024)。这些发现意味着生成式 AI 在心理干预和情感支持方面的应用前景越发重要。

除了标准化的心理健康干预,生成式 AI 还展现出灵活的场景适应性,能够根据不同用户需求、年龄群体和交互模式提供定制化的情感支持。这种适应性在跨年龄段的应用中得到了充分验证。

在儿童情感教育领域,Seo et al.(2024)开发了面向 8—12 岁儿童的情感对话机器人 CHACHA,发现借助大语言模型能鼓励并引导孩子讲述亲身经历,进行情绪表达。孩童倾向于将 CHACHA 视为密友并主动分享旅行、成就等故事。这就验证了利用大语言模型设计安全的、适应情感分享的助手的可行性,也为儿童情绪教育提供了范本。

在成人的自律培养方面,生成式 AI 技术展现出更为创新的应用形式。C. Fang et al.(2025a)开发的 ESV 系统运用情感语言模型和语音克隆技术,生成用户本人声音以模仿未来的自己。在为期三个月的实验中,虽然 ESV、文本、想象三种方式都能提升参与者的韧性、信心和目标承诺,但 ESV 因其个性化和沉浸感被认为最具优势。特别是在培养个人习惯方面,ESV 展现出独特价值,能

够在关键时刻进行精准干预。

然而,这些积极效果的背后也隐藏着不容忽视的风险。高频率使用不仅可能产生依赖,还可能损害用户真实的人际关系。C. Fang et al.(2025b)基于四周的随机对照实验,比较了文本、中性语音、热情语音三种人机交互模型以及开放、人际、非人际三类话题的效果。研究发现除低频使用,语音优势随使用量增多而消失。此外,高使用量在所有模式下都会增加孤独感、降低真实社交参与,并出现依赖和成瘾使用行为。研究显示出用户越依恋或信任生成式 AI 模型,其孤独感就越高。Zhang et al.(2025)进一步发现,虽然多数用户会与伴侣式聊天机器人建立类人际情感关系,但这种社交取向的使用整体上与较低的心理福祉相关。当机器人无法提供人际互惠与实质情感支持时,反而可能放大脆弱者的情感缺口。

总的来说,生成式 AI 机器人不仅具备临床级别的心理干预能力,还能够不同场景构建个性化情感支持。但与此同时,它也存在高频使用导致情感依赖并替代真实人际关系的风险。这种复杂性要求我们在推广生成式 AI 情感支持应用的同时,也应注意构建合理的使用指导原则。这对于适应个体需求,设计更安全、更有效的人机情感交互系统具有重要意义。

表 2 生成式 AI 影响的核心文献:个体层面

领域	文献	识别策略	实验地点	样本情况
(一) 效率和质量影响				
办公场景	Dillon et al.(2025)	自然实地实验	66 家大型企业	7 137 名员工(6 个月)
售后客服	Brynjolfsson et al. (2025)	准自然实验	单一企业	5 172 名代理(多个月)
售后客服	Ni et al.(2024)	自然实地实验	淘宝电商平台	5 940 名客服员工(2 个月)
视频摘要	Kim et al.(2024)	准自然实验	视频平台	106 个跨平台频道/979 对 站内视频
法律实务	Choi et al.(2024)	人造实地实验	明尼苏达大学法学院	60 名法学院学生
软件开发	Peng et al.(2023)	人造实地实验	GitHub 平台	95 名开发者
软件开发	Cui et al.(2026)	自然实地实验	3 家科技巨头公司	4 867 名开发者
创意任务	Lee and Chung (2024)	实验室实验	Prolific 和 MTurk 线 上实验平台	不同实验共 1 586 名受试者
创意任务	Urban et al.(2024)	实验室实验	捷克科学院实验室	145 名大学生
创意任务	Boussioux et al. (2024)	框架实地实验	Freelancer 众包平台	125 名众包解题者和 300 名 评价者
办公写作	Noy and Zhang (2023)	人造实地实验	Prolific 平台针对性 招募	453 名受过高等教育的专业 人士

(续表)

领域	文献	识别策略	实验地点	样本情况
艺术创作	Zhou and Lee(2024)	准自然实验	在线艺术分享平台	5.3 万名艺术家的 400 多万篇作品
医疗决策	Wang et al.(2025)	框架实地实验	武汉大学中南医院	196 名执业医师和 120 名学生
医疗决策	Hager et al.(2024)	性能评估		2 400 个真实病例
医疗决策	Shekar et al.(2025)	实验室实验	Prolific 线上实验平台	300 名非专业受试者
金融素养	Lu(2025)	自然实地实验	中国大型券商平台	2.88 万名投资者
(二) 认知债务积累				
教育学习	Bastani et al.(2025)	框架实地实验	土耳其高中课堂	50 个班级约 1 000 名高中生
学习任务	Kosmyna et al.(2025)	实验室实验	麻省理工学院媒体实验室	54 名佩戴脑电图设备的受试者
艺术创作	Chen et al.(2025)	框架实地实验	插画师个人工作室	219 名专业插画师
(三) 个体表现的异质性				
咨询任务	Dell'Acqua et al.(2026)	框架实地实验	波士顿咨询公司	758 名咨询顾问
知识创造	Li et al.(2024)	实验室实验	清华大学行为实验室	435 名(实验 1), 139 名(实验 2)
设计任务	Lysyakov and Viswanathan(2023)	准自然实验	在线创意平台	9 280 名设计师的 42.5 万份作品
(四) 情感支持和交互				
心理干预	Heinz et al.(2025)	框架实地实验	全国性临床招募	210 名临床症状患者持续 4 周干预
心理干预	Sharma et al.(2024)	框架实地实验	大型心理健康网站	1.5 万名真实用户
情感支持	Yin et al.(2024)	实验室实验	Prolific 线上实验平台	540 名初始参与者
情感支持	C. Fang et al.(2025b)	人造实地实验	CloudResearch 线上实验平台	981 名用户持续 4 周记录
情感支持	Zhang et al.(2025)	横截面分析	Character.AI	1 131 名受访者/41 万条消息

注：个体层面研究的识别策略整理参考 Harrison and List(2004)，将实验设计按“现场程度”从低到高划分为四类：实验室实验指在高度控制的实验环境中，通常招募学生或线上受试者，使用抽象任务与人为设定规则来考察因果机制；人造实地实验则在保留这种控制性和抽象框架的前提下，转向更加“真实世界”的非学生样本，如职场专业人士、开发者或患者等；框架实地实验在使用自然样本的基础上，引入与实际市场、组织或政策情境相匹配的语言、任务与产出指标，使实验任务更贴近被试日常决策环境；自然实地实验则完全嵌入真实的工作或生活场景中，被试在不知情的情况下参与随机化干预，研究者在尽量不干扰其正常行为的条件下收集行为与绩效数据。

四、生成式 AI 在组织层面的影响

生成式 AI 在组织层面同样展现出丰富的应用。在团队层面,它既可以通过整合不同意见提升协作效率与决策质量,也可能因共同的预训练生成模式而导致团队创意趋同。在企业等大型组织层面,生成式 AI 通过劳动力替代对企业价值产生积极作用。本部分还纳入了更广泛的人工智能影响研究,展示 AI 在促进企业产品创新与增强风险抵御能力等方面的作用。本部分最后总结了生成式 AI 在组织层面影响的核心文献(见表 3)。

(一) 团队协作增效与创意趋同

相较于个人工作,团队合作的成效在很大程度上依赖于对成员多样性的协调与利用,生成式 AI 的引入也为这一过程带来了新的机遇与挑战。

大语言模型强大的学习与整合能力,使其在信息收集与分析方面展现出独特优势。例如,它通过充分考虑团队中不同个体的偏好而有效促进成员间的协调,推动达成共识。Bakker et al.(2022)的研究表明,经微调的大语言模型能够生成兼顾多元背景的表达,从而凝聚群体意见。这使大语言模型在处理具体的协作事务(如会议时间和地点的协调)时,能够整合多方偏好、引导相互理解,不仅降低了沟通成本,还提升了成员满意度(Papachristou et al., 2025)。除了提供整合性建议,生成式 AI 还可以在团队决策过程中扮演“魔鬼代言人”(devil's advocate)的角色,推动团队作出更全面完善的决策。魔鬼代言人特指在讨论中为激发更深层次思考、确保全面审视问题而故意提出异议的人。Chiang et al.(2024)的研究发现,在 AI 辅助的群体决策任务中,如果大语言模型以互动方式持续对 AI 提出的建议进行质疑,可以有效避免团队对算法的盲目信任,显著提升最终决策的准确性。

然而,生成式 AI 的预训练特性也意味着,一旦模型训练完成,其知识库和参数便相对固定,这导致其在执行创造性任务时,倾向于遵循特定的生成模式和潜在立场。因此,当团队成员普遍依赖生成式 AI 来获取建议和解决方案时,他们各自的产出可能会变得更加相似,从而降低了团队整体的创意多样性。Doshi and Hauser(2024)的实验表明,在生成式 AI 的辅助下,参与者所创作的短篇故事在新颖性与质量上显著提高,然而其作品之间的差异性却随之降低。Meinke et al.(2025)在一项创意设计任务研究中也证实了这一点。他们基于 Lee and Chung(2024)关于 ChatGPT 对个体创造力影响的研究结果,在集体层面进行了进一步分析,发现由 ChatGPT 辅助产生的设计思路在集体层面表现出显著的趋同性。上述研究共同揭示了生成式 AI 在团队协作中带来的一个潜

在困境：其应用在提升个体效率的同时，可能导致团队最终产出的多样性与创造力下降。

（二）生成式 AI 对企业的影响

企业是个体实现协作与分工的一种组织形式。现有文献通过度量企业受生成式 AI 的冲击程度来探究其影响。

常见的做法是通过生成式 AI 能力与企业任务的匹配度来推算企业的“生成式 AI 暴露度”。Eisfeldt et al.(2023)以及 Labaschin et al.(2025)提供了衡量企业大语言模型暴露度的方法。他们均基于 Eloundou et al.(2024)的研究框架，将职业层面的大语言模型暴露度与由 Revelio Labs 提供的企业员工职业构成数据相结合，通过加权平均构建了美国上市公司的企业层面大语言模型暴露度指标。他们的研究发现，在 ChatGPT 发布后，受其影响程度较高的公司市值显著上升，而核心机制则是劳动替代效应：高暴露度企业中员工的核心任务更易被 ChatGPT 替代，从而使企业能够以 AI 技术替代部分劳动力，降低成本并提升利润(Eisfeldt et al., 2023)。他们的另一项研究则发现，大型企业普遍受到大语言模型的冲击，且企业暴露度与内部技术人员和具备 AI 技能员工的数量成正相关，这表明技术人员在推动大语言模型的推广与应用中发挥着关键作用(Labaschin et al., 2025)。

也有研究通过调查问卷或企业员工信息直接获取生成式 AI 技术的采用情况。例如，美国人口普查局的商业趋势与展望调查(Business Trends and Outlook Survey, BTOS)就包含了企业对各类 AI 技术，包括大语言模型应用的问卷。Bonney et al.(2024)的研究基于该调查指出，大语言模型的推出降低了 AI 的使用门槛，显著提升了小型企业中 AI 的整体使用率。

由于生成式 AI 的普及时间较短，直接以生成式 AI 为对象的企业实证研究仍相对有限。相较而言，关于 AI 应用对企业影响的研究已形成一定的积累。考虑到在企业层面上，生成式 AI 的应用机制及潜在影响与传统 AI 具有一定相似性，这里我们结合相关文献对 AI 在企业层面的影响进行综述。

作为一种通用目的技术(general-purpose technology)，AI 最直接的影响是提高效率和促进企业成长。Czarnitzki et al.(2023)利用 2018 年欧盟社区创新调查(Community Innovation Survey, CIS)的德国部分，研究了 AI 应用对企业生产率的影响，发现 AI 作为一种无形资本可显著提高企业生产率、销售额及附加值。Babina et al.(2024)的研究将就业履历数据库 Cognism 中的员工简历数据与美国上市公司数据库 Compustat 中的企业运营数据相匹配，同样证实了 AI 投入对企业增长的积极作用，并进一步发现这种增长更多来源于产品创新(而非流程改进)带来的效率提升。Antoniades et al.(2025)结合中国企业的海

关数据和在线招聘数据发现,营销、策略、分析等非生产岗位的 AI 应用可以帮助企业降低因消费者偏好信息不对称带来的摩擦,推动企业制定更有效的定价策略,推动出口扩张。Cockburn et al.(2018)则强调, AI 不仅是一种通用目的技术,更是作为研究工具深刻重塑着创新的流程。通过对论文和专利数据的实证分析,他们认为在人工智能时代,数据驱动的应用型创新将更为重要。

不过,识别 AI 对企业增长的积极影响需要注意可能存在的内生性问题。已有研究表明, AI 的采用主要集中于规模较大、生产率较高的企业,这一现象在多个发达国家的样本中均得到验证(Acemoglu et al., 2025; Calvino and Fontanelli, 2023、2024; Aghion et al., 2025)。Acemoglu et al.(2023)进一步指出,由于采用与未采用人工智能的企业在特征上存在显著差异,准确识别 AI 对企业增长的因果效应需格外谨慎。他们的研究发现, AI 采用与企业规模增长之间的相关性很大程度上是一种“选择效应”,往往是那些原本规模较大、增长较快的企业更倾向于引入以 AI 为代表的新技术。Hampole et al.(2025)则利用企业在 AI 普及前的毕业生招聘生源与相应大学后续 AI 人才的增长情况构建工具变量,证明了 AI 技术采用与企业销售额、生产率、利润及总就业增长之间的因果关系。

同时, AI 独特的预测能力还可以降低企业经营的波动性,增强其风险抵御能力。Babina et al.(2025)的研究强调了 AI 作为预测技术在降低企业经营风险、稳定业绩方面的作用。他们发现 AI 投入使企业在销售、盈利、现金流等方面的业绩波动性均显著下降。该团队的另一项研究(Babina et al., 2023)则发现, AI 投入反而提升了企业的系统性市场风险(即贝塔系数)。这是因为 AI 投入为公司创造了未来的增长机会,使企业更具“成长股”特性和顺周期性,导致经营状况与市场整体趋势的关联更为紧密。Han et al.(2025)的研究还发现除了应对常规市场波动之外, AI 在面对自然灾害等突发冲击时也能帮助企业表现出更强的韧性,使其更快从灾害损失中恢复。

表 3 (生成式)AI 影响的核心文献:组织层面

(一) 团队影响				
领域	文献	识别策略	实验地点	样本情况
团队决策	Papachristou et al.(2025)	智能体仿真、 人造实地实验	美国科技公司内	45 名公司员工
团队决策	Chiang et al.(2024)	实验室实验	Prolific 平台	350 名参与者
团队创意 生成	Doshi and Hauser(2024)	实验室实验	Prolific 平台	英国的 293 名写作者 与 600 名评估者
团队创意 生成	Meincke et al.(2025)	实验室实验	来自 Lee and Chung (2024)的实验数据	不同实验共 1 586 名受 试者

(续表)

(二) 企业影响				
领域	文献	识别策略	时间范围	样本情况
企业市值	Eisfeldt et al.(2023)	事件研究法	2022—2023年	美国2518家上市公司
企业生产率	Czarnitzki et al.(2023)	工具变量法	2011—2018年	德国5851家企业
企业创新	Babina et al.(2024)	工具变量法	2010—2018年	美国1993家非科技上市公司
企业采用事实	Acemoglu et al.(2025)	特征事实	2016—2018年	美国ABS调查 ^a 30万家企业
企业规模	Aghion et al.(2025)	双重差分法	2014—2023年	法国868家私营企业
企业规模	Acemoglu et al.(2023)	队列分析	1978—2018年	美国ABS调查、LBD数据库 ^b 企业
企业业绩表现	Babina et al.(2025)	长差分回归	2008—2020年	美国非科技上市公司1亿条员工信息

注：a为美国年度商业调查(Annual Business Survey, ABS);b为纵向商业数据库(Longitudinal Business Database, LBD)。

五、生成式AI在社会层面的影响

本部分将聚焦生成式AI在社会层面的三方面影响。首先,考察生成式AI对劳动力市场的冲击,探讨可能对现有的技能、资历与性别不平等的放大或削弱效应。其次,分析生成式AI在社会偏见问题中的双重作用:既具有消解刻板印象的潜力,也可能复制并放大既有偏见。最后,评估其在意识形态领域的影响,展示生成式AI如何凭借强大的说服能力塑造公众观念,并为虚假信息治理带来新挑战。^① 本部分最后总结了生成式AI在社会层面影响的核心文献(见表4)。

(一) 劳动力异质性冲击与不平等

生成式AI通过在个人与企业层面提高劳动生产率,其影响最终会传导至整个劳动力市场。如何在整体职业层面精确测度生成式AI的影响,并分析受其影响的核心群体,是当前研究的核心议题之一,众多学者提出了不同的方法。

Felten et al.(2023)沿用并拓展了Felten et al.(2021)的方法,系统性地评估了哪些职业最容易受到生成式AI的冲击。该研究将语言建模和图像生成等关键能力与O*NET数据库中的52种职业能力相匹配,为每种职业构建了一

^① 生成式AI的应用已渗透到众多领域,其影响呈现出显著的异质性。Capraro et al.(2024)指出,在(错误)信息、工作、教育和健康等领域,生成式AI在提高效率的同时,也带来了加剧社会不平等的风险,这为政策制定提出了新的挑战。本文对其已详细探讨的内容不再赘述。

个受生成式 AI 影响的暴露度指标。来自 OpenAI 的 Eloundou et al.(2024) 则利用大语言模型自身来评估其提升不同任务完成速度的潜力,构建了职业层面的生成式 AI 暴露度。上述研究一致发现,与过去的自动化浪潮不同,高学历、高薪酬的工作受到生成式 AI 的冲击更大。

多项基于在线招聘平台的实证研究为上述结论提供了有力证据。张丹丹等(2025)借鉴 Eloundou et al.(2024)的思路,测算了中国劳动力市场中各职业的大语言模型技术暴露度,发现暴露度较高的主要是对受教育程度要求较高和薪资较高的白领职业。该研究进一步揭示,技术暴露度与劳动力需求呈负相关:高暴露度职业的薪资增幅下降,职业内部的薪资差距拉大,同时对从业者的教育和工作经验要求也相应提高。Hui et al.(2024)、Teutloff et al.(2025)和 Demirci et al.(2025)基于自由职业招聘平台的研究均表明,生成式 AI 的出现导致写作、设计、编程等技能的需求和从业者收入显著下降。

也有研究通过问卷调查或挖掘企业招聘数据等途径,探究生成式 AI 的应用在社会层面的影响。Humlum and Vestergaard(2024,2025)对丹麦劳动者的大语言模型应用情况进行了两轮调查。其初步调查发现,ChatGPT 在工作中的应用已相当普遍,并且可能加剧现有的不平等,因为在不同群体间 ChatGPT 的使用率并不均衡:先前收入更高的劳动者更可能在工作中使用 ChatGPT (Humlum and Vestergaard, 2024)。Humlum and Vestergaard(2025)基于第二轮调查进一步发现,尽管生成式 AI 的使用已较为普遍,但其对劳动者收入与工作时长影响并不显著。Hosseini and Lichtinger(2025)从员工资历视角出发,发现自 ChatGPT 发布以来,在采纳生成式 AI 技术的企业中,初级员工的就业显著减少,而资深员工的就业相应增加。机制分析表明,这一现象主要源于这些企业对新员工招聘的大幅下降。Aldasoro et al.(2024)关注生成式 AI 应用的性别差异,研究基于纽约联储消费者预期调查的数据发现,男性的生成式 AI 使用率显著低于女性,且这一差异主要源于对生成式 AI 技术了解程度的不同。这一证据也印证了生成式 AI 对个人影响的新趋势:AI 素养日渐成为关键的人力资本,其差异正在劳动力市场中产生新的分化。

生成式 AI 的浪潮,是 AI 技术持续冲击劳动力市场的最前沿动态。尽管生成式 AI 有其独特性,但关于技能替代、薪酬分化和生产力提升等核心议题,早已在过往对更广泛 AI 技术的研究中得到了深入探讨。因此,为了更深刻地理解当前变革的全貌,本节也将整合更广泛的 AI 影响的研究,旨在为读者呈现一个更系统、更全面的分析。

人工智能对劳动力市场的影响,主要取决于其对劳动力的“替代”与“增强”这两种竞争性效应之间的权衡,不同研究由于侧重点与数据颗粒度各异,会得出不尽相同的结论(Frank et al., 2019)。有研究指出,从总体上看,AI 对劳动

力市场的影响可能相对有限。Acemoglu et al.(2022)的研究显示,在职业或行业层面上,AI的应用并未对就业与工资产生显著影响。Hampole et al.(2025)针对就业层面给出了一种解释,指出虽然AI暴露度较高的职业需求有所下降,但AI带来的整体生产率提升促进了经济总体的就业增长,这两种效应大体上相互抵消。

与此同时,大量研究强调了AI对特定职业的替代性冲击。Webb(2019)通过分析职业任务描述与AI专利文本的相似度,构建了一项职业受AI影响的指标。他指出,与传统软件或自动化技术不同,AI主要针对高技能任务,因此可能对高技能职业的就业产生负面冲击。Dominski and Lee(2025)利用2021—2025年间前沿AI技术能力的变化构建了动态的AI暴露度指标。他们基于美国当前人口调查(Current Population Survey, CPS)发现,职业层面的AI暴露度增加与总就业人数下降、失业率上升以及主要工作时长减少显著相关。Minniti et al.(2025)基于欧洲的地区级数据发现,区域性的AI创新活动导致中、高技能工人的工资受到压缩,并显著降低了地区劳动收入份额。陈琳等(2024)的研究则重点关注AI对不同任务类型劳动力的影响。他们的研究发现AI显著降低了企业对常规任务劳动力的需求,同时增加了对非常规任务(如管理与技术岗位)劳动力的需求,且AI暴露度较高的非常规型职业享有更高的工资溢价。这一发现也与本文在个体层面得到的证据相一致,即AI在常规任务上表现优异,但仍难以完全替代从事复杂非常规任务的个体。

还有一些文献更细致地关注AI对劳动者工作集约边际(如工作时长)的影响,但结论并不一致。与前文提到的Dominski and Lee(2025)不同,Jiang et al.(2025)的研究基于美国时间利用调查(American Time Use Survey, ATUS)发现,由于AI与人类工作的互补性,从事高AI暴露度职业的工人最终工作时间更长、休闲时间更少,这意味着AI的应用反而可能损害工作与生活的平衡。

AI的异质性影响不仅体现在技能与暴露度的差异上,现有文献还从性别和地区等多个维度探讨了AI的异质性影响可能会进一步放大既有的经济不平等。

从性别视角来看,AI技术浪潮可能为女性带来新的机遇。Cazzaniga et al.(2025)在六个发达和新兴经济体的研究中发现,女性更集中于受AI影响较大的认知型岗位,因此比男性更多地暴露于AI技术之下。这既带来了更高的替代风险,也意味着AI的互补性可能让女性获得更大的受益潜力。Albanesi et al.(2025)则基于欧洲国家的劳动力数据发现,AI暴露度越高的职业,其女性就业份额的增长也越显著,且这种积极效应在女性教育水平和劳动参与率较高的国家更为明显。但与此同时,女性在AI技术的采用中处于相对劣势地位,受限于认知差异等因素,她们利用AI工具提升表现的机会可能受到限制,从而放大了性别在机会层面的不平等(Humlum and Vestergaard, 2024; Aldasoro

et al., 2024)。

从地区视角来看, AI可能重塑区域间的联系与差距。Andreadis et al. (2025)利用美国的招聘数据考察了AI相关工作在各县的分布与扩张,他们发现AI工作的地理分布极不均衡,而郊区或创新活跃地区更有可能在AI浪潮中受益。Liu et al. (2025)使用中国的在线招聘数据评估了生成式AI对劳动力市场的影响,发现其对高薪白领岗位造成了显著的替代效应,且这种效应在发达的大城市更为突出,因此亦可能有助于缩小区域间的就业差距。Hunt et al. (2024)的研究则表明地理距离是AI就业增长的制约因素:与AI学术中心的距离会减缓技术的应用和相关岗位的增长,这可能阻碍非技术热点地区的人工智能发展。

(二) 社会观念塑造与偏见弥合

除了在具体工作任务上影响人类表现,生成式AI同样能在思想层面引导用户的价值观与社会观念。我们首先讨论生成式AI在社会偏见问题中的应用与潜在影响。

由于AI是人为设计的计算工具,生成式AI可以用于构建不具有偏见的算法与形象,从而使大众摆脱传统刻板印象的束缚。在围棋等传统棋类运动中,男性选手的成绩普遍优于女性,这一现象常常导致人类教师产生性别刻板印象。Bao et al. (2024)的研究利用生成式AI,将一位以卡通人物形象呈现的AI教师引入围棋教学。他们发现, AI教师不仅总体上加速了学生的围棋水平提升;由于其绝对中立的立场,它还能减轻有性别刻板印象的人类教师对女生在情绪等方面的负面影响,从而有效消除了实验前长期存在的棋力性别差异,促进了性别平等。在另一项研究中, Yuan et al. (2025)构建了一个由大语言模型驱动的互动小说框架,参与者在与大语言模型的互动过程中,通过代入基层体力劳动者(如清洁工、护工)的视角进行换位思考,亲身体会到这些工作的复杂性、压力与风险。这一过程显著加深了参与者对这些职业的理解,催生了高度的同情心与亲近感,有效减少了相关职业的污名化。此外,生成式AI技术的进步本身亦使人们的社会观念变得更为科学、理性。Jackson et al. (2023)的实验让不同宗教背景的参与者阅读介绍ChatGPT等AI技术或科学前沿进展的文章。结果发现,与阅读其他科学文章的对照组相比,了解了ChatGPT技术进步的参与者,其宗教信仰程度显著降低。

然而,预训练的AI很难彻底摆脱歧视。由于大语言模型在包含人类社会已有偏见的海量文本上进行预训练,这就不可避免地会习得并复制这些偏见。大量研究已指出,大语言模型存在性别、职业、种族和宗教等领域的刻板印象与

偏见(Nadeem et al., 2020; Abid et al., 2021; Li and Bamman, 2021; Tamkin, 2023; Cheng et al., 2023)。大语言模型在传播故事时还会像人类一样,倾向于关注、记忆和传递符合刻板印象的内容(Acerbi and Stubbersfield, 2023),这可能导致偏见被进一步放大。

因此,我们若忽视大语言模型自身固有的偏见,则可能酿成严重的潜在后果。Hofmann et al.(2024)指出,基于人类反馈的对齐训练仅仅掩盖了模型表层的种族主义言论。当不明确提及种族时,使用非裔美国人英语(African American English, AAE)的口吻进行交流,会激发大语言模型表现出比现实更为极端的隐性种族主义。若利用此类模型进行决策辅助,那么在就业能力分析中,AAE使用者可能被更多地推荐从事欠体面的工作;在模拟犯罪判决中,他们也更容易被判定有罪。隐性的偏见还可能引致群体间的效率差异与不平等。Agarwal et al.(2025)的研究指出,大语言模型提出的写作建议更符合西方写作风格,与参与实验的美国人相比,单条建议对印度参与者写作效率的提升幅度较小,同时大语言模型的使用还使其放弃了原有的表达习惯。医疗场景同样面临风险,Zack et al.(2024)的研究显示,GPT-4在进行诊断推理时,会过分夸大因种族或性别导致的患病率差异,并在设计诊疗方案时延续这种偏见。它会根据患者的种族和性别给出截然不同的诊断重点和治疗建议,这种做法可能延续并放大有害的社会偏见,严重破坏医疗公平性。

(三) 意识形态与虚假信息治理

在政治议题和公共政策的宣传中,大语言模型凭借其强大的说服作用,亦可以产生不可忽视的效果。

大语言模型的强大说服力来自其独特的说服策略和互动方式。^①首先,大语言模型生成的文本难以与人类作品区分(Jakesch et al., 2023b; Kreps et al., 2022),且其生成的说服力信息具有证据丰富、逻辑严密、分析充分、语气冷静的特点(Bai et al., 2025; Karinshak et al., 2023),在论证时擅长增强细节、承认局限、总结要点(Zhou et al., 2023),这使其更能引起共鸣、更具说服力,能够有效促成参与者发生态度与观念转变。其次,大语言模型能以极低的成本进行对话,其基于参与者背景生成的个性化说服信息可以显著提高说服力(Matze et al., 2024; Salvi et al., 2025)^②,甚至在沟通中能够有针对性地找到人类论述中的漏洞进行反驳(Costello et al., 2024),从而达到更好的说服效果。

^① Rogiers et al.(2024)对大语言模型的说服力进行了较为详细的综述。

^② 也有研究对个性化说服策略对政治说服力的提升效果表示否定(Hackenburg and Margetts, 2024; Argyle et al., 2025)。

一方面,这一能力有利于政府与公众之间进行高效的沟通和信息传递。现有研究发现,大语言模型在多个领域中可以发挥类似甚至超越人类的说服力,特别是在政治宣传(Goldstein et al., 2024; Palmer and Spirling, 2023; Hackenburg et al., 2025; Bai et al., 2025)和公共卫生(Spitale et al., 2023; Karinshak et al., 2023; Hou et al., 2025)议题当中。通过与大语言模型的对话,不仅可以改变大众在一些不成熟的新兴话题上(如网络隐私、太空旅行)的观念(Durmus et al., 2024),对于可能有根深蒂固立场的且高度两极分化的争议性议题(如武器禁令、碳税、疫苗和各类政治阴谋论等),大语言模型同样能有效改变参与者的态度,且这种影响可能在大语言模型辅助创作的过程中潜移默化地发生(Jakesch et al., 2023a)。

但另一方面,生成式 AI 的强大能力也为虚假信息的传播打开了方便之门。以 Deepfake 为代表的生成影像技术能够合成以假乱真的人脸和声音。大语言模型又天然存在幻觉问题,无法确保生成内容的真实性。生成式 AI 内容的广泛传播可能导致个人无所适从,如果被恶意用于欺诈性和误导性的信息生成与传播,将对公共舆论和政治生态构成强大的潜在威胁(Radsch, 2022; Capraro et al., 2024)。

在生成式 AI 引发的信息泛滥与不确定性增加的环境下,大众对“可信来源”的信息需求上升,更加注重来源权威性与可验证性。而传统、可靠的新闻媒体可以有效削弱人们对虚假信息的信念(Lazer et al., 2018; Altay et al., 2024),这可能提升主流新闻机构的社会信任与市场影响力。此外,我们也应以动态发展的视角看待大语言模型。随着技术的持续迭代与新方法的涌现,其现有局限有望被逐步克服。例如,X. Zhou et al. (2024)构建的 MUSE(multimodal misinformation correction)事实核查系统就揭示了大型语言模型在信息核查领域的巨大潜力。

表 4 生成式 AI 影响的核心文献:社会层面

(一) 劳动力市场冲击				
研究方向	文献	识别策略	时间范围	样本情况
	Felten et al.(2021)	指标构建	2010—2019 年	美国 O*NET 数据库 774 个 6 位码职业
劳动力 市场	Eloundou et al.(2024)	指标构建	2020—2022 年	美国 O*NET 数据库 923 个数据级职业
	张丹丹等(2025)	工具变量法	2018—2024 年	中国招聘平台的 125 万 招聘启事

(续表)

研究方向	文献	识别策略	时间范围	样本情况
	Humlum and Vestergaard(2025)	双重差分法	2023—2024年	丹麦11个职业的2.5万名劳动者和7000家机构
	Hosseini and Lichtinger(2025)	双重差分法、三重差分法	2015—2025年	美国28.5万家公司的2.45亿条招聘启事
	Acemoglu et al.(2022)	移动-份额工具变量法	2010—2018年	美国 Burning Glass 平台的招聘启事 美国上市公司5800万份
	Hampole et al.(2025)	工具变量法	2010—2023年	个人简历、1400万份招聘启事
	Webb(2019)	指标构建、外推预测	1980—2010年	美国 O* NET 数据库 964个职业
劳动力市场	Minniti et al.(2025)	自回归分布滞后模型	2000—2017年	273个欧洲地区(NUTS2)
	陈琳等(2024)	工具变量法、双重差分法	2016—2020年	中国3700多家上市公司的458万条招聘启事
	Jiang et al.(2025)	双重差分法	2004—2023年	美国12万名受访者的时间利用记录
	Albanesi et al.(2025)	特征事实	2011—2019年	16个欧洲国家
	Humlum and Vestergaard(2024)	特征事实、框架实地试验	2023—2024年	丹麦11个职业的1.8万名劳动者
	Andreadis et al.(2025)	特征事实	2014—2023年	美国各县级行政区在 Lightcast 平台的招聘启事

(二) 社会观念

任务领域	文献	识别策略	实验地点	样本情况
性别观念	Bao et al.(2024)	准自然实验	中国某围棋教学机构,2020—2021年	287名学生与36名教师
宗教信仰	Jackson et al.(2023)	纵向研究、人造实地实验、实验室实验	来自68个国家的样本,2006—2020年;在线实验	300多万样本;1371名参与者
种族倾向	Hofmann et al.(2024)	智能体仿真	基于不同方言的文本进行性格、能力、司法分析	2023—2024年12个版本的LLM

(续表)

任务领域	文献	识别策略	实验地点	样本情况
疾病诊断	Zack et al.(2024)	智能体仿真	对 NEJM Healer 等来源的医疗病例进行诊断	GPT-4(2023年)
LLM 生成内容辨别	Jakesch et al.(2023)	实验室实验	Lucid, Prolific 平台	美国 4 600 名参与者
(三) 意识形态				
任务领域	文献	识别策略	实验地点	样本情况
公共政策	Bai et al.(2025)	实验室实验	Prolific, Cloud-Research 平台	美国 4 829 名成年人
公共政策	Salvi et al.(2025)	实验室实验	Prolific 平台	美国 900 名参与者
虚假信息	Costello et al.(2024)	实验室实验	CloudResearch 平台	美国 2 190 名参与者
虚假信息	Spitale et al.(2023)	实验室实验	Facebook 平台	英、澳、加、美等国的 697 名参与者
公共卫生	Karinshak et al.(2023)	实验室实验	Amazon MTurk 平台	2 348 名参与者
公共卫生	Hou et al.(2025)	框架实地实验	中国上海和安徽的学校	180 个班级的 2 671 名家长

六、进一步的思考

(一) 偏好问题: 认知和心理的不一致

生成式 AI 技术能力的快速提升与大众接受度之间存在显著差距, 因此人类在面对 AI 生成内容时, 表现出复杂的情感。

现有文献印证了人类对生成式 AI 创作内容的负面偏见。Karinshak et al. (2023) 和 Lim and Schmäzle (2024) 的研究发现, 即使大语言模型生成的内容在客观质量上优于人类创作, 一旦披露其 AI 来源, 受众的接受度和说服效果便显著下降。这反映出人类在面对智能技术时, 存在理性认知与情感直觉的分离。也有研究认为人们并非对 AI 作品持有偏见, 而是更偏爱人类创作 (Zhang and Gosline, 2023)。这种对人类独特性的偏好在 Lecamwasam and Chaudhuri (2025) 关于音乐情感的研究中得到了进一步验证。人类创作虽然不完美, 但蕴

含着情感、经验和目的，这是生成式 AI 的作品所不具备的特质。实际上，这种对生成式 AI 的态度甚至更为复杂。Shen et al.(2024) 在一项大规模实验中发现，尽管参与者对人类作品的共情始终显著高于 AI 作品，透明地披露 AI 身份反而能显著提升人们对 AI 作品的共情意愿。当用户知道对方是 AI 时，他们会调整自己的评判标准，从而产生更为宽容和理解的态度。

我们应该怎么理解人类对生成式 AI 的偏好？从动态视角来看，当前这种“摇摆不定”的接受度可能只是过渡阶段的表现。随着人们对生成式 AI 特性的深入了解，这种不接受的态度或许就会消散，正如对先前的蒸汽机、电器以及互联网技术那样。

更有趣的是，人类对 AI 的复杂态度只是故事的一半。Laurito et al.(2025) 的研究揭示了一个更为深层的风险，即 AI 系统本身也表现出对同类生成内容的系统性偏好。大语言模型在评估内容时更倾向于正面评价由 AI 模型生成的文本，而非人类原创内容。这种偏见可能导致生成式 AI 对自己内容偏好的不断强化。随着 AI 系统越来越多地参与内容评估、筛选和推荐，这种“AI-AI 偏见”可能逐渐构建一个排斥人类原创内容的信息生态系统。在这个系统中，AI 的内容因为符合算法的评估标准而获得更高的评分和更广的传播，进而训练出更偏向 AI 内容的新一代模型。

我们应该如何在接纳生成式 AI 的同时，保持人类自身特质，不被 AI“带偏”？这些问题目前文献并没有太多突破，有待未来的研究继续探索。更深层次地，这也进一步触及了人类价值和主体性问题。

（二）人本价值和主体性

随着生成式 AI 技术逐渐发展，与 AI 共处是大势所趋。已有大量文献告诉我们，仅依赖 AI 可能产生认知依赖等危害，丧失自我的主体性和长期能力培养。为了实现生成式 AI 赋能而非对人类的替代，我们需要在工作设计、能力培养两个层面进行设计。

1. 工作设计层面：渐进式增强的智慧

现有的大量生成式 AI 工具专注于将复杂工作拆解为细碎任务并自动化完成，往往忽略了使用者在任务中的参与。正如前文分析所示，这种注重短期效率的模式可能导致对生成式 AI 工具的依赖，替代了使用者本应思考分析的过程，带来认知能力降低等危害。

引导式的生成式 AI 或许是一个可行的选项。Dhillon et al.(2024) 在写作实验中控制生成式 AI 的提示字数，发现随着字数增加，其辅助效果呈现“U”形曲线——低强度辅助效果有限，中等强度辅助甚至可能产生负面影响，但高强度的结构化辅助（如逐段提示）能显著提升质量与速度，特别是对于非熟练用

户。这种逐段辅助的方式并未增加认知负担,保持了使用者写作过程的主导权。

这种辅助方式的核心是渐进式增强。一键式输出结果并不符合人类的认知习惯,生成式 AI 辅助要能让使用者循序渐进,拾级而上。正如古人所言,“授人以鱼不如授人以渔”。我们应当让生成式 AI 工具从基础层面开始,随着用户熟练度和任务复杂度的提升而逐步增强支持力度,始终保持“辅助而不取代人类认知”的平衡,确保底层认知能力不会因为依赖而退化。

2. 能力与情感培养:人类价值的守护

在 AI 能够高效完成越来越多认知任务的年代,能力的培养将逐渐从具体技能转向思维层面的训练,例如提升批判性思维和 AI 素养。当生成式 AI 成为日常工作伙伴时,我们的工作有了质的转变:从信息收集转向了信息验证,从问题解决转向结果整合,从直接执行转向监督并调试结果。这种转变看似只是工作内容的调整,实际上正在潜移默化地影响我们的底层能力。现实表明,人们的批判性思维已经呈现出明显的情境依赖性。在重要任务中,大多数人还会仔细评估生成式 AI 的输出质量;但在次要任务、不熟悉的或者责任边界模糊的场景中,人们往往会放松警惕,无条件接受生成式 AI 的建议(Lee et al., 2025)。

这是对我们可能丧失独立判断力的提醒。生成式 AI 可以帮我们替代掉枯燥而烦琐的工作,如填报系统、汇总报表、收集信息。这是为了让人们把更多时间放在创造性工作、深入分析和真正热爱的事业中。在这个过程中,真正值得投入时间的不是某项技能,而是能够不断滋养我们的底层能力。

在许多基础任务不再由我们亲手完成的时候,批判性思维显得尤其重要。它能帮助我们质疑和改进生成式 AI 的输出,始终保持在设计层面纠错和干预的能力。然而,对大部分人而言,生成式 AI 输出的结果往往像一个黑箱,让我们容易因为生成式 AI 输出的流畅性和权威感而放松警惕,忽视了对结果的独立验证。

那么,如何在享受生成式 AI 的同时保持独立判断能力呢? Bućinca et al. (2025)认为良好交互设计是做到这一点的途径。例如通过对比式交互,我们既能知道生成式 AI 的输出结果,也能理解决策中涉及的原因。这种交互方式在提供生成式 AI 的便利建议的同时,也训练人类的独立思考能力。当然这只是其中一种方式,可供我们探索的空间远不止于此。

同样重要的是 AI 素养。生成式 AI 的能力在不断扩展,但真正决定人机协作效果的是使用者的 AI 素养水平(Li et al., 2024)。即使是最先进的模型,如果使用不懂得如何与之协作,效果也会大打折扣。相反,具备 AI 素养的人即使使用相对简单的工具,也能发挥出惊人的效率。这提醒我们教育和培训的重心应该从追求更高级的工具转向培养 AI 素养——理解 AI 的能力边界、掌握有

效的提示技巧、学会验证和优化 AI 输出。

在 AI 越来越强大的时代,人类的价值不在于与智能设备竞争算力和储备,而在于发挥独特的价值判断、情绪感知与创造力。这正是批判性思维和 AI 素养所要守护的核心。

综合上述分析,有两个层面可以让生成式 AI 服务于人。首先是设计理念的转变。生成式 AI 辅助设计上应做到渐进式增强,始终让人类保持认知的主导地位,而不是完全挤出使用者的思考。其次是能力培养的重点转移。我们需要着重提升批判性思维和 AI 素养,在生成式 AI 时代发挥人类的比较优势。

(三) 社会影响和制度安排

最后,生成式 AI 影响着社会发展和分配格局,这对我们的制度安排提出了新的挑战。

生成式 AI 在组织机构转化成可部署产品的门槛高于预期。首先,部署生成式 AI 本身需要相关的算力支持以及适配模型和维护产品的技术团队。如果缺少这些条件,行业或公司可能在部署生成式 AI 时遇到技术障碍,抑或推出大模型产品可控性差、效果不如人意等问题,这些问题直接影响生成式 AI 的使用效率。其次,不同地区之间 AI 利用的资源分布极不平等,机会主要聚集在少数发达城市。对没有相应技术和认知储备的地区来说,采纳生成式 AI 成为一种系统性的障碍。上述分化可能加深地区间的不平等格局。生成式 AI 时代,我们如何能让社会各部分平等享受到技术进步的成果,这是一个有待探索的问题。

生成式 AI 的采用可能意味着什么?一方面,作为知识公共品,生成式 AI 理论上降低了创新创业门槛。对于有创收前景的项目,以往需要大量人力资源,现在小团队甚至个人都可能凭借生成式 AI 完成,极大地降低了前期的创业成本。从这个角度看,生成式 AI 充当了人力和资本的替代品,缩小了初创企业和龙头企业之间的鸿沟,促进了市场竞争。但另一方面,随着越来越多市场参与者应用相似的生成式 AI 工具和训练数据,各自产出的内容趋于相似。我们已经注意到,生成式 AI 在复杂任务处理中的表现并不理想。这种情况表明,生成式 AI 更多地推动了现有产品的维护和边际改进,而不是带来真正突破性的创新(Yeverechyahu et al., 2024)。这时我们需要思考两个问题:未来生成式 AI 是否应该从通用大模型转向个性化定制的垂域小模型,怎么获取更加细颗粒度的数据?生成式 AI 的普及如何才能全面促进重大发明的突破?

此外,生成式 AI 内容的爆发式增长带来了新的治理挑战。生成式 AI 批量产出的内容正充斥各个平台,涉及知识社区答疑、AI 文章、AI 视觉内容甚至是技术支持。虽然这些内容不乏优质作品,但更多是千篇一律的同质化创作。也

许随着模型的不断迭代,作品质量会得到改善。但是生成式 AI 本身的技术特征决定了它的量产优势和同质化特征。此时独创作品可能被淹没,这是否构成不公平竞争?从安全角度考虑,AI 生成内容可能造成谣言传播或安全隐患。那么在不标注来源的情况下,谁是这些危机的最终负责人?披露内容创作来源可以部分缓解以上问题,这不仅关乎消费者的知情权,也是对传统创作者的公平保护。如何披露,怎么保护,部分研究给出了实验证据(Epstein et al., 2023),更多的则有待未来研究进一步探索。

这些思考可以归纳为一个更底层的问题,生成式 AI 是否属于一种可持续的技术?随着我们对生成式 AI 负面影响认识的加深,例如认知退化、虚假信息、创造力衰退,不排除未来在某些领域可能限制甚至禁止生成式 AI 的使用。如果生成式 AI 普及率甚高,这意味着对生产效率和社会秩序的巨大冲击(Kreitmeir and Raschky, 2024; Bertomeu et al., 2025)。那些深度依赖生成式 AI 的个人、企业和行业,其工作环节已大量由生成式 AI 替代,相关人力资本也可能逐渐流失。面对这种风险,我们是否已经有足够的制度准备?如何保持技术独立性和风险缓冲能力?面对这些未知,我们需要的不是盲目乐观,而是更有前瞻性的思考。

七、总 结

ChatGPT 自问世以来的短短三年间,用户数量的爆发性增长与跨行业的快速渗透,使这一工具迅速融入公众的日常工作生活。从个人助理到创意伙伴,从数据分析到内容创作,以 ChatGPT 为代表的生成式 AI 得到广泛应用,标志着技术进步与社会应用之间的一次深度融合。

面对这样一个极具变革性的工具,人类社会对生成式 AI 仍有诸多未知之处。各学科的研究者们正从各自的领域出发,尝试探究生成式 AI 影响的不同侧面。本文在经济学逻辑下,整合了多学科关于生成式 AI 的核心研究成果,并通过构建涵盖个体、组织与社会多层次的跨学科分析框架,对生成式 AI 的作用机理与现实影响进行了系统性的综述。

生成式 AI 的作用首先体现在个体层面,即对人类能力的延伸与补充。无论是认知性任务中的智能辅助,还是情感层面的稳定与支持,生成式 AI 已在基础性工作中充分展现价值,并且正在逐步渗透至更复杂的认知任务。然而,生成式 AI 在提升短期效率的同时可能削弱长期认知能力,积累认知债务。大量用户快速采纳生成式 AI 技术,主要是因为其在效率提升方面的显著优势。但是高效便捷的使用体验背后,用户实际上付出了相应的认知成本,包括批判性思考能力下降、对任务掌控感和自我调控能力的缺失。更重要的是,生成式 AI

在简单任务中缩小了技能差距,却在复杂任务中因技术认知与使用方式的差异而加剧不平等。在专业技能回报下降的背景下,批判性认知、AI素养与价值判断等底层能力正成为新的核心竞争力。这一转变提示我们,应重新审视个体能力的价值排序,并着力培养能在AI时代保持长期回报的关键能力。这或许也能提醒我们,人文社会科学在AI时代将变得更为重要。

生成式AI在个体层面的影响进一步传导至劳动力市场。现有研究显示,AI不仅在赋能不同产业和岗位,还首次对高技能群体产生直接影响。与此同时,不同群体在AI工具可及性上的差异也可能加剧原有的不平等。这一现实提醒我们,技术的采用不可避免地伴随社会结构的相应调整。持续关注由此带来的分化与不平等,应成为学界与政策制定者的共同责任。

当然,我们也应关注生成式AI自身的技术局限。预训练机制既是生成式AI能力的来源,也决定了其内在局限。AI生成结果的多样性、公正性与客观性背后,潜藏着趋同性、隐含偏见以及幻觉错误等风险。这要求我们,在推动生成式AI应用的同时,更需以审慎的态度对待技术的发展,做到趋利避害。

鉴于以上发现,生成式AI未来的发展路径需要重点关注三个方向。在技术设计方面,生成式AI的辅助需要从“一步到位”转向“渐进式增强”,确保人类始终保持认知主导权;在能力培养方面,批判性思维和AI素养应成为核心竞争力,这是保持人类主体性的关键保障机制;在制度准备方面,需要系统解决技术可及性差异、内容治理挑战和风险防范机制,确保技术发展的包容性和可持续性。

当我们讨论生成式AI的未来时,实际上也是在探讨人类社会的发展方向。在这个意义上,生成式AI的影响已远超技术本身,映射出人类对效率和能力、平等和分化、创新与保守的价值取向。这种价值选择在技术高速发展的背景下显得日益重要。

生成式AI技术正在加速迭代,其社会影响亦将随着模型能力增强与应用场景渗透而逐渐加深。展望未来,我们应坚持以人的价值为核心,在充分发挥技术优势的同时保持人的主体性,使技术切实服务于社会与人类福祉。

参考文献

- [1] Abid, A., M. Farooqi, and J. Zou, "Persistent Anti-Muslim Bias in Large Language Models", *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021.
- [2] Acemoglu, D., G. Anderson, D. Beede, C. Buffington, E. Childress, E. Dinlersoz, L. Foster, N. Goldschlag, J. Haltiwanger, Z. Kroff, P. Restrepo, and N. Zolas, "Advanced Technology Adoption: Selection or Causal Effects?", *AEA Papers and Proceedings*, 2023, 113, 210-214.
- [3] Acemoglu, D., G. W. Anderson, D. N. Beede, C. Buffington, E. E. Childress, E. Dinlersoz, L. S.

- Foster, N. Goldschlag, J. C. Haltiwanger, Z. Kroff, P. Restrepo, and N. Zolas, "Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey", In: Basu, S., L. El-dridge, J. Haltiwanger, and E. Strassner (eds.), *Technology, Productivity, and Economic Growth*. Chicago: University of Chicago Press, 2025, 13-56.
- [4] Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo, "Artificial Intelligence and Jobs: Evidence from Online Vacancies", *Journal of Labor Economics*, 2022, 40(S1), 293-340.
- [5] Acerbi, A., and J. M. Stubbersfield, "Large Language Models Show Human-Like Content Biases in Transmission Chain Experiments", *Proceedings of the National Academy of Sciences*, 2023, 120(44), e2313790120.
- [6] Agarwal, D., M. Naaman, and A. Vashistha, "AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances", *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025.
- [7] Aghion, P., S. Bunel, X. Jaravel, T. Mikaelson, A. Roulet, and J. Sogaard, "How Different Uses of AI Shape Labor Demand: Evidence from France", *AEA Papers and Proceedings*, 2025, 115, 62-67.
- [8] Albanesi, S., A. Dias da Silva, J. F. Jimeno, A. Lamo, and A. Wabitsch, "AI and Women's Employment in Europe", *AEA Papers and Proceedings*, 2025, 115, 46-50.
- [9] Aldasoro, I., O. Armantier, S. Doerr, L. Gambacorta, and T. Oliviero, "The Gen AI Gender Gap", *Economics Letters*, 2024, 241, 111814.
- [10] Altay, S., R. K. Nielsen, and R. Fletcher, "News Can Help! The Impact of News Media and Digital Platforms on Awareness of and Belief in Misinformation", *The International Journal of Press/Politics*, 2024, 29(2), 459-484.
- [11] Andreadis, L., E. Kalotychou, M. Chatzikonstantinou, C. Louca, and C. A. Makridis, "Local Heterogeneity in Artificial Intelligence Jobs Over Time and Space", *AEA Papers and Proceedings*, 2025, 115, 29-34.
- [12] Antoniadis, A., C. He, Z. Liang, and M. Xu, "Learning by AI: The Impact of AI Hiring on Chinese Exporters", SSRN 5361273, 2025.
- [13] Argyle, L. P., E. C. Busby, J. R. Gubler, A. Lyman, J. Olcott, J. Pond, and D. Wingate, "Testing Theories of Political Persuasion Using AI", *Proceedings of the National Academy of Sciences*, 2025, 122(18), e2412815122.
- [14] Autor, D. H., F. Levy, and R. J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration", *The Quarterly Journal of Economics*, 2003, 118(4), 1279-1333.
- [15] Babina, T., A. Fedyk, A. He, and J. Hodson, "Artificial Intelligence and Firms' Systematic Risk", SSRN 4868770, 2023.
- [16] Babina, T., A. Fedyk, A. He, and J. Hodson, "Artificial Intelligence, Firm Growth, and Product Innovation", *Journal of Financial Economics*, 2024, 151, 103745.
- [17] Babina, T., A. Fedyk, A. He, and J. Hodson, "Artificial Intelligence Makes Firm Operating Performance Less Volatile", *AEA Papers and Proceedings*, 2025, 115, 35-39.
- [18] Bai, H., J. G. Voelkel, S. Muldowney, J. C. Eichstaedt, and R. Willer, "LLM-Generated Messages Can Persuade Humans on Policy Issues", *Nature Communications*, 2025, 16(1), 6037.
- [19] Bakker, M., M. J. Chadwick, H. R. Sheahan, M. H. Tessler, L. Campbell-Gillingham, J. Bal-

- agner, N. McAleese, A. Glaese, J. Aslanides, M. M. Botvinick, and C. Summerfield, “Fine-Tuning Language Models to Find Agreement Among Humans with Diverse Preferences”, *Advances in Neural Information Processing Systems*, 2022, 35, 38176-38189.
- [20] Bao, L., D. Huang, and C. Lin, “Can Artificial Intelligence Improve Gender Equality? Evidence from a Natural Experiment”, *Management Science*, 2024, 72(1), 474-494.
- [21] Bastani, H., O. Bastani, A. Sungu, H. Ge, Ö. Kabakçı, and R. Mariman, “Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics”, *Proceedings of the National Academy of Sciences*, 2025, 122(26), e2422633122.
- [22] Bertomeu, J., Y. Lin, Y. Liu, and Z. Ni, “The Impact of Generative AI on Information Processing: Evidence from the Ban of ChatGPT in Italy”, *Journal of Accounting and Economics*, 2025, 101782.
- [23] Bonney, K., C. Breaux, C. Buffington, E. Dinlersoz, L. S. Foster, N. Goldschlag, J. C. Haltiwanger, Z. Kroff, and K. Savage, “Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey”, *National Bureau of Economic Research Working Paper*, 2024, w32319.
- [24] Boussioux, L., J. N. Lane, M. Zhang, V. Jacimovic, and K. R. Lakhani, “The Crowdless Future? Generative AI and Creative Problem-Solving”, *Organization Science*, 2024, 35(5), 1589-1607.
- [25] Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language Models Are Few-Shot Learners”, *Advances in Neural Information Processing Systems*, 2020, 33, 1877-1901.
- [26] Brynjolfsson, E., D. Li, and L. Raymond, “Generative AI at Work”, *The Quarterly Journal of Economics*, 2025, 140(2), 889-942.
- [27] Buçinca, Z., S. Swaroop, A. E. Paluch, F. Doshi-Velez, and K. Z. Gajos, “Contrastive Explanations That Anticipate Human Misconceptions Can Improve Human Decision-Making Skills”, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025.
- [28] Calvino, F., and L. Fontanelli, “A Portrait of AI Adopters Across Countries”, *Documents de travail de l'OCDE sur la science, la technologie et l'industrie*, 2023.
- [29] Calvino, F., and L. Fontanelli, “AI Users Are Not All Alike: The Characteristics of French Firms Buying and Developing AI”, *CESifo Working Paper*, 2024.
- [30] Cao, Y., S. Li, Y. Liu, Z. Yan, Y. Dai, P. Yu, and L. Sun, “A Survey of AI-Generated Content (AIGC)”, *ACM Computing Surveys*, 2025, 57(5), 1-32.
- [31] Capraro, V., A. Lentsch, D. Acemoglu, S. Akgun, A. Akhmedova, E. Bilancini, J. -F. Bonnefon, P. Brañas-Garza, L. Butera, K. M. Douglas, J. A. C. Everett, G. Gigerenzer, C. Greenhow, D. A. Hashimoto, J. Holt-Lunstad, J. Jetten, S. Johnson, W. H. Kunz, C. Longoni, P. Lunn, S. Natale, S. Paluch, I. Rahwan, N. Selwyn, V. Singh, S. Suri, J. Sutcliffe, J. Tomlinson, S. Van Der Linden, P. A. M. Van Lange, F. Wall, J. J. Van Bavel, and R. Viale, “The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making”, *PNAS Nexus*, 2024, 3(6), pgae191.

- [32] Cazzaniga, M., A. Pantoni, L. Li, C. Pizzinelli, and M. M. Tavares, "A Gender Lens on Labor Market Exposure to AI", *AEA Papers and Proceedings*, 2025, 115, 56-61.
- [33] 陈琳、高悦莲、余林徽, "人工智能如何改变企业对劳动力的需求? ——来自招聘平台大数据的分析", 《管理世界》, 2024 年第 6 期, 第 74—93 页。
- [34] Chen, Y., J. Gong, J. Li, and Z. Zhao, "AI and Creative Process: The Time-Quality Relationship and Its Implications", SSRN 5208163, 2025.
- [35] Chen, Z., and J. Chan, "Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise", *Management Science*, 2024, 70(12), 9101-9117.
- [36] Cheng, M., E. Durmus, and D. Jurafsky, "Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models", *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023.
- [37] Chiang, C.-W., Z. Lu, Z. Li, and M. Yin, "Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate", *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 2024.
- [38] Choi, J. H., A. B. Monahan, and D. Schwarcz, "Lawyering in the Age of Artificial Intelligence", *Minnesota Law Review*, 2024, 109, 147-218.
- [39] Choi, J. H., and D. Schwarcz, "AI Assistance in Legal Analysis: An Empirical Study", *Journal of Legal Education*, 2024, 73, 384-420.
- [40] Cockburn, I. M., R. Henderson, and S. Stern, "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis", In: Agrawal, A., J. Gans, and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*. Chicago: University of Chicago Press, 2018, 115-146.
- [41] Costello, T. H., G. Pennycook, and D. G. Rand, "Durably Reducing Conspiracy Beliefs through Dialogues with AI", *Science*, 2024, 385(6714), eadq1814.
- [42] Cui, Z. K., M. Demirel, S. Jaffe, L. Musolf, S. Peng, and T. Salz, "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers", *Management Science*, 2026, forthcoming, <https://pubsonline.informs.org/doi/10.1287/mnsc.2025.00535>.
- [43] Czarnitzki, D., G. P. Fernández, and C. Rammer, "Artificial Intelligence and Firm-Level Productivity", *Journal of Economic Behavior and Organization*, 2023, 211, 188-205.
- [44] Dell'Acqua, F., E. McFowland, E. R. Mollick, H. Lifshitz-Assaf, K. C. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality", *Organization Science*, 2026, 37(2), 403-423.
- [45] Demirci, O., J. Hannane, and X. Zhu, "Who Is AI Replacing? The Impact of Generative AI on Online Freelancing Platforms", *Management Science*, 2025, 71(10), 8097-8108.
- [46] Dhillon, P. S., et al., "Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-Writing with Language Models", *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024.
- [47] Dillon, E. W., S. Jaffe, N. Immorlica, and C. T. Stanton, "Shifting Work Patterns with Generative AI", *National Bureau of Economic Research Working Paper*, 2025, w33795.
- [48] Dominski, J., and Y. S. Lee, "Advancing AI Capabilities and Evolving Labor Outcomes", *ArXiv*

- Preprint ArXiv*: 2507.08244, 2025.
- [49] Doshi, A. R., and O. P. Hauser, “Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content”, *Science Advances*, 2024, 10(28), eadn5290.
- [50] Durmus, E., L. Lovitt, A. Tamkin, S. Ritchie, J. Clark, and D. Ganguli, “Measuring the Persuasiveness of Language Models”, *Anthropic*, 2024.
- [51] Eisfeldt, A. L., G. Schubert, and M. B. Zhang, “Generative AI and Firm Values”, *National Bureau of Economic Research Working Paper*, 2023, w31222.
- [52] Eloundou, T., S. Manning, P. Mishkin, and D. Rock, “GPTs Are GPTs: Labor Market Impact Potential of LLMs”, *Science*, 2024, 384(6702), 1306-1308.
- [53] Epstein, Z., A. A. Arechar, and D. Rand, “What Label Should Be Applied to Content Produced by Generative AI”, *Unpublished Manuscript*, 2023.
- [54] Essel, H. B., D. Vlachopoulos, A. B. Essuman, and J. O. Amankwa, “ChatGPT Effects on Cognitive Skills of Undergraduate Students: Receiving Instant Responses from AI-Based Conversational Large Language Models (LLMs)”, *Computers and Education: Artificial Intelligence*, 2024, 6, 100198.
- [55] Eulerich, M., A. Sanatizadeh, H. Vakilzadeh, and D. A. Wood, “Is It All Hype? ChatGPT’s Performance and Disruptive Potential in the Accounting and Auditing Industries”, *Review of Accounting Studies*, 2024, 29(3), 2318-2349.
- [56] Fan, Y., L. Tang, H. Le, K. Shen, S. Tan, Y. Zhao, Y. Shen, X. Li, and D. Gašević, “Beware of Metacognitive Laziness: Effects of Generative Artificial Intelligence on Learning Motivation, Processes, and Performance”, *British Journal of Educational Technology*, 2025, 56(2), 489-530.
- [57] Fang, C. M., P. Chua, S. Chan, J. Leong, A. Bao, and P. Maes, “Leveraging AI-Generated Emotional Self-Voice to Nudge People toward Their Ideal Selves”, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025a.
- [58] Fang, C. M., A. R. Liu, V. Danry, E. Lee, S. W. T. Chan, P. Pataranutaporn, P. Maes, J. Phang, M. Lampe, L. Ahmad, and S. Agarwal, “How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Randomized Controlled Study”, *ArXiv Preprint ArXiv*: 2503.17473, 2025b.
- [59] Fang, H., M. Li, and G. Lu, “Decoding China’s Industrial Policies”, *National Bureau of Economic Research Working Paper*, 2025, w33814.
- [60] Felten, E., M. Raj, and R. Seamans, “Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses”, *Strategic Management Journal*, 2021, 42(12), 2195-2217.
- [61] Felten, E. W., M. Raj, and R. Seamans, “Occupational Heterogeneity in Exposure to Generative AI”, *SSRN* 4414065, 2023.
- [62] Filippova, K., “Controlled Hallucinations: Learning to Generate Faithfully from Noisy Data”, *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020.
- [63] Frank, M. R., D. Autor, J. E. Bessen, E. Brynjolfsson, M. Cebrian, D. J. Deming, M. Feldman, M. Groh, J. Lobo, E. Moro, and D. Wang, “Toward Understanding the Impact of Artificial Intelligence on Labor”, *Proceedings of the National Academy of Sciences*, 2019, 116(14), 6531-6539.
- [64] Fu, G., Q. Zhao, J. Li, D. Luo, C. Song, W. Zhai, S. Liu, F. Wang, Y. Wang, L. Cheng, J.

- Zhang, and B. X. Yang, "Enhancing Psychological Counseling with Large Language Model: A Multifaceted Decision-Support System for Non-Professionals", *ArXiv Preprint ArXiv*: 2308.15192, 2023.
- [65] Goldstein, J. A., J. Chao, S. Grossman, A. Stamos, and M. Tomz, "How Persuasive Is AI-Generated Propaganda?", *PNAS Nexus*, 2024, 3(2), pgae034.
- [66] Goodfellow, I. J., "On Distinguishability Criteria for Estimating Generative Models", *ArXiv Preprint ArXiv*: 1412.6515, 2014.
- [67] Graves, A., *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin; Heidelberg: Springer, 2012.
- [68] Hackenburg, K., L. Ibrahim, B. M. Tappin, and M. Tsakiris, "Comparing the Persuasiveness of Role-Playing Large Language Models and Human Experts on Polarized US Political Issues", *AI & Society*, 2025, 1-11.
- [69] Hackenburg, K., and H. Margetts, "Evaluating the Persuasive Influence of Political Microtargeting with Large Language Models", *Proceedings of the National Academy of Sciences*, 2024, 121(24), e2403116121.
- [70] Hager, P., F. Jungmann, R. Holland, K. Bhagat, I. Hubrecht, M. Knauer, J. Vielhauer, M. Makowski, R. Braren, G. Kaissis, and D. Rueckert, "Evaluation and Mitigation of the Limitations of Large Language Models in Clinical Decision-Making", *Nature Medicine*, 2024, 30(9), 2613-2622.
- [71] Hampole, M., D. Papanikolaou, L. D. W. Schmidt, and B. Seegmiller, "Artificial Intelligence and the Labor Market", *National Bureau of Economic Research Working Paper*, 2025, w33509.
- [72] Han, M., H. Shen, J. Wu, and X. Zhang, "Artificial Intelligence and Firm Resilience: Empirical Evidence from Natural Disaster Shocks", *Information Systems Research*, 2025, 36(4), 2116-2133.
- [73] Harrison, G. W., and J. A. List, "Field Experiments", *Journal of Economic Literature*, 2004, 42(4), 1009-1055.
- [74] He, H., and J. Xia, "Enhancing AI-Assisted Physician Decision-Making via Information Accuracy and Transparency: A Field Experiment", *Unpublished Manuscript*, 2025.
- [75] Heinz, M. V., D. M. Mackin, B. M. Trudeau, S. Bhattacharya, Y. Wang, H. A. Banta, A. D. Jewett, A. J. Salzhauer, T. Z. Griffin, and N. C. Jacobson, "Randomized Trial of a Generative AI Chatbot for Mental Health Treatment", *The New England Journal of Medicine Artificial Intelligence*, 2025, 2(4), AIoa2400802.
- [76] Hinton, G. E., S. Osindero, and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computation*, 2006, 18(7), 1527-1554.
- [77] Hochreiter, S., and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, 1997, 9(8), 1735-1780.
- [78] Hofmann, V., P. R. Kalluri, D. Jurafsky, and S. King, "AI Generates Covertly Racist Decisions About People Based on Their Dialect", *Nature*, 2024, 633(8028), 147-154.
- [79] Hosseini, S. M. and G. Lichtinger, "Generative AI as Seniority-Biased Technological Change: Evidence from US Résumé and Job Posting Data", *SSRN* 5425555, 2025.
- [80] Hou, Z., Z. Wu, Z. Qu, L. Gong, H. Peng, M. Jit, H. J. Larson, J. T. Wu, and L. Lin, "A Vaccine Chatbot Intervention for Parents to Improve HPV Vaccination Uptake among Middle School Girls: A Cluster Randomized Trial", *Nature Medicine*, 2025, 31(6), 1855-1862.

- [81] Hu, K., “ChatGPT Sets Record for Fastest-Growing User Base-Analyst Note”, *Reuters*, 2023.
- [82] 胡诗云、易君健, “智能之光: 人机协作的经济管理研究新时代”, 工作论文, 2025。
- [83] Huang, L., W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, and T. Liu, “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions”, *ACM Transactions on Information Systems*, 2025, 43(2), 1-55.
- [84] Hui, X., O. Reshef, and L. Zhou, “The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market”, *Organization Science*, 2024, 35(6), 1977-1989.
- [85] Humlum, A., and E. Vestergaard, “The Unequal Adoption of ChatGPT Exacerbates Existing Inequalities among Workers”, *Proceedings of the National Academy of Sciences*, 2024, 122(1), e2414972121.
- [86] Humlum, A., and E. Vestergaard, “Large Language Models, Small Labor Market Effects”, *National Bureau of Economic Research Working Paper*, 2025, w33777.
- [87] Hunt, J., I. M. Cockburn, and J. Bessen, “Is Distance from Innovation a Barrier to the Adoption of Artificial Intelligence?”, *National Bureau of Economic Research Working Paper*, 2024, w33022.
- [88] Jackson, J. C., K. C. Yam, P. M. Tang, C. G. Sibley, and A. Waytz, “Exposure to Automation Explains Religious Declines”, *Proceedings of the National Academy of Sciences*, 2023, 120(34), e2304748120.
- [89] Jakesch, M., A. Bhat, D. Buschek, L. Zalmanson, and M. Naaman, “Co-Writing with Opinionated Language Models Affects Users’ Views”, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023a.
- [90] Jakesch, M., J. T. Hancock, and M. Naaman, “Human Heuristics for AI-Generated Language Are Flawed”, *Proceedings of the National Academy of Sciences*, 2023b, 120(11), e2208839120.
- [91] Jelinek, F., “Self-Organized Language Modeling for Speech Recognition”, *Readings in Speech Recognition*, 1990, 450-506.
- [92] Ji, Z., N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung, “Survey of Hallucination in Natural Language Generation”, *ACM Computing Surveys*, 2023, 55(12), 1-38.
- [93] Jiang, W., J. Park, R. Xiao, and S. Zhang, “AI and the Extended Workday: Productivity, Contracting Efficiency, and Distribution of Rents”, *National Bureau of Economic Research Working Paper*, 2025, w33536.
- [94] Jing, Y., H. Wang, X. Chen, and C. Wang, “What Factors Will Affect the Effectiveness of Using ChatGPT to Solve Programming Problems? A Quasi-Experimental Study”, *Humanities and Social Sciences Communications*, 2024, 11(1), 1-12.
- [95] Kaplan, J., S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, “Scaling Laws for Neural Language Models”, *ArXiv Preprint ArXiv: 2001.08361*, 2020.
- [96] Karinshak, E., S. X. Liu, and J. S. Park, “Working with AI to Persuade: Examining a Large Language Model’s Ability to Generate Pro-Vaccination Messages”, *Proceedings of the ACM on Human-Computer Interaction*, 2023, 7(CSCW1), 1-29.
- [97] Kim, A., Y. Lu, T. Ma, and Y. Tan, “Less Is More? Impact of AI-Generated Summaries on User

- Engagement of Video-Sharing Platforms”, SSRN 5020305, 2024.
- [98] Korinek, A., “Generative AI for Economic Research: Use Cases and Implications for Economists”, *Journal of Economic Literature*, 2023, 61(4), 1281-1317.
- [99] Kosmyna, N., E. Hauptmann, Y. T. Yuan, J. Situ, X. -H. Liao, A. V. Beresnitzky, I. Braunstein, and P. Maes, “Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task”, *ArXiv Preprint ArXiv*: 2506. 08872, 2025.
- [100] Kreitmeir, D., and P. A. Raschky, “The Heterogeneous Productivity Effects of Generative AI”, *ArXiv Preprint ArXiv*: 2403. 01964, 2024.
- [101] Kreps, S., R. M. McCain, and M. Brundage, “All the News That’s Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation”, *Journal of Experimental Political Science*, 2022, 9(1), 104-117.
- [102] Labaschin, B., T. Eloundou, S. Manning, P. Mishkin, and D. Rock, “Extending ‘GPTs Are GPTs’ to Firms”, *AEA Papers and Proceedings*, 2025, 115, 51-55.
- [103] Laurito, W., B. Davis, P. Grietzer, T. Gavenčiak, A. Böhm, and J. Kulveit, “AI-AI Bias: Large Language Models Favor Communications Generated by Large Language Models”, *Proceedings of the National Academy of Sciences*, 2025, 122(31), e2415697122.
- [104] Lazer, D. M. J., M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain, “The Science of Fake News”, *Science*, 2018, 359(6380), 1094-1096.
- [105] Lecamwasam, K., and T. R. Chaudhuri, “Exploring Listeners’ Perceptions of AI-Generated and Human-Composed Music for Functional Emotional Applications”, *ArXiv Preprint ArXiv*: 2506. 02856, 2025.
- [106] Lee, B. C., and J. Chung, “An Empirical Investigation of the Impact of ChatGPT on Creativity”, *Nature Human Behaviour*, 2024, 8(10), 1906-1914.
- [107] Lee, H. -P., A. Sarkar, L. Tankelevitch, I. Drosos, S. Rintel, R. Banks, and N. Wilson, “The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers”, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025.
- [108] Li, L., and D. Bamman, “Gender and Representation Bias in GPT-3 Generated Stories”, *Proceedings of the Third Workshop on Narrative Understanding*, 2021.
- [109] Li, N., H. Zhou, W. Deng, J. Liu, F. Liu, and K. Mikel-Hong, “When Advanced AI Isn’t Enough: Human Factors as Drivers of Success in Generative AI-Human Collaborations”, SSRN 4738829, 2024.
- [110] Lim, S., and R. Schmäzle, “The Effect of Source Disclosure on Evaluation of AI-Generated Messages”, *Computers in Human Behavior: Artificial Humans*, 2024, 2(1), 100058.
- [111] Liu, L., J. Hu, A. Wang, and Z. Weng, “Generative Artificial Intelligence Causes Displacement for White-Collar Workers but Reduces Regional Opportunity Disparities”, *Research Square Preprint*, 2025.
- [112] Long, D., and B. Magerko, “What Is AI Literacy? Competencies and Design Considerations”, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.

- [113] Lu, F., “Robo-Advising Meets Large Language Models: Educating Investors on Alpha and Beta of Mutual Funds and Stocks”, SSRN 5083305, 2025.
- [114] Lysyakov, M., and S. Viswanathan, “Threatened by AI: Analyzing Users’ Responses to the Introduction of AI in a Crowd-Sourcing Platform”, *Information Systems Research*, 2023, 34(3), 1191-1210.
- [115] Markov, A. A., “An Example of Statistical Investigation of the Text ‘Eugene Onegin’”, *Bulletin of the Imperial Academy of Sciences of St. Petersburg*, 1913, 7(3), 153-162.
- [116] Maslej, N., L. Fattorini, R. Perrault, Y. Gil, V. Parli, N. Kariuki, E. Capstick, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J. C. Niebles, Y. Shoham, R. Wald, T. Walsh, A. Hamrah, L. Santarlasci, J. B. Lotufo, A. Rome, A. Shi, and S. Oak, “Artificial Intelligence Index Report 2025”, *ArXiv Preprint ArXiv*: 2504.07139, 2025.
- [117] Matz, S. C., J. D. Teeny, S. S. Vaid, H. Peters, G. M. Harari, and M. Cerf, “The Potential of Generative AI for Personalized Persuasion at Scale”, *Scientific Reports*, 2024, 14(1), 4692.
- [118] Maynez, J., S. Narayan, B. Bohnet, and R. McDonald, “On Faithfulness and Factuality in Abstractive Summarization”, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.
- [119] Meincke, L., G. Nave, and C. Terwiesch, “ChatGPT Decreases Idea Diversity in Brainstorming”, *Nature Human Behaviour*, 2025, 9(6), 1107-1109.
- [120] Minniti, A., K. Prettnner, and F. Venturini, “AI Innovation and the Labor Share in European Regions”, *European Economic Review*, 2025, 105043.
- [121] Mo, H., and S. Ouyang, “(Generative) AI in Financial Economics”, SSRN 5287110, 2025.
- [122] Nadeem, M., A. Bethke, and S. Reddy, “StereoSet: Measuring Stereotypical Bias in Pretrained Language Models”, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 2020.
- [123] Ni, X., Y. Wang, T. Feng, L. Xiaoyuan Lu, Y. Wang, and C. Zhou, “Generative AI in Action: Field Experimental Evidence on Worker Performance in E-Commerce Customer Service Operations”, SSRN 5012601, 2024.
- [124] Niloy, A. C., S. Akter, N. Sultana, J. Sultana, and S. I. U. Rahman, “Is ChatGPT a Menace for Creative Writing Ability? An Experiment”, *Journal of Computer Assisted Learning*, 2024, 40(2), 919-930.
- [125] Noy, S., and W. Zhang, “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence”, *Science*, 2023, 381(6654), 187-192.
- [126] Ouyang, L., J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, “Training Language Models to Follow Instructions with Human Feedback”, *Advances in Neural Information Processing Systems*, 2022, 35, 27730-27744.
- [127] Pais, C., J. Liu, R. Voigt, V. Gupta, E. Wade, and M. Bayati, “Large Language Models for Preventing Medication Direction Errors in Online Pharmacies”, *Nature Medicine*, 2024, 30(6), 1574-1582.
- [128] Palmer, A., and A. Spirling, “Large Language Models Can Argue in Convincing Ways About Poli-

- tics, but Humans Dislike AI Authors: Implications for Governance”, *Political Science*, 2023, 75(3), 281-291.
- [129] Papachristou, M., L. Yang, and C. -C. Hsu, “Leveraging Large Language Models for Collective Decision-Making”, *Proceedings of the ACM on Human-Computer Interaction*, 2025, 9(7), 1-4.
- [130] Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer, “The Impact of AI on Developer Productivity: Evidence from GitHub Copilot”, *ArXiv Preprint ArXiv*: 2302. 06590, 2023.
- [131] Prather, J., B. Reeves, J. Leinonen, S. MacNeil, A. S. Randrianasolo, B. Becker, B. Kimmel, J. Wright, and B. Briggs, “The Widening Gap: The Benefits and Harms of Generative AI for Novice Programmers”, *Proceedings of the 2024 ACM Conference on International Computing Education Research*, 2024.
- [132] Prather, J., B. N. Reeves, P. Denny, B. A. Becker, J. Leinonen, A. Luxton-Reilly, G. Powell, J. Finnie-Ansley, and E. A. Santos, “It’s Weird That It Knows What I Want: Usability and Interactions with Copilot for Novice Programmers”, *ACM Transactions on Computer-Human Interaction*, 2023, 31(1), 1-31.
- [133] Rabiner, L. R., “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, *Proceedings of the IEEE*, 1989, 77(2), 257-286.
- [134] Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, “Learning Transferable Visual Models from Natural Language Supervision”, *Proceedings of the 38th International Conference on Machine Learning*, *Proceedings of Machine Learning Research*, 2021, 139, 8748-8763.
- [135] Radsch, C., “AI and Disinformation: State-Aligned Information Operations and the Distortion of the Public Sphere”, *OSCE Representative on Freedom of the Media, Organization for Security and Co-Operation in Europe*, 2022.
- [136] Rogiers, A., S. Noels, M. Buyl, and T. De Bie, “Persuasion with Large Language Models: A Survey”, *ArXiv Preprint ArXiv*: 2411. 06837, 2024.
- [137] Salvi, F., M. H. Ribeiro, R. Gallotti, and R. West, “On the Conversational Persuasiveness of GPT-4”, *Nature Human Behaviour*, 2025, 9(8), 1645-1653.
- [138] Seo, W., C. Yang, and Y. -H. Kim, “Chacha: Leveraging Large Language Models to Prompt Children to Share Their Emotions About Personal Events”, *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024.
- [139] Shannon, C. E., “A Mathematical Theory of Communication”, *The Bell System Technical Journal*, 1948, 27(3), 379-423.
- [140] Sharma, A., K. Rushton, I. W. Lin, T. Nguyen, and T. Althoff, “Facilitating Self-Guided Mental Health Interventions Through Human-Language Model Interaction: A Case Study of Cognitive Restructuring”, *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024.
- [141] Shekar, S., P. Pataranutaporn, C. Sarabu, G. A. Cecchi, and P. Maes, “People Overtrust AI-Generated Medical Advice despite Low Accuracy”, *The New England Journal of Medicine Artificial Intelligence*, 2025, 2(6), AIoa2300015.
- [142] Shen, J., D. DiPaola, S. Ali, M. Sap, H. W. Park, and C. Breazeal, “Empathy Toward Artificial Intelligence Versus Human Experiences and the Role of Transparency in Mental Health and Social

- Support Chatbot Design: Comparative Study”, *JMIR Mental Health*, 2024, 11(1), e62679.
- [143] Si, C., T. Hashimoto, and D. Yang, “The Ideation-Execution Gap: Execution Outcomes of LLM-Generated Versus Human Research Ideas”, *ArXiv Preprint ArXiv*: 2506.20803, 2025.
- [144] Spatharioti, S., D. Rothschild, D. G. Goldstein, and J. M. Hofman, “Effects of LLM-based Search on Decision Making: Speed, Accuracy, and Overreliance”, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025.
- [145] Spitale, G., N. Biller-Andorno, and F. Germani, “AI Model GPT-3 (Dis) Informs Us Better Than Humans”, *Science Advances*, 2023, 9(26), eadh1850.
- [146] Stadler, M., M. Bannert, and M. Sailer, “Cognitive Ease at a Cost: LLMs Reduce Mental Effort but Compromise Depth in Student Scientific Inquiry”, *Computers in Human Behavior*, 2024, 160, 108386.
- [147] Sun, S., Z. A. Li, M. -D. Foo, J. Zhou, and J. G. Lu, “How and for Whom Using Generative AI Affects Creativity: A Field Experiment”, *Journal of Applied Psychology*, 2025.
- [148] Takayanagi, T., K. Izumi, J. Sanz-Cruzado, R. McCreddie, and I. Ounis, “Are Generative AI Agents Effective Personalized Financial Advisors?”, *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025.
- [149] Tamkin, A., A. Askill, L. Lovitt, E. Durmus, N. Joseph, S. Kravec, K. Nguyen, J. Kaplan, and D. Ganguli, “Evaluating and Mitigating Discrimination in Language Model Decisions”, *ArXiv Preprint ArXiv*: 2312.03689, 2023.
- [150] Teutloff, O., O. Reshef, and L. Zhou, “Winners and Losers of Generative AI: Early Evidence of Shifts in Freelancer Demand”, *Journal of Economic Behavior & Organization*, 2025, 106845.
- [151] Urban, M., F. Dëchtërenko, J. Lukavský, V. Hrabalová, F. Svacha, C. Brom, and K. Urban, “ChatGPT Improves Creative Problem-Solving Performance in University Students: An Experimental Study”, *Computers and Education*, 2024, 215, 105031.
- [152] Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need”, *Advances in Neural Information Processing Systems*, 2017, 30.
- [153] Wang, Z., L. Wei, and L. Xue, “Overcoming Medical Overuse with AI Assistance: An Experimental Investigation”, *Journal of Health Economics*, 2025, 103043.
- [154] Webb, M., “The Impact of Artificial Intelligence on the Labor Market”, *SSRN* 3482150, 2019.
- [155] Wei, J., M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, “Finetuned Language Models Are Zero-Shot Learners”, *ArXiv Preprint ArXiv*: 2109.01652, 2021.
- [156] Wei, J., Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus, “Emergent Abilities of Large Language Models”, *ArXiv Preprint ArXiv*: 2206.07682, 2022.
- [157] Wu, S., Y. Liu, M. Ruan, S. Chen, and X. Xiao-Yun, “Human-Generative AI Collaboration Enhances Task Performance but Undermines Human’s Intrinsic Motivation”, *Scientific Reports*, 2025, 15(1), 15105.
- [158] Xie, Y., and S. Avila, “The Social Impact of Generative LLM-Based AI”, *Chinese Journal of Sociology*, 2025, 11(1), 31-57.

- [159] 杨扬、孙乐轩、陈良源、林建浩,“人工智能时代的行为科学:三个值得关注的研究方向”,《计量经济学报》,2025年第6期,第1491—1508页。
- [160] Yeverechyahu, D., R. Mayya, and G. Oestreicher-Singer, “The Impact of Large Language Models on Open-Source Innovation: Evidence from GitHub Copilot”, *ArXiv Preprint ArXiv*: 2409.08379, 2024.
- [161] Yin, Y., N. Jia, and C. J. Wakslak, “AI Can Help People Feel Heard, but an AI Label Diminishes This Impact”, *Proceedings of the National Academy of Sciences*, 2024, 121(14), e2319112121.
- [162] Yuan, X., J. Wang, Q. Wan, and S. Hu, “A Day in Their Shoes: Using LLM-Based Perspective-Taking Interactive Fiction to Reduce Stigma Toward Dirty Work”, *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 2025.
- [163] Zack, T., E. Lehman, M. Suzgun, J. A. Rodriguez, L. A. Celi, J. Gichoya, D. Jurafsky, P. Szolovits, D. W. Bates, R. -E. E. Abdunour, A. J. Butte, and E. Alsentzer, “Assessing the Potential of GPT-4 to Perpetuate Racial and Gender Biases in Health Care: A Model Evaluation Study”, *The Lancet Digital Health*, 2024, 6(1), 12-22.
- [164] 张丹丹、于航、李力行、胡佳胤、莫怡青、李泓宇,“中国人工智能技术暴露度的测算及其对劳动需求的影响——基于大语言模型的新证据”,《管理世界》,2025年第7期,第59—75页。
- [165] Zhang, Y., and R. Gosline, “Human Favoritism, Not AI Aversion: People’s Perceptions (and Biases) Toward Generative AI, Human Experts, and Human-GAI Collaboration in Persuasive Content Generation”, *Judgment and Decision Making*, 2023, 18, e41.
- [166] Zhang, Y., D. Zhao, J. T. Hancock, R. Kraut, and D. Yang, “The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being”, *ArXiv Preprint ArXiv*: 2506.12605, 2025.
- [167] Zhao, W. X., K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J. Y. Nie, and J. R. Wen, “A Survey of Large Language Models”, *ArXiv Preprint ArXiv*: 2303.18223, 2023.
- [168] Zhou, E., and D. Lee, “Generative Artificial Intelligence, Human Creativity, and Art”, *PNAS Nexus*, 2024, 3(3), 52-59.
- [169] Zhou, J., Y. Zhang, Q. Luo, A. G. Parker, and M. Choudhury, “Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions”, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.
- [170] Zhou, L., W. Schellaert, F. Martínez-Plumed, Y. Moros-Daval, C. Ferri, and J. Hernández-Orallo, “Larger and More Instructable Language Models Become Less Reliable”, *Nature*, 2024, 634(8032), 61-68.
- [171] Zhou, X., A. Sharma, A. X. Zhang, and T. Althoff, “Correcting Misinformation on Social Media with a Large Language Model”, *ArXiv Preprint ArXiv*: 2403.11169, 2024.
- [172] Zöllner, N., J. Berger, I. Lin, N. Fu, J. Komaraveni, G. Barabucci, K. Laskowski, V. Shia, B. Harack, E. A. Chu, V. Trianni, R. H. J. M. Kurvers, and S. M. Herzog, “Human-AI Collectives Most Accurately Diagnose Clinical Vignettes”, *Proceedings of the National Academy of Sciences*, 2025, 122(24), e2426153122.

The Impact of Generative Artificial Intelligence: Individuals, Organizations, and Society

CHEN Zhao ZHANG Zhou ZHONG Yuelin*

(Fudan University)

Abstract: Generative artificial intelligence is rapidly advancing across domains, drawing growing attention for its broad social implications. This study employs an empirical and interdisciplinary framework to assess its multifaceted impacts at the individual, organizational, and societal levels. At the individual level, generative AI enhances routine task efficiency but shows limitations in complex tasks and may foster cognitive dependence. Organizationally, it improves collaboration and productivity while potentially constraining creativity. Societally, it reshapes labor markets and influences inequality, bias, ideology and misinformation. We argue that individuals should strengthen AI literacy and balance efficiency with critical thinking, while policies should promote cognitive autonomy and social equity.

Keywords: generative artificial intelligence; large language model; multifaceted impacts

JEL Classification: O33, J24, Z13

* Corresponding Author: ZHONG Yuelin, China Center for Economic Studies, Fudan University, No. 579 Guoquan Road, Yangpu District, Shanghai 200433, China; Tel: 86-13978454088; E-mail: 24110680030@m.fudan.edu.cn.