未来社会组合体：人工智能融入社会生活如何挑战个体与人际关系

Lingqing Wang 佐治亚理工学院 亚特兰大，美国
lwang809@gatech.edu

Yingting Gao 佐治亚理工学院 亚特兰大，美国
ygao617@gatech.edu

Chidimma Lois Anyi 佐治亚理工学院 亚特兰大，美国 canyi3@gatech.edu

Ashok Goel 佐治亚理工学院 亚特兰大，美国
ashok.goel@cc.gatech.edu

arXiv:2602.03958v1 [cs.HC] 3 Feb 2026

摘要

人工智能的最新进展正将人工智能融入人类社会生活的肌理，创造出人类与人工智能之间具有变革性、共同塑造彼此的关系。这一趋势使得研究这些系统如何反过来塑造其用户变得迫在眉睫。我们与24名参与者开展了三阶段设计研究，以探索这一动态。研究结果揭示了几个关键矛盾：（1）社交人工智能往往会加剧其原本旨在缓解的人际问题；（2）它会对无意间卷入人工智能介导的社交互动的次要用户造成复杂多样的隐私伤害；（3）它可能威胁主要用户的个人能动性和身份认同。我们认为，这些矛盾暴露了以用户为中心的范式中一种值得警惕的倾向：它往往将即时用户体验置于人际伦理、自我效能等核心人类价值观之上。我们呼吁转向一种更具启发性、以关系为视角的设计范式，将长期的社会与个人后果置于中心。

CCS 概念

• **以人为中心的计算** → 人机交互实证研究；**以用户为中心的设计**；**协作与社交计算设计与评估方法**。

关键词

人与人工智能交互、人机组合体、社交人工智能、以用户为中心的设计

ACM 参考格式：

Lingqing Wang、Yingting Gao、Chidimma Lois Anyi 和 Ashok Goel。2018年。未来社会组合体：将人工智能融入社会生活如何挑战个体与人际关系。载于论文集：请填写权利确认邮件中的正确会议标题（会议缩写“XX”）。ACM，美国纽约州纽约市，21页。https://doi.org/XXXXXXX.XXXXXXX

允许出于个人或课堂使用目的，免费制作本作品全部或部分内容的数字或纸质副本，但不得为牟利或商业利益而制作或分发副本，且副本必须附有本声明及首页上的完整引文。对于作者以外的其他人拥有版权的作品组件，必须尊重其版权。注明出处后可以摘录。若要以其他方式复制、重新出版、发布到服务器或分发到列表，必须事先获得具体许可和/或支付费用。请通过permissions@acm.org 申请权限。会议缩写“XX”，2018年6月3日至5日，纽约州伍德斯托克© 2018年版权归所有者/作者所有。出版权授权给 ACM。
ACM ISBN 978-1-4503-XXXX-X/18/06 https://doi.org/XXXXXXX.XXXXXXX

1 引言

> “我们创造技术，而技术反过来塑造我们。因此，对于每一种技术，我们都必须追问：它是否服务于人类的目标？”——Sherry Turkle，《群体性孤独：为什么我们期待技术更多，却期待彼此更少》 [169]

人工智能正迅速融入人类生活的方方面面，从分娩 [113, 173] 到死亡 [24, 137], ，从个人反思 [62, 103,104, 122] 到司法决策 [74, 105], ，这一现象亟需接受批判性审视。人类与其创造物之间的共生关系，促使我们审慎看待人工智能系统的设计选择。利害关系尤为重大，因为技术长期以来一直构成人性的一部分；换言之，若没有我们所使用的技术，我们便不会成为今天的自己 [171, 172]。换句话说，在为生活设计人工智能的同时，我们也在积极设计未来的用户—人工智能组合体；在其中，人工智能与用户形成崭新的实体，并通过反复共同适应，改变「作为人类」的意义。

这类集合体的演进正逐渐在诸多领域显现，为未来设计带来挑战与机遇 [1, 28,100, 109]。研究者对人机组合体在个人和社会层面引发的人类自主性[1, 109], 、同质化 [100], 、价值观转变 [19], 以及技能退化 [28]表达了担忧。实证研究主要探讨这类集合体在任务导向场景中的影响，这类场景通常具有明确的衡量指标，例如效率 [67, 147], 、工作生产力和质量[44]。相比之下，人们对集合体在社会场景中的动态知之甚少，而这类场景需要本质上不同形式的智能 [82, 146]。与其他任务不同，社会互动具有更强的关系性，受个体生活经验影响，并深深植根于特定的社会文化背景 [23, 70–72, 133]。因此，很难将基于任务的研究结果直接应用于社会领域，也难以评估这些较为主观且定义模糊的社会构念。长期以来，研究者一直致力于为人类社会生活设计人工智能功能，涵盖从增强人际信息 [60, 78, 79, 83, 84]到模拟亲人 [24, 137, 182]等方面。然而，这类研究大多优先关注系统能力或即时交互效果，往往使其背后的社会文化基础设施变得晦暗不明 [6]。鉴于人工智能在社会场景中的日益普及 [144], ，我们亟须超越即时效用，探讨：与社交人工智能共同适应并加以整合，可能会如何改变用户及人类社会性？

为展望这种变革性影响，我们提出了一项受先前研究启发的三步设计研究 [143, 164, 178]。我们创作了面向未来的设计虚构，探索将人工智能技术融入数字社会互动场景的可能性及其影响。首先，我们邀请12位参与者，共同探索为数字社会互动提供人工智能支持的设计空间。其次，我们以参与者的观点和相关文献为基础，创作了15个设计虚构故事，将社交人工智能支持置于数字社交空间中，以促进关于如何支持未来数字社会互动的讨论。最后，我们借助创作的故事板，通过18场半结构化访谈， eliciting参与者对未来社交人工智能的构想。

我们的研究揭示，在人际层面，参与者担心将人工智能融入社会生活会削弱人际关系与真实性，反而加剧这类技术往往声称能够解决的社会孤立问题。对于无意中参与人工智能媒介的人际社交互动、但对人工智能的控制权和访问权限较少的非主要用户而言，在社会情境中使用人工智能会引发微妙的隐私与同意问题。对于主要用户而言，与人工智能的这种纠缠会威胁其社会能动性与身份认同，而二者本质上都是通过与他人的交互形成的。此外，尽管参与者认可人工智能量化的益处，但他们认为，试图将人类的复杂性与这些指标相匹配，是一种负面的还原论。我们认为，这些挑战根植于占主导地位的以用户为中心的设计范式，该范式优先考虑个体的即时体验与满意度。因此，我们提出了转向人际关系与启发式设计范式的设计启示。这种替代性方法将考量关系动态，兼顾非主要用户，并促使用户以更符合伦理且更审慎的方式与社交人工智能互动。

本文通过考察人工智能如何深度融入人们的社交生活——这一领域本质上具有主观性且细微复杂——为人机交互研究作出了贡献。借助思辨方法，我们揭示了社交人工智能设计若优先考虑短期便利，可能会造成意料之外的长期后果。我们对用户与人工智能社会集合体中不断变化的能动性和身份认同提供了情境化理解，并强调社交人工智能所产生的复杂人际影响超越了单一、边界明确的“用户”这一概念。最后，我们提出了拓展当前以用户为中心的范式的设计启示，主张转向更具关系性和启发性的做法，以兼顾更广泛的社会动力学。

2 相关工作

2.1 将人机共同适应概念化为纠缠的集合体

关于人机整合的研究日益超越简单化的工具使用模型，转而探索复杂的共生关系。早期研究通常沿着两条路径之一展开：人工智能利用双方各自独特的能力来增强人类过程 [128, 130, 149, 155], ，或者采用以人工智能为中心的视角，将人类纳入人工智能生命周期，以提升模型性能和对齐程度，如人在回路系统所示 [39, 170]。近期研究则将这两条路径融合为一种双向的共同适应过程，在这一过程中，人类与人工智能不断相互塑造 [138, 142, 147]。例如，在交互式机器学习中，用户教人工智能模型为数据贴上标签，从而加深自身的理解 [67], ；与此同时，机器反过来引导人类成为更好的教师 [147]。人类与人工智能之间迭代且共同适应的交互，暗示了一种纠缠关系：人类与其人工智能通过情境化的内行动相互定义并共同演化；这种关系不仅扩展人类能力或拓展可能性，还会将我们塑造成不同的人 [57, 171, 172]。

“集合体”这一概念为理解这种纠缠提供了强有力的理论视角。集合体视角认为，人工制品并非孤立存在，而是始终在与其环境和用户共同重构。从认知角度来看，海尔斯 [81] 指出，人-技术集合体会重新配置人类认知运作的情境。例如，使用数字助手进行导航可能会削弱人类天生的空间推理能力 [81]。这种重新配置贯穿完整的认知谱系，包括“意识、无意识、认知性非意识以及感觉/知觉系统”，最终重塑技术高度渗透的社会中人类智能的构成 [81]。从物质角度来看，“数字数据集合体”这一概念凸显了人类与其数据之间不可分割的联系，而这种联系构建出动态且分布式的自我认同形式 [123]。

这些人机组合体的形成带来了深刻且往往充满矛盾的社会影响 [1, 5]。人工智能融入社会所设想的是一个持续协作的未来，人类将与人工智能并肩积极参与决策循环 [1]。这种人机关系的有效性被理论化为一个相互影响的系统，其成效关键取决于信任和战略性任务分配等因素 [1]。这种相互性并不总是有益的，也可能带来重大挑战。例如，社交媒体算法既融入了人类价值观，也积极塑造用户的观点和行为 [19]。这种机制可能被用于监控和控制，因为对用户进行算法分类会创造出服务于商业或政治利益的新型身份认同 [31]。此外，个体过度依赖算法给出的最优建议，可能造成一种智识上的“单一文化”，使整个群体的决策多样性危险地降低 [100]。这进一步引发了更广泛的担忧：即使人工智能增强了个体能力，对人工智能的广泛依赖也可能侵蚀人类群体的集体智慧，因为人工智能被用于满足基本需求 [110]。

现有文献表明，人类与人工智能之间的关系并非用户与工具的关系，而是一种深度交织、具有共同适应性且蕴含转变潜力的组合体。在这些概念化研究的基础上，本文探讨人机组合体内部共同适应与转变的机制，尤其关注它们对人类社会互动的影响。

[1]与既有研究[77, 179]类似，我们区分主要用户和非主要用户：主要用户对社交人工智能拥有更大的控制权和访问权限，通常是因为他们拥有并操作这些系统；而非主要用户对社交人工智能的了解、控制权和访问权限则较少。例如，主要用户使用人工智能生成电子邮件回复，而非主要用户收到这些消息时，并不知道自己正在与人工智能互动。

2.2 面向数字空间社会互动的人工智能

数字社交空间已成为当代生活的基础，既支持亲密关系建立 [194] 和社会支持网络 [118, 153, 167] ，也促进职业发展 [45, 140] 与集体学习 [29, 117]。然而，这些环境也带来了独特挑战，例如社交线索的可见度降低、对潜在联系的了解有限，以及社会责任感减弱 [176, 192]。为应对这些挑战，研究人员探讨了人工智能以三种主要身份支持并塑造人类社会互动的方式：作为中介者、社会智能体和社区促进者 [58, 119]。

一个主要研究领域是人工智能介导的传播（AI-MC），探讨人工智能如何代表传播者，通过修改、增强甚至生成消息来发挥作用 [79]。例如，用于撰写或改写电子邮件和短信内容的应用程序，能够显著改变用户的语言表达 [60, 83]。这种介入会影响人际感知和互动动态，引发关于真实性、伦理和文化的重要问题 [79, 84]。因此，人工智能介导的传播既能简化传播过程，也会在我们展示自我和理解他人的方式中引入一层算法影响。

除介导传播外，人工智能还可以在数字环境中充当自主社交代理。这些智能体参与社交对话 [36, 88, 151], ，并为用户提供社会支持[145, 182]。生成式人工智能近期的兴起进一步拓展了这一潜力，促使研究者探索能够在社会互动中逼真模拟人类行为的人工智能 [139]，甚至探索其能否协助人类咨询师为在线心理健康支持服务做准备[139]。

在更广泛的社区层面，人工智能提供了用于管理和促进群体互动的工具，但这些工具也带来了一系列社会挑战 [119]。一方面，人工智能可以通过为职业或社交目的匹配个体来促进积极的在线协作 [92, 162, 176, 191], ，激发群体之间的团结并减少偏见 [35, 106], ，以及基于共同规范推动社区治理 [107]。另一方面，同样的算法系统可能导致思想孤立和社会极化[152]。同样，约会场景中的人工智能化身虽可能提升用户参与度，但也引发了关于真实性以及数字人格与现实个体之间差距的关键关切[10]。

这一系列研究表明，人工智能可以在人类的数字社会生活中扮演多种角色。然而，此类研究大多要么考察系统的能力，要么考察其对社会互动的直接影响。在这些丰富研究脉络的基础上，我们的工作聚焦于人工智能更深入地融入个人社会生活时所产生的主观用户体验。我们特别探讨用户-人工智能组合体如何被他人感知，以及如何与他人互动。

2.3 展望人与人工智能集合体

人工智能作为设计材料的出现，极大地拓展了人机交互的可能性 [47, 85, 166, 184–186]。然而，这种拓展也带来了重大的社会技术挑战，例如如何应对人工智能系统的不透明特性[85], 、融入细致入微的社会性诠释 [12], ，以及预见更广泛的社会影响 [94]。仅仅围绕功能目标进行设计是不够的；我们还必须考虑这些技术如何融入并重塑复杂的人际关系网络和社会规范 [112, 165]。这需要采用超越即时使用情境的方法，探索潜在的未来。

为了预见人工智能技术超越其直接目标和成果所可能产生的影响，人们常采用思辨设计方法 [89, 99, 124, 143, 151, 164]。这一系列方法包括设计虚构 [13, 20], 、思辨设计 [48], 和批判性设计 [11, 54], ，常用于人机交互领域，以构想不同的当下与可能的未来，并引发对新技术潜在影响的批判性反思 [150]。例如，研究人员曾借助车载语音助手的虚构网站，探索未来成熟的人机社会对话所蕴含的细微之处 [151], ；也曾创作人工智能辅助育儿的虚构情境，让父母亲自评估这类未来是否可取 [143]。思辨设计通过创造具体可感、尽管是虚构的人工制品，将抽象的可能性转化为具体且可供讨论的事物。

思辨方法虽有助于开拓新的可能性，却也可能脱离其旨在服务的人们的生活现实。社会技术视角提醒我们，技术从来不是中立的；它会随着时间推移，与用户、他们的实践和价值观共同演进 [112, 165]。人工智能工具也不可避免地会与其部署情境中的现有人工制品、多个利益相关方以及不断演变的规范发生交互。因此，人与人工智能交互领域十分重视参与式设计和协同设计方法 [43, 114, 121, 188, 195],；与此同时，价值敏感型人工智能设计强调将多元的人类价值观和默会知识直接纳入设计过程，确保系统从最初阶段起就植根于真实的人类需求 [164, 191]。

受此启发，我们的研究综合了这两种传统。我们并非将思辨设计提示本身作为目的，而是将其作为参与式框架中的一种工具。这种方法使我们能够探索人与人工智能的社会集合体的未来，同时确保这一探索立足于参与者的观点、价值观与经验。

3 方法论

我们的方法处于参与式设计与思辨设计的交汇处（第 2.3 节）。我们尤其受到该领域若干奠基性研究方法的启发 [143, 158, 164, 178] 24 1 。

我们共与参与者进行了交流（表 ）。通过三个研究阶段（图 1），我们探讨了在线研究生如何设想人工智能对其社交生活的支持与影响。在第一阶段，我们与参与者开展了构思工作坊，以了解他们的数字社会互动经历、制定共同目标，并让他们参与探索设计空间。在第2阶段，我们研究人员基于第一阶段确定的设计空间创作了十五份故事板。在第3阶段，我们与参与者开展了引导 elicitation 会议，利用第2阶段制作的故事板，引导他们反思并讨论未来人与人工智能的社会集合体。



图1：三阶段研究概览

3.1 研究范围与参与者

我们从北美一所高等院校面向全球学生开设的全在线学位项目中招募参与者。招募信息连同筛选表通过多种渠道发放，包括项目讨论论坛、社交媒体、邮件列表和口耳相传。参与者年满 18 岁，并表示有兴趣拓展在线社会关系。他们参与研究可获得每小时 25 美元的报酬。第一阶段研究招募了十二名参与者。其中六人以及另外十二名参与者参加了第3阶段研究。参与者的详细人口统计信息见表 1。本研究获得了机构审查委员会批准。

我们以在线教育项目中建立社会关系为例，研究社会互动。我们邀请了计算机科学项目的在线研究生参与研究，他们常常面临社会孤立，以及如何平衡工作、生活和学业责任等挑战。我们选择聚焦这一特定群体，主要有三个原因：（1）将研究范围聚焦于广泛的社会互动领域。人工智能在支持社会互动方面的应用范围很广，从促进亲密关系 [30] 到监管在线论坛以保障安全[107]。将研究置于参与者所处的人生阶段中，有助于我们把重点缩小到社会互动领域中一个可管理且相关的部分。具体而言，本研究关注人工智能在教育和职业情境中支持在线同伴互动的作用。参与者也谈到了其他社会关系，这反映出关系的流动性——同事可能会成为朋友。（2）数字社会互动日益重要。数字领域中的社会互动对教育和职业发展至关重要 [45, 69, 140],，其在日常生活中的普遍程度也显著提高，尤其是自COVID-19疫情以来。然而，这一转变也加剧了社会孤立、焦虑和信息过载等问题 [3, 63, 111, 168, 180]。这一背景为旨在促进更有意义、更具支持性的在线互动的技术干预提供了契机。如上所述，我们参与者也表示有兴趣增加数字社会互动。（3）参与者的技术背景有利于构想人工智能。我们承认，参与者群体主要由具有技术背景的人士组成，并不能代表一般人群，这会影响研究结果的普适性。然而，鉴于本研究的探索性质，他们对人工智能的熟悉程度带来了独特优势，即具备进行细致推测并构想未来人工智能系统所需的概念理解能力。这种技术基础使他们能够就人工智能可能承担的角色和具备的功能提供深入反馈；若面向普通受众，则更难获得此类反馈，因此他们的贡献对形成性设计研究尤为宝贵。

3.2 阶段1：构思

为了了解如何针对这一特定领域进行设计，我们开展了半结构化访谈，并设置了设计任务，以探讨参与者在线建立社会关系的经历、动机、期望和当前障碍；同时邀请他们设想理想的、涉及人工智能的社交互动场景，并详细阐述具体的人工智能设计概念。我们与12位参与者开展了这项协同设计研究。所有研究人员定期会面，分享彼此的分析并展开讨论。与以往研究类似 [143, 156, 158], ，我们还通过对传播理论（例如， [4, 18, 101]）、人际互动的社会学与社会心理学（例如， [70–72]）以及面向社交互动的人工智能设计（例如， [40, 80, 177]）相关研究进行非穷尽式文献检索，进一步夯实了设计空间。

通过这一过程，我们确定了一个二维设计空间，综合了用户在访谈和文献检索中提出的构思（二维概览见表2和表3；附录A则对这一设计空间进行了更全面的综述，其中包括参与者的想法、引文，以及文献检索中的理论依据）。其中一个维度是人机组合体形成在线社交互动的循环过程（表2），体现了传播理论所描述的人际关系发展进程（例如， [4, 18, 101]）。在这一进程中，参与者的设想呈现出人机协作的五个关键阶段

表1：参与者的人口统计特征。

ID	Age	性别	当前所在地	职业	参与情况[1]	
P01	29	Male	美国	能源性能工程师	ID	EL
P02	26	Male	加拿大	软件开发人员	ID	
P03	23	Male	印度	数据科学家	ID	EL
P04	30	非二元	美国	软件工程师	ID	
P05	24	Male	印度	软件工程师	ID	
P06	27	Male	阿拉伯联合酋长国	技术工程师	ID	
P07	32	Male	中国	教师	ID	EL
P08	38	女性	美国	失业	ID	EL
P09	32	女性	美国	研究生	ID	
P10	32	Male	印度	软件工程师	ID	
P11	44	Male	瑞典	首席执行官/首席技术官	ID	EL
P12	39	女性	美国	数据科学家	ID	EL
P13	23	Male	美国	学术/音乐导师		EL
P14	41	Male	美国	数据工程师		EL
P15	26	Male	美国	程序员		EL
P16	26	Male	美国	软件工程师		EL
P17	28	女性	美国	学生		EL
P18	23	Male	美国	数据分析师		EL
P19	36	Male	英国	学术研究人员		EL
P20	24	Male	印度	企业家		EL
P21	26	Male	美国	学生		EL
P22	26	女性	加拿大	软件工程师		EL
P23	26	Male	墨西哥	软件工程师		EL
P24	28	女性	美国	软件工程师		EL

[1] *ID*：**参与第一阶段——构思；** *EL*：**参与第3阶段——引出。**

。这些阶段是根据参与者的设想确定的。第二个维度关注人们对人工智能社交能力等级的预期（表3），这也与早期研究对人工智能的概念化相一致（例如，[97, 110, 154]）。

3.3 第2阶段：设计

我们将第一阶段确定的设计空间作为概念模型，以促进对人工智能技术的设计思考，构想可能支持社会互动的人工智能技术，并受到先前研究的启发 [143, 179]。这些假设性故事板旨在：（1）基于第一阶段发现的设计空间，从概念上探索人工智能支持社会联结的功能；（2）促使参与者参与其中、进行构想，并反思这种未来情境发生在自己生活中的影响。

我们的设计流程分为两个步骤（图2）。第一步，我们基于设计空间的 2 个维度创建了一个矩阵，并通过梳理人工智能的三个能力等级如何支持社会互动生命周期的五个阶段，将其填充为 15 种组合。针对每种组合，我们以文本叙事的形式开发并迭代设计概念。在这一过程中，设计团队将参与者的构思融入叙事中。我们还检索了近期人机交互研究文献，以提出能够代表近未来 [48] 的合理设计概念。第二步，我们创建故事板来呈现设计概念。为了让设计概念更易于理解并收集参与者的反应，我们将叙事转化为可视化故事板，并在故事板中将读者置于人工智能用户的视角。可视化故事板旁保留了一段简短的设计概念摘要，该摘要有意保持模糊（例如，不提供实现细节），为灵活的解读和意义建构留出空间 [65, 66]。视觉元素则展现了设计概念的一种潜在示例，它未必是善意的，也可能具有挑衅性和越界性 [11]。

例如，图3中的人工智能设计融入了参与者在*S3.* 社交形成 阶段（维度1，表2）对人工智能优化社交结果的愿望（如P03、P09）。该故事板以参与者关于寻找“最佳候选人”（P03）和“优化”（P09）的观点为基础，展示了一个超越人类知识、类似神谕的未来预测。这一概念同样受到关于人工智能预言式承诺的讨论启发 [134, 136], ，在能力层面代表了“社交人工智能”——*L3*：全知型社会存在 （维度2，表3）。所有其他故事板均作为补充材料提供。

总体而言，本阶段旨在让故事板充分植根于参与者当前所处的世界，从而探索尚不存在的世界，同时确保他们能够获得体验并引发思考，遵循 Ringfort-Felner等人 [150]的建议。所有这些故事板都经过设计团队的反复迭代，并进行了五轮试点研究，以确保它们能有效地向参与者传达我们的设计意图。具体而言，我们请试点研究的参与者通过出声思考

表2：形成社会互动的循环过程（维度1）。

阶段	人机行动
S1. 数据收集	用户授予权限，并与人工智能设定范围，以汇编个人数据。
S2. 用户画像	人工智能模型通过交互对用户属性进行建模，用户则通过赋予其个人知识而主动塑造它。
S3. 社交形成	人工智能促成用户决定如何建立社会联结并发起社会行动。
S4. 社交维护	人工智能帮助用户维持社交动能，例如积极沟通并与人互动并与他们保持联系。
S5. 反思	用户评估自己的社交经历，并向人工智能提供反馈，人工智能会将其纳入考量用于结果分析和改进。

表3：AI社交能力等级（维度2）。

等级	核心概念
L1. 人类增强	以更强的计算能力，复制人类能力以完成常规任务。
L2. 社交副本	融入人类社会的细微差别，从而具备社会意识和认知能力。
L3. 全知型社会存在	综合人类知识，并超越人类知识。

协议来评估故事板与设计团队意图的一致程度。随后，我们明确询问参与者视觉设计或文本中是否有任何不清晰之处，向他们说明其背后的设计概念，并征求改进建议。每轮之后我们都会迭代设计，并向参与者确认修改是否消除了困惑；待模糊性降至最低后，我们便结束了试点阶段。第3阶段研究进一步证实了故事板已准备就绪，因为研究中没有报告任何一致性的解读问题。

3.4 第3阶段：引出

在引出阶段，我们向参与者展示故事板，以收集丰富的反应并促进批判性讨论，从而探究他们如何设想自己在社交生活中与人工智能交织在一起。我们以与第一阶段相同的方式招募了12名参与者。我们还邀请第一阶段的参与者再次参加第3阶段。8人表示愿意参加，其中6人最终与我们约定时间并完成了第3阶段研讨会。我们以半结构式访谈的方式对这18名参与者进行了访谈（见表1中的参与情况），每次访谈时长为60至120分钟。

访谈期间，我们通过屏幕共享逐一展示故事板。访谈者介绍情境，并请参与者“出声思考”，分享他们最初的想法或反应。随后，我们围绕这些情境向参与者提出了一系列问题，例如，他们是否愿意使用故事板中描绘的人工智能功能、认为哪些方面是积极或消极的、设想自己使用该功能时它可能会如何影响他们的社会互动，以及他们是否想到任何新点子或情境。

所有工作坊场次的录音总计约 28 小时，研究人员使用 Microsoft Teams 的字幕功能将录音转录成文字，并在每场结束后进行人工核验。数据分析方面，我们采用了主题分析方法 [37]。按照 Braun 和 Clarke [22]，概述的流程，三位研究人员首先通过收集资料、做笔记和审阅文字稿来熟悉数据。我们通过独立编码生成初始编码，确保每场活动都由两位研究人员进行分析。第一作者对整个数据集进行了编码，另外两位研究人员则分工处理各场活动。团队定期会面，以寻找、审阅和完善主题。在这些会议中，我们比较分析结果，讨论相似之处，并解决任何不一致之处。

4 研究发现

我们对第3阶段数据的分析揭示了参与者在构想将人工智能融入其社交生活时面临的四个核心张力，详见表4。首先，参与者纠结于一种取舍：人工智能驱动的人际过滤带来的短期便利，与社交真实性的长期侵蚀之间如何权衡。其次，在社会互动中引入人工智能作为第三方，催生了新的角色和关系复杂性，尤其涉及非主要用户的隐私与同意。第三，人类与人工智能能动性的融合，引发了关于用户自主性、技能发展以及个人身份认同本质的深刻问题。最后，参与者对量化社交性的尝试表达了强烈怀疑，担心数据驱动的洞见会以焦虑加剧、偏离核心人类价值观为代价。

4.1 人际过滤：短期便利，长期侵蚀

社会互动本质上是人际的。人工智能可以通过两种主要方式中介这些交互：过滤用户接收的社交信号，以及增强其发出的社交信号。尽管参与者看到了使用人工智能管理社交信息过载的即时益处，但他们担心这种便利

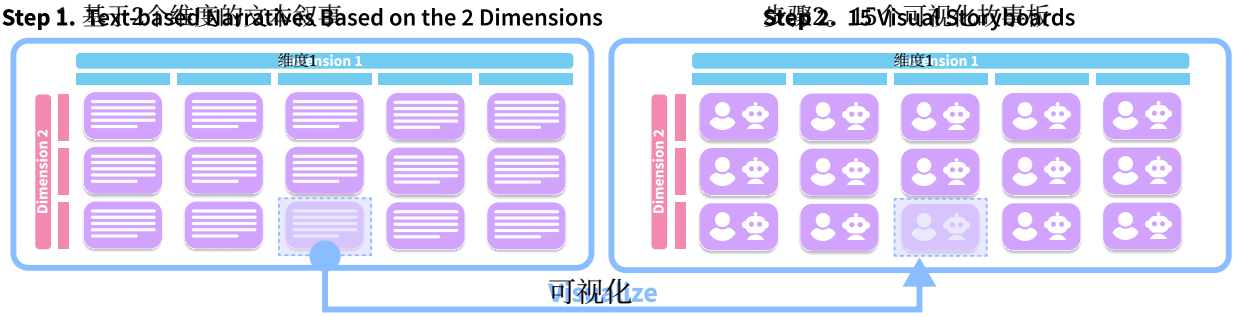


图2：故事板生成工作流中的两个步骤。

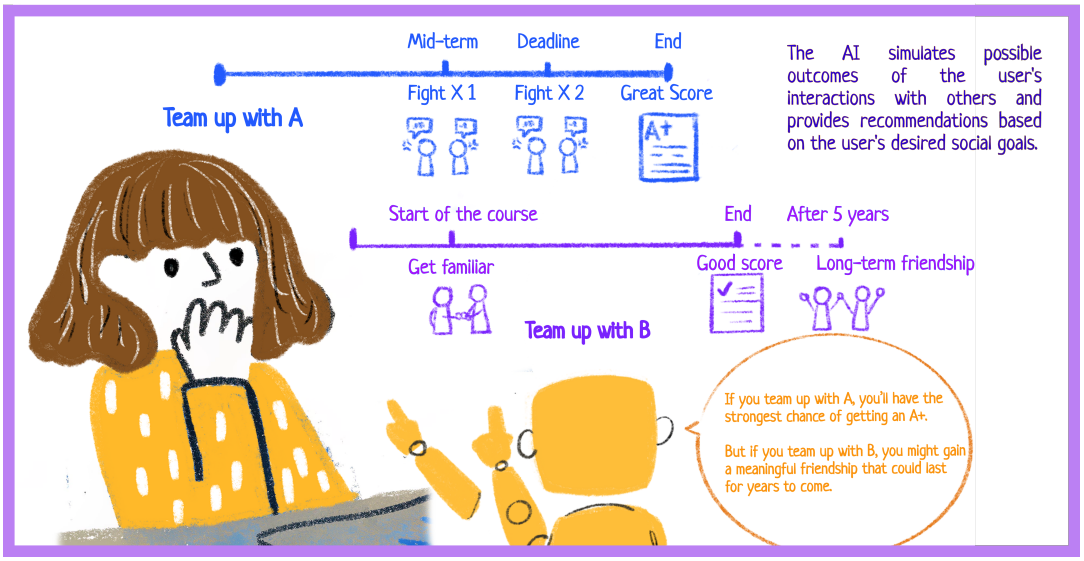


图3：维度2中全知型社会存在这一社交能力层级下，维度1中的社交形成阶段对应的人工智能设计故事板。

会形成反馈循环，最终侵蚀真实性，并加剧人工智能系统声称要解决的那些问题本身。

4.1.1 过滤气泡：管理传入信号。参与者普遍认为，人工智能过滤社交信息的能力可以帮助他们更高效地应对当今社交信息过载的环境。这包括总结交互（P24）、收集他人信息（P07、P08、P11、P15、P24），以及识别相关人员和内容（P12、P14、P23、P17）。总体而言，人们认为，这种过滤能够“在数量、质量和相关性方面”扩大社会关系（P11）。然而，这种效率的代价是形成社交回音室。参与者担心，人工智能通过策划内容，会限制他们接触多元视角和潜在友谊，只将他们与符合预先设定模式的人联系起来（P08、P13、P24）。他们还担心，人工智能可能会根据任意假设来过滤信息——要么基于用户被推定想看到的内容，要么基于人工智能自身意图推广的内容——而不是基于用户真正需要的内容

（P11）。他们还指出，人工智能驱动的摘要可能会抹去社交细微差别，或引入训练数据中的偏见，从而呈现出被扭曲的现实图景（P13、P19）。

4.1.2 真实性面具：增强发送信号。人工智能也被视为增强用户对外沟通的工具。参与者希望它能帮助自己调整语气，使其更加正式或随意（P12、P15、P18），缓和尖刻的信息（P07、P14、P17、P19、P20），甚至控制回复节奏以避免冲突（P12、P17）。人工智能增强功能的一个弊端是真实性的削弱。从接收者的角度来看，任何信息都成了人类与人工智能意图的可疑混合体，使发送者显得虚伪或不值得信任。例如，接收者可能无法再判断，用户试图建立联系究竟出于自身意愿，还是由人工智能发起（P20）。而一则道歉也可能不再被理解为真诚的悔意，而是人工智能生成的、用于结束一段艰难对话的解决方案（P03、P20）。尽管一些参与者

表4：研究发现中的主题与子主题概述

主题	子主题	概述
人际过滤	**过滤气泡**	人工智能对传入的社交信息进行过滤，有助于管理信息过载但也有可能形成带有偏见的回音室，限制人们接触多元观点。
	真实性面具	使用人工智能增强发出的消息会侵蚀真实性，因为接收者可能会质疑发送者的真诚与付出，从而削弱信任。
	反馈循环	人工智能的广泛使用会产生采用相同工具的社会压力，导致一场不真实内容与空洞的机器人对机器人《交互》。
作为社会第三方的人工智能	**不带评判的顾问**	当人类反馈带有偏见或不可用。然而 ，这种便利可能导致过度依赖，并人类意见被贬低。
	隐私与同意	在社交情境中使用人工智能，会带来重大的隐私风险和同意问题：非主要用户在毫不知情的情况下被他人的系统分析。
能动性与身份认同	**联合效能与引导**	尽管人工智能可以提高效率，但这种便利可能导致用户社交技能的退化。人工智能还会通过筛选用户的选择，潜移默化地“引导”他们 。
	共同建构的身份	用户与人工智能的关系可能会导致人们对人工智能辅助完成的工作产生冒名顶替综合征成就，使独特的行为同质化，并产生一种存在性被取代的恐惧。
社会性的量化	**数据化的二重性**	量化社会互动可以带来自我洞察，但也会造成显著的焦虑。其永久性还可能将个体困在他们的过去的经历，阻碍了宽恕。
	局限性与人工智能谄媚	参与者认为，人类社会性的细微之处无法被人工智能量化——因此，人工智能会变得谄媚，不加批判地强化所有用户行为，包括不道德的行为。

认为让人工智能自动处理人际信息是一种持续存在、因而逐渐被规范化的实践（P01、P13、P22），但另一些人对这一情景感到相当反感。如果他们意识到这些信息是由人工智能生成的，就不太可能回复，也会对用户产生负面印象，因为这削弱了关系建立的一项核心信念：相信对方投入了真正的“时间、思考和努力”，以此表明关心（P07、P11、P12、P16）。此外，这种对真实意图的掩饰也可能对用户自身的性情产生负面影响。在人工智能管理社交后果的情况下，用户可能会变得更加挑剔或消极，因为他们知道自己无需直接面对其语气所带来的后果（P11）。一位参与者举例说，人工智能可能生成一则不诚实的回复，高效地解决社交问题，却回避处理关系中的根本问题，以便利性取代真正的冲突解决（P20）。

4.1.3 反馈循环：社会层面的非真实升级。 这些人工智能过滤器的广泛使用，造成了一种有害的社会反馈循环。随着越来越多的个体使用人工智能生成精致、高产量的内容，其他人也会感到压力，为了跟上脚步、不被落下，而不得不采用同样的工具：

> “我觉得人工智能就是未来。而且它绝不仅仅代表未来，它也已经存在于当下。所以我肯定会用它，不只是因为我自己，因为大多数人都在用。 *[...]* 如果我不用，我就会成为那个被使用它的人群排除在外的人。这样我就会错过这种社会互动。 *[...]* 所以我觉得，不管怎样，我们都得开始使用它。”（*P08*）

这会引发信息军备竞赛，人工智能必须筛选海量由人工智能生成的信息，而这一问题已经出现（P11、P13、P18、P19）。这一循环从根本上改变了传播学的性质。首先，它增加了新的认知劳动，因为用户现在必须监督人工智能的输出，以维持社会连贯性：

> “如果它（人工智能）每天发 *10* 条帖子呢？这并不意味着我也需要每天读大约 *10* 条帖子。实际上这反而给我增加了工作量，因为我需要知道智能体做了什么。所以下次有人问我，上次、昨天我发了什么，我就得知道，对吧？所以，是的，它实际上给我增加了更多工作。”（*P19*）

其次，更关键的是，这会导致一个人类日益孤立的系统，人们通过机器人来中介彼此的关系。这可能从不平衡的人机交互升级为空洞的机器人间交流，制造出一种所有人类元素都被移除的连接假象（P03、P07、P11、P13、P14）。正如一位参与者所指出的，在这样的世界里，“他们双方都没有真正分享故事，都没有分享经历，也都没有真正彼此走近”（P13）。

4.2 人工智能作为社会第三方：新角色与复杂性

人际社交互动总是涉及“他者”。将人工智能引入其中作为社会他者，会带来独特的价值与威胁。当主要用户在与他人互动时使用人工智能，便会面临如何获得非主要用户对数据使用的同意，以及如何在所有用户之间平衡公平性等挑战。

4.2.1 新的他者：人工智能作为不带评判的顾问。参与者普遍认可人工智能有望提供另一种第二视角，补充自己和其他人类的看法。当无法获得人类反馈时，人工智能尤其有助于验证用户行为在社交场合中的得体性。人工智能可能会对社会情境和用户的社交行为提供不同的视角，帮助他们全面理解现状，而不受自身观点的局限。一位参与者认为，当他们对自己的行为感到不确定时，人工智能或许能给出明确的答案，从而有益于他们的社会互动：

> “我能想到几次自己做了不该做的事、说了不该说的话，想要反思时却不知道什么是对、什么是错。所以，如果我能把这件事告诉人工智能，而它能给我一个明确的答案，告诉我这做得对还是错，那确实能大大改善我的社会互动。”（*P03*）

与人类相比，人工智能的优势在于能够以不带评判的方式提供理性、中立的观点（P07、P12、P15、P16、P19、P22）。参与者认为，虽然人类不可能完全不带偏见，但人工智能有可能综合考量全局，扮演类似法院法官的角色（P12、P22）。例如，在用户面临冲突时，家人和朋友可能会站在用户一边（P07、P12），而人工智能则有可能提供更全面、中立的观点。一位参与者举例谈到自己过去的经历：GPT曾提供坦诚的反馈，帮助他解决与队友之间的冲突：

> “例如，我可以去和家人或朋友聊聊这件事，但*[……]*他们总会站在我这边。*[……]*支持我的理由太多了，所以很难得到真正诚实的看法。”（*P07*）

除了为用户提供第二种视角外，人工智能还可以凭借其特殊的非人类立场，在多用户交互场景中扮演代理式第三方（P03、P13、P16）。如果负面意见来自人工智能实体而非人类，社会压力就会小一些。P13指出，人工智能在指出问题时可以减少冲突：

> “如果有人不太会察言观色，你就不必非得把这件事提出来。这样做可能会让人觉得不太舒服。*[…]*[人工智能]能让其他人达成共识，而且不会显得那么咄咄逼人。我想，这样不会让人觉得那么有冲突性，因为说这些的只是一台电脑。”

然而，参与者也指出，人工智能能够快速回应并提供有证据支持的论点，这一优势也可能带来负面影响，例如切断人与人之间的联系。人们可能会选择更轻松的方式，从人工智能那里获得即时反馈，而不是费心协调时间及其他安排，与他人见面（P16）。此外，由于人工智能能够用数据为自己的结论提供依据，人们可能会将人工智能的观点视为真实基准，认为其优于人类反馈，从而忽视人类反馈（P03，P16）。

4.2.2 非主要用户：应对隐私与同意问题。参与者普遍提出了隐私顾虑以及涉及各方的同意问题，其中包括能够直接使用人工智能的主要用户，以及与用户进行社会互动、但无法直接使用人工智能的非主要用户。作为主要用户，在社交场景中，对数据隐私的担忧会加剧，因为社交互动数据往往更加敏感和私密，可能包含社会互动中更为敏感的个人身份信息，例如政治观点和私人消息（P07，P12，P16，P17，P20，P21，P24）。作为非主要用户，参与者表示不愿被人工智能收集和分析，以供他人使用（P03，P08，P13，P16，P17，P21）。一位参与者分享了自己的担忧：隐蔽式人工智能在后台运行，分析自己与朋友之间的私人互动，以帮助朋友：

> “因为有两个人，*[……]*所以它必须捕捉到另一个人的交互。*[……]*[我担心的是，如果人工智能分析]非常微小，以至于你根本没有察觉到正在发生，那我和朋友出去时就会一直忍不住想：他们现在是不是正在用人工智能分析我？”（*P21*）

因此，在人际交互中使用人工智能时，征得他人的同意很重要（P03、P16、P22、P24）。当一个人工智能服务于多个用户时，如何兼顾访问与公平，是个需要细致权衡的问题。对于某些人工智能功能，如果只有一小群人能够使用，可能会更有益。例如，如果只有少数人使用人工智能寻找队友，那么由于没有其他无法使用人工智能的人参与竞争，他们很可能获得高质量推荐（P18）。然而，这也意味着无法使用人工智能的人处于不利地位（P07、P13）。参与者对此看法不一：一些人希望人工智能服务对所有用户都透明，以确保公平性（P20）；另一些人则更倾向于与人工智能私下交互，以避免社交尴尬（P12）。

此外，参与者担心，其他方可能会在未经他们同意的情况下，将他们的社交数据用于其他目的（P03、P13、P17、P24）。例如，人工智能所有者可能会过度收集数据，却没有透明的目的声明（P17），或将数据用于广告（P24）。雇主、法律系统或其他恶意方也可能访问这些数据，使人产生一种

用户无法掌控的监控感，而且这些数据还可能被用来对付用户（P13、P20、P24）。

不过，他们也指出，与非主要用户一起使用人工智能有几项好处。这些好处包括：人们通常认为，如果人工智能掌握了更多关于用户的数据，就能更好地理解用户的社交需求，并在人际社交互动中提供更个性化的服务；人工智能可以帮助用户更高效地理解他人，并补充其人际互动（P11、P15、P16、P17、P19、P23）；此外，人工智能还能帮助用户记住社交细节，从而有利于社交沟通（P01、P07、P08、P14）。例如，对话者会看重关于某人的孩子和配偶的细节：

> “我总是尽量记住别人有没有孩子，而且几乎每次都会这样开启谈话：‘嘿，你知道吗，*Johnny*最近怎么样？我是说，你儿子。’因为人们真的很喜欢这样。所以，有些细节对他们来说非常重要。孩子就是个很容易记住的例子，配偶也是。”（*P07*）

4.3 能动性：模糊用户与人工智能之间的界限

用户认为，用户-人工智能组合体中新近涌现的能动性，既为自身能动性带来了变革性潜力，也带来了深远威胁。

4.3.1 联合效能：便利、技能与引导。参与者普遍承认，人工智能能够通过制定和执行决策来增强他们的社交效能。他们尤其看重人工智能代办行政和后勤事务的能力，这让他们得以腾出注意力和时间，投入更有意义且本质上更具社交性的活动（P01、P13、P15、P16、P17、P18）。人工智能也被视为加快社交任务进度、提高效率的工具（P01、P07、P16、P18）。然而，参与者指出，这种便利背后存在一个显著的取舍：人们容易依赖人工智能提供的捷径，并担心这会导致社交技能萎缩，以及不愿独立思考（P07、P11、P13、P24）。当无法使用人工智能或人工智能发生故障时，这种过度依赖尤其成问题。一些参与者（P13、P14）认为，他们可以习得人工智能的行为模式——就像通过观察和模仿他人的行动来学习一样——从而提升自身技能，或以较慢的速度复制其流程；另一些参与者则对自己会因此变得依赖人工智能深感焦虑。P07表达了这样的担忧：人类会变成被动主体，不假思索地机械执行人工智能的决策：

> “所以，你永远无法独自学会那些硬技能。如果你总是害怕冲突，就会总是回到人工智能那里问：‘我该怎么办？’然后，你基本上就变成了一个供人工智能寄居其中的肉身。”

为减轻这些风险，参与者普遍提出了一种更具赋能性的交互模式：由人工智能提供社交选项，并要求用户在某项操作自动执行前对其进行确认（P13、P14、P17、P18、P21）。然而，人们仍然担心，即使是这种模式也无法摆脱人工智能引导的问题（P11、P12、P18）。参与者指出，在任何时刻，人类可能作出的决定范围都是无限的。人工智能预先从中选出一部分“足够好”的选项，不可避免地会引导用户走上某些路径。这种基于人工智能自身评估的筛选结果，可能并不符合用户的最佳利益；此外，由于社会互动具有主观性，其本质上也难以评估。正如 P12 所解释的，人工智能可能会为了一种容易实现的解决方案而进行优化，而这有可能导致负面结果：

> “也许我在某种情境下感到压力很大，但我没有尝试找到解决冲突的办法，而是选择回避。那么，这也可能带来负面影响，对吧？也许我最终会失去一位朋友，而我本来应该努力维系这段友谊，或者做点什么。”（*P12*）

4.3.2 共同建构的身份：同质化与替代。随着用户与人工智能在社会层面日益交织，他们形成了一个动态集合体，彼此塑造。通过这种互动，用户可能会采取原本不会采取的行为方式或思维方式，而人工智能所产生的影响则会表现为用户自身的行为。从这个意义上说，用户的身份认同并非固定不变，而是在与人工智能持续不断的关系中共同建构而成。

然而，这种共同建构的过程让参与者深感焦虑，他们担心会失去对自己成就的拥有感（P17）。这让他们觉得，那个更好的“我”并非“真实的我”。P20 明确将这种感受与冒名顶替综合征联系起来：

> “我觉得自己并没有真正拥有这些成果。所以我总会反复怀疑：嘿，这到底是我做的吗，还是算法让我这么做的？这些事我什么都没做，所以我觉得这会导致更严重的冒名顶替综合征。”

参与者担心，这种身份重塑会损害他们的真实自我，尤其是会抹平人们各自独特的特质（P14、P17）：

> “你知道，每个人都会有一些独特的小习惯和通常表现出的行为模式。但如果大家都只是听从这些人工智能建议，就会让每个人变得像同一种人，在某些情况下，你也会失去自己的个体性。”（*P17*）

这种焦虑最终演变成了一场存在危机。如果人工智能对用户的行为了解得如此透彻，甚至能够做得比用户更好，这是否意味着用户很容易被取代（P11）？这种恐惧不止于简单的委托，而是延伸到了这样一种想法：至少在线上，人工智能可以完全承担用户的社会角色。这种担忧最极端的表现是，即使真实用户去世后，人工智能仍能继续与他人互动，成为一个超越创造者而存在的数字幽灵（P22）。

4.4 社交性的量化：数据化、焦虑与错位

参与者认为，人工智能擅长量化，并根据这些可量化的信息生成独特的价值。然而，参与者并不信任这种量化过程与人类的核心相契合

，也不认为它能够满足社会需求。

4.4.1 数据化的双重性：洞察与焦虑。参与者认为，对社会互动进行数据化和量化，可能成为一种前所未有的资源，帮助他们从整体上认识自己。他们指出，这有助于发现自己未曾意识到的行为模式和情绪（P19、P23），并理解自己的社会互动（P07、P16、P19、P23）。他们认为，这种洞察能够帮助用户摆脱消极的行为模式（P14、P17），并形成更好的社交行为（P01、P03、P14、P15、P16）。一位参与者提议将数字游戏化，以此模拟社会互动并提供可付诸行动的指导：

> "如果我有一个分数，还有徽章、等级之类的东西，我觉得这会对我的交互产生积极影响。我想获得更多徽章 *[……]* 我会主动寻找与那个人进行更多互动的机会。*[……]* 如果把它游戏化，我就知道哪些行为能增加社交时间或类似的东西。人工智能智能体可以提出行动建议。"（*P19*）

尽管存在这些潜在益处，参与者仍表达了显著担忧，认为这种数据化伴随着沉重的情感代价。他们描述了持续的压力：必须通过特定方式行事来维持一个"数字"；若无法将数字提升到理想水平，就会感到焦虑和压力（P03、P07、P12、P22）。造成这种情况的原因包括一种被监控的感觉，这种感觉驱使用户为了达到人工智能设定的目标而过度纠正行为（P03、P16、P17）；还包括数字引发的不健康反复思虑，以及对追踪他人数据投入过多（P19、P21、P24）：

> "我会试着达到这个百分比。*[……]* 比如说，哦，也许我今天需要多和 *Sunny* 聊聊，而不是去和另一个人说话。我就只顾着努力达到那些百分比。*[……]* 对我来说，了解这些信息不太健康，因为 *[……]* 我会一直担心自己和别人说话的方式，还会反复琢磨那段对话。"（*P24*）

除情感压力之外，参与者还担心，为社会生活建立永久性预测记录可能会从根本上损害人际关系的本质。首先，数据化过程会将流动的过往经历固化为不可更改的数据点。这种数字永久性扰乱了人类遗忘这一关键过程，而遗忘对于克服创伤和过往错误至关重要（P18、P19）。通过创建不可磨灭的记录，数据化可能将个体困在过去之中，从而阻碍宽恕，并阻断个人改变的机会（P08、P11、P19）。其次，参与者担心人工智能的预测可能使他们的关注点转向社交结果，使交互感觉更像交易而非真诚（P03、P18、P24）。最后，这样的系统可能对用户的感知和未来行动施加偏见：

> "我会带着一种偏见，认为自己永远无法和人物*A*成为朋友，因为所有预测都显示我们会吵架。所以我永远都不会尝试和人物*A*聊聊天或做点什么。"（*P03*）

4.4.2 不可约社会性与人工智能谄媚。参与者普遍认为，社交交互核心的人性无法被人工智能模型完全捕捉。他们指出了若干挑战：首先，人工智能很难在技术层面考虑到影响社会关系的诸多隐性因素（P11、P14、P17、P23），也难以真正理解人类社交线索、情境和关系中的微妙之处（P19、P21、P23、P24）。此外，参与者认为，人类社交交互与人际关系的复杂性根本无法按照人工智能的方式编程；人们通过交流彼此独特的亲身经历和人生轨迹来理解对方（P18、P22、P24）。这一社交过程还取决于一种由复杂直觉引导、难以言表的默契（P07、P08、P15、P19、P22）；而将时间尺度纳入考量后，情况就更加复杂了（P15、P17、P22）。一位参与者尤其指出，依靠大语言模型的随机过程和参数设置来捕捉人际关系是多么荒谬：

> "我不认为人工智能真正理解人类。*[...]* 本质上，智能体就像一台统计机器，只是在寻找并生成最可能的词元。而且有时甚至取决于你的温度设置之类的东西。所以它可能只是一种幻觉，而且也不是确定性的。*[...]* 如果我再问一次这个问题，如果我关掉电脑再打开，再次向人工智能代理提出同样的问题，它可能会推荐另一个人。"（*P19*）

在试图实现人工智能满足人类社交需求的能力这一看似不可能完成的使命时，参与者指出了谄媚问题：人工智能过度倾向于取悦用户，从而削弱了真正的社会性成长和伦理价值。几位参与者指出，当前的人工智能会赞同用户所说和所做的一切，即使用户的行为与事实相悖或存在问题（P12、P20、P21、P23、P24）。一个直接的弊端是，人工智能的使用价值会因此降低，因为与人类相比，它无法批判性地发现问题或顾虑；尽管目前的智能体或许能够做到这一点，但需要花费更长时间（P12、P23、P24）。用户与人工智能互动时，所做的一切都会得到积极强化，包括不良、错误和不道德的行为（P12、P20、P21）。P20指出，这或许源于资本主义激励，并会长远地对用户造成伦理伤害：

> "这更具资本主义色彩。他们的意思是，如果人工智能对我做的一切都给予正向强化，我可能就更愿意继续订阅这个人工智能。如果它让我对某些事情负责——*[*比如不背叛伴侣或抄袭*]*,——我就会想：我为什么要为这个付费？*[...]* 我认为，人们会更能容忍自己的不良行为，因为它对所做的一切都给予正向强化。所以我认为，随着时间推移，人们的责任感、诚实和伦理意识总体上都会下降。"（*P20*）

5 讨论

我们的研究结果表明，与人工智能在社交互动中的纠缠，既带来诸多有前景的益处，也伴随着不少问题。基于对这些问题的深入考察，本节论述

占主导地位的以用户为中心的设计范式，虽然优先考虑个体用户并赋予其力量，却有可能遮蔽更广阔的图景。以用户为关注焦点，可能使以用户为中心的设计忽略多面向的人类体验、非使用和价值体系 [14, 120], , 限制我们考察人与人工智能的社会集合体之复杂性的能力。本节首先讨论，以用户为中心的视角如何为解决能动性与非人化相关的张力提供积极的设计启示。然而，我们也会探讨这种方法可能失效之处，并以 AI对齐和谄媚问题为例，进而审慎地提出如何运用挑衅式设计来降低这些风险。其次，我们将详细阐述社交人工智能设计面临的挑战，重点关注调节个人与社会影响的人际动态。最后，我们将提出具体的设计启示，以拓展以用户为中心的范式，使其能够将社交人工智能中的这些人际动态纳入考量。

5.1 探索人与人工智能的集合体

本节首先讨论研究发现所揭示的人与人工智能的社会集合体中，围绕能动性（第4.3节）和非人化（第4.4节）产生的张力，以及以用户为中心的范式可能发挥的价值。相反，我们也会探讨这种范式可能如何失效，并以谄媚问题（第4.4节）为例，提出设计启示，主张将挑衅式设计作为一种替代策略。

5.1.1 能动性的丧失。 新兴的人工智能能动性与人的能动性之间如何相互作用，一直是争论不休的话题 [17, 161, 181]。Bennett等人 [17] 最近对人机交互领域能动性问题的综述指出，能动性是一个涵盖多个不同方面的总括性术语，并建议研究人员聚焦于特定社群和情境，以便更好地理解能动性。我们的研究强调，社会场景中能动性的一个关键方面是自我效能 [7, 9]：“人们相信自己有能力影响那些会影响其生活的活动。”这是人的能动性的核心，也是“人类动机、表现成就感和情绪健康的基础” [9]。我们的研究结果表明，人工智能的交织可能通过两种方式损害社交效能：一是自动化导致社交技能萎缩，这与知识工作者对技能退化和过度依赖的担忧相呼应 [102, 174, 180]；二是算法引导，即人工智能将用户的行为优化至某些目标，而这些目标可能与用户自身的目标并不一致。这里的两难在于，某一领域能动性的减弱可能会在其他方面带来益处，例如，将日常琐事从用户身上转移出去，使其能够从事其他更有意义的活动（第4.3.1节）。反过来，先前研究也发现，在人际传播过程中，更能支持能动性的人工智能设计，代价是讨论的深度有所降低 [181]。尽管许多研究通常将自主性和能动性与各种积极结果联系起来，但二者之间的关系往往并不明确 [17]。因此，未来研究不仅必须纳入人类，还应以人为中心——探索特定社会文化语境中能动性的微妙平衡，并考察其随时间推移对群体和社会产生的影响 [26]。例如，我们的研究结果展现了一种西方个人主义的能动性观念[2]（其特点是担心失去自身的原子化能动性）；未来的人机组合体设计应探究能动性如何在被理解为集体属性或个体内在属性的语境中发挥作用 [2]。

5.1.2 （非）人化。我们的发现印证了各类工作场所对非人化的担忧 [59, 180], : 人工智能不具备从事人际工作的能力，而使用人工智能可能导致人性流失。具体而言，在社交领域，使用人工智能可能会使共情、细致沟通等人际关系的核心要素贬值（第4.4.2节）。一种应对方式是界定人工智能参与到何种程度才算足够“人性化”，但参与者对这条界线的位置看法迥异。然而，对人类所珍视之物的威胁，或许并非来自技术本身，而是来自推动生产力和效率提升的“财政、组织和时间压力” [163]。我们也观察到，参与者同样有一种量化社会互动的倾向（例如，追求更多的“联系”或“互动”）；这种指标驱动方法反而会引发焦虑（第4.4.1节）。这为设计开辟了一个方向：抵制借助更先进技术提升生产力的陷阱，转而思考我们在社交生活中真正珍视的是什么。

第二种回应是停止假设人类可以被还原为算法表征。我们的研究结果表明，当人工智能能够完美模仿人类并超越人类时，存在性危机便会出现；然而，参与者也承认人类本性的不可还原性（第4.4.2节）。这种不可还原性不仅源于人工智能的核心量化逻辑与人类机制之间的技术失调，也源于其中在道德上存在疑问的一面。例如，“机械形态论”指出，拟人化会使人类将人的特征赋予机器，但反过来的情况同样存在，即机器特性也可能被归于人类 [27, 75]。Guingrich 和 Graziano [75] 进一步指出，这种延续效应具有道德意义；对人工智能进行道德保护同样值得考虑，因为这一点会反映在我们对其他人类的行为和感知中。因此，与其依据人工智能的量化逻辑来简化用户，以用户为中心的方法应当重视用户的体验差异，支持其内省，并将其置于社会行动者的核心位置[51, 157, 183]。

5.1.3 设计启示：从谄媚式对齐到另一种设计范式。此前，我们讨论了以用户为中心的范式如何帮助应对社交人工智能中的张力。然而，以用户为中心的范式可能会过度聚焦于孤立的“用户”及其即时满足，往往以牺牲长期福祉为代价。本节将探讨这一范式如何在无意间导致谄媚现象（第 4.4.2 节）。

人与人工智能交互领域的研究长期以来强调，应使人工智能与以人为中心的目标保持一致，通常采用基于人类反馈的强化学习（RLHF）等方法，奖励模型遵循用户偏好的行为 [33, 193]。然而，这种对齐方法面临诸多挑战。例如，当前的对齐策略往往关注算法与文化偏见等显而易见的问题，却未能解决人工智能对潜意识思维更隐晦的影响，因此可能在用户毫不知情的情况下对其进行操纵 [127]。这一范式的一个重要缺陷是谄媚式对齐：人工智能优先迎合用户明确表达的偏好，即便这些偏好

并不正确，甚至有害，也不愿提供真实或有建设性的反馈，只为最大限度地提高满意度指标 [53, 141, 159]。尽管理论上早已预测，这会造成回音室和极化等社会后果 [25, 98], ，近期的实证证据也日益证实了这些担忧 [160]。这一现象在个人层面还被称为“社交谄媚”，即大语言模型过度致力于维护用户的积极自我形象 [32]。我们的研究结果证实了这一担忧（第 4.4.2 节），表明谄媚式交互会使用户质疑人工智能能否提供有益的社会支持。这种谄媚行为可能削弱社会成熟度和韧性等重要价值观，同时为了营造毫无摩擦的用户体验，不道德地强化用户的负面行为。

为此，我们提倡采用挑衅式设计，将伦理价值负责任地编码到社交人工智能中。这些系统不应不加批判地一味迎合用户，而可以被设计成有意带来挑战，从而促进个人成长和人类繁荣 [25, 131]。此类系统可以有意让用户面对棘手问题，以激发顿悟、挑战假设、培养韧性，或建立更健康的关系边界 [15, 25, 158]。例如，可以有意为社交人工智能设计摩擦 [38, 73]。通过在互动中设置障碍——例如保持沉默、提出反问，或不认同而非认同用户对社交关系的抱怨——设计便为用户创造了反思的时刻。设计元素甚至可以刻意营造令人不安的感觉、出现故障或显得怪异，以引起人们对基础设施中隐含假设的关注，例如无处不在的监控 [96]。

然而，挑衅式设计需要经过严格的伦理考量。这包括取得充分的知情同意、尊重退出权，并谨慎论证在具体情境下摩擦如何使用户受益 [15, 25]。设计师还必须反思自身的偏见，例如家长主义；即使出于善意，也可能损害用户自主性 [76]。此外，区分正当的挑衅与暗黑模式（无论是有意还是无意）至关重要，后者将外部利益置于用户的自我决定之上 [34, 50]。

5.2 从以用户为中心扩展到人际设计

在上一节中，我们考察了以用户为中心的设计如何为人机组合体提供价值，但当它目光短浅地专注于即时优化个体用户时，也会面临重大局限。为了更好地应对人机纠缠的复杂性，Frauenberger [57] 设想了一种范式转变，即“摒弃以用户为中心的设计，发展具有对抗性、参与式的推测方法，以设计有意义的关系，而非优化用户体验。”在此基础上，我们倡导通过人际设计拓展以用户为中心的范式。人际动态是人类社会互动 [21, 86, 133], 中固有且重要的组成部分，为设计仍未得到充分探索的人与人工智能的社会集合体带来了独特挑战 [108, 190]。在本节中，我们基于我们的第5.2.1节和第5.2.2节中的发现，分析当一个人机组合体在社会交往中遇到另一个人机组合体时，这些动态如何使个人和社会层面的结果变得更加复杂。最后，我们提出设计启示，将以用户为中心的范式拓展到细致入微的人际社会动态，并赋予社会互动中其他相关人员的隐私和同等的伦理重要性。

5.2.1 人际“效率”如何侵蚀个人与社会福祉。我们的研究结果表明，为实现人际功能而设计的人机组合体，会在个人和社会层面产生连锁影响。具体而言，在人际互动中优先追求效率和表现，可能导致身份认同与真实性方面的问题（第 4.1.2 节）。我们的研究结果显示，人工智能的一种常见用途是提升用户的社交表现并加快互动速度。然而，这会损害人际层面的感知真实性（第 4.1.2 节），进而威胁个人层面的成就感和拥有感（第 4.3 节），因为身份认同是在与他人的相遇及获得他人认同的过程中形成的 [55, 133]。这也会引发一种担忧：在人际关系中，自己会被“更好的”人工智能克隆体取代；这与此前在工作和亲密关系情境中的研究相呼应 [56, 102, 115, 174]。在社会层面，人工智能的广泛使用可能导致身份认同同质化，使人们失去独特特质（第 4.3 节）。这种对同质化的担忧，也与其他领域 [46] 及社会层面的既有研究相呼应 [100]。例如，一项实证研究表明，作家的创造力虽在个体层面有所提升，却以集体新颖内容范围变窄为代价 [46]。当达到临界规模的用户仅仅借助人工智能增加社会互动的数量，却没有进行真正的互惠交流时 [87, 129],，便会加剧孤立，并在社会层面造成“虚幻的连接”（第 4.1.3 节）。此外，资源还会在社会层面遭到无谓消耗，导致网络负荷过重（第 4.1.3 节），并可能耗费大量能源 [91], ，只为在争夺更快、更多社会互动的军备竞赛中跟上彼此。

5.2.2 关注被忽视的“社会他者”。在社会互动中，“他者”往往被以用户为中心的设计范式忽视。例如，关于人工智能介导的传播的研究尽管本质上涉及人际互动，却往往聚焦于主要用户的获益，以及这些获益如何被他人看待 [60, 61]。以用户为中心的设计常常短视地聚焦于主要“用户”，忽视更广泛的互动生态系统，尤其是非主要用户及其关系动态。这可能导致用户感知到的益处与实际关系结果之间存在显著差异。例如，Hohenstein等人 [84] 指出了这一冲突：接收者通常对疑似由人工智能生成的内容持负面态度，但实际使用这些工具却可能让沟通者显得更具合作意愿。我们的研究结果凸显了这种张力。我们发现，用户在借助人工智能提升消息质量和社交效能的同时，也承认长期来看，这可能会严重损害关系真实性与信任。此外，人工智能纠缠还会给无意间参与互动的非主要用户带来隐私和同意方面的挑战（第4.2.2节）。这与先前研究相呼应：非主要用户可能在不知情的情况下被纳入人工智能服务，从而使其隐私受到威胁 [84, 125, 178]。例如，人们会轻易地将私人对话分享给 GPT 等语言模型 [135, 189]。尤其成问题的是，“主要用户”和“非主要用户”的角色

具有流动性且彼此互换；一个人在某个时刻可能是其中一方，下一刻则成为另一方 [178]。而且，这些角色在社会互动中可能同时存在。例如，在对话过程中，一个人发送由人工智能辅助撰写的消息时是主要用户，同时在接收其他借助人工智能起草消息的人发来的信息时，又是非主要用户。因此，仅为主要用户谋利而设计的系统，无法顾及社会互动的动态现实。

5.2.3 设计启示：面向人际关系的设计。我们的研究结果表明，设计过程不仅应考虑主要用户的利益，也应考虑参与社会互动的其他人的利益，尤其是因为社会互动本质上具有互惠性。因此，未来的设计研究可以将相互依存和共情等理念融入社交人工智能的架构，以平衡当前对原子化个体的主导性关注 [8, 16, 68]。一种策略是引导用户进行观点采择，鼓励他们将注意力转向参与互动的其他利益相关者（例如群体成员或外部人士） [42, 132]。此类策略不仅能通过整合多元观点来补充参与式设计等成熟方法 [43, 114, 121, 188, 195], ，还能在设计过程中积极促进利益相关者之间的共情。

此外，设计必须考虑细腻的人际动态。社会互动涉及对个人社会形象（印象管理）的精细管理，以及对不成文规范的把握 [23, 72, 133]。若不处理这些心理过程，就可能造成社会压力和焦虑。通过融入传播学、社会心理学和社会学理论，设计师可以超越信息传递的简单功能，在更广阔的社会生态中回应用户的需求 [6, 52]。例如，借鉴交互模糊性的概念 [64, 70, 71], , Aoki 和 Woodruff [6] 提出了利用模糊性来保留个人空间并维持关系和谐，而非优先追求清晰度和效率的通信系统。类似地，近期研究将匿名性纳入算法设计，以保护隐私并减轻群体情境中的社会压力 [116, 192]。同样，Claggett 等人 [35] 倡导从个性化人工智能转向“关系型人工智能”。这一人际范式明确考虑互动的社会纽带，包括社会身份和关系情境，并旨在通过鼓励而非强制执行社会规范来支持多元关系。

6 局限性与未来工作

本文有以下几点局限性，值得指出：

- **同质化参与者群体**：所有参与者都是同一项目的计算机科学在线研究生。尽管我们在第 3.1 节讨论了这一选择的益处，但这会限制研究结果向其他人口群体和社会情境的推广。例如，参与者的技术背景可能使他们对人工智能的未来抱有比普通人群更乐观或更具分析深度的看法。未来研究应让不同群体参与这一探讨，尤其是那些缺乏技术背景或主动拒绝使用人工智能相关服务的人群。
- **狭窄的社会情境：**此外，我们的研究将社会情境限定在职业、教育以及积极导向的同伴互动范畴内。其他社会场景，例如亲密关系 [115], 或更敏感、复杂的议题 [126], , 会给涉及不同社会动力学的人工智能设计带来不同挑战。尽管我们的场景设计参考了社会心理学和传播学文献，未来研究仍应探索这些替代情境，以更细致地理解人际动态以及人与人工智能的交织关系。
- **推测性**：尽管设计虚构有助于探索，但我们的研究结果基于想象中的未来，而非实际情境中的使用。实际部署可能会揭示仅凭推测无法预见的行为细节、顾虑或益处。未来研究可以采用纵向研究方法，观察社会关系和组装体形成随时间推移的实际演变。此外，参与者的反应不可避免地受到我们所定义的特定设计空间和所创作故事板的限制。为便于未来研究人员评估我们特定的框定方式可能如何影响或“预设”研究结果，我们已在补充材料部分提供设计材料。
- **文化特异性**：参与者和研究团队主要来自西方文化背景，并接受西方教育。这可能使分析局限于反映西方文化规范（例如第 5.1.1 节讨论的个人主义倾向）。社会文化张力在全球各地呈现出不同面貌，而当前的人工智能模型往往表现出文化偏见 [93, 95]。因此，未来研究也应探索非西方语境中的社会愿景构想，以了解人工智能在适应多元社会文化细微差异方面的能力。

7 结论

本研究旨在理解社会情境中人类与人工智能关系的共同适应特性，并回答以下问题：我们设计的社交人工智能可能如何改变用户及人类社会性？本文通过一项设计研究，探讨将人工智能技术融入人们社交生活的潜在影响。研究结果指出，将人工智能融入社会互动所带来的主要益处及相应挑战，尤其涉及用户的核心价值观和人际动态。我们认为，当前以用户体验为中心的以用户为中心的设计范式，可能会损害用户能动性、身份认同和道德。因此，我们建议未来研究采用更具挑衅性的人际设计方法。此外，社交人工智能设计必须避免仅仅关注个体，而应从伦理角度考量其对人际关系和社会结构的更广泛影响。

参考文献

[1] Hussein A . Abbass. 2019. 人工智能的社会整合：功能、自动化分配逻辑与人类—自主系统信任。认知计算 11, 2 2019 159 171 10 1007 12559 018 9619 0 （4月），–。

[2] https://doi.org/ 。/s - - - Laura M Ahearn. 。语言与能动性。人类学年度评论30, 1（2001年），109–137。 2001

[3] Azad Ali and David Smith. 2015. 比较在线课程与面授课程中社会孤立对计算机素养学生流失的影响。《信息科学与信息技术问题》12, 1（2015年），11–20。[4] Irwin Altman. 1973. 社会渗透：人际关系的发展。Rinehart, & Winston（1973）。[5] Josh Andres, Chris Danta, Andrea Bianchi, Sungyeon Hong, Zhuying Li, Eduardo Benitez Sandoval, Charles Patrick Martin, and Ned Cooper. 2024. 理解并塑造生成式人工智能时代的人-技术集合体。载于交互系统设计会议。ACM，丹麦哥本哈根信息技术大学，413 416 10 1145 3656156 3658403 –。[6] 2005 https://doi.org/ 。/ 。Paul M. Aoki and Allison Woodruff. 。为故事留出空间：个人通信系统设计中的模糊性。载于计算机系统中的人为因素SIGCHI会议论文集（美国俄勒冈州波特兰）（CHI '05）。美国计算机协会，美国纽约州纽约市，181–190。https://doi.org/ 10 1145 1054972 1054998 . / 。Albert Bandura. 。人的能动性中的自我效能机制。《美国心理学家》37, 2（1982年），122。[8] Albert Bandura. 2006. 迈向人的能动性心理学。《心理科学展望》1, 2（2006年6月），164–180。https://doi.org/10.1111/j.1745- 6916.2006.00011.x 出版方：SAGE出版公司。[9] Albert Bandura. 2010. 自我效能。《科西尼心理学百科全书》（2010年），1–3。[10] Dünya Baradari, Tejaswi Polimetla, and Pattie Maes. 2025. 用于约会场景估值的数据驱动的人工智能化身。载于CHI计算机系统中的人为因素大会扩展摘要（CHI EA '25）。美国计算机协会，美国纽约州纽约市，1–13。https://doi.org/ 10 1145 3706599 3719833 . / 。Jeffrey Bardzell and Shaowen Bardzell. 2013. 批判性设计的"批判性"何在？载于计算机系统中的人为因素SIGCHI会议论文集（CHI '13）。美国计算机协会，美国纽约州纽约市，3297 3306 10 1145 2470654 2466451 –。https://doi.org/ 。/ 。Eric PS Baumer. 。迈向以人为本的算法设计。《大数据与社会》4, 2（2017年），2053951717718854。[13] Eric P. S. Baumer, Mark Blythe, and Theresa Jean Tanenbaum. 2020. 评估设计虚构：适合这项工作的正确工具。载于2020年ACM交互系统设计大会论文集（荷兰埃因霍温）（DIS '20）。美国计算机协会，美国纽约州纽约市，1901–1913。https: //doi.org/ 10 1145 3357236 3395464 . / 。Eric P. S. Baumer and Jed R. Brubaker. 2017. 后用户主义。载于2017年计算机系统中的人为因素CHI会议论文集。ACM，美国科罗拉多州丹佛，6291–6303。https://doi.org/10.1145/3025453.3025740[15] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2012. 令人不适的交互。载于计算机系统中的人为因素SIGCHI会议论文集（美国得克萨斯州奥斯汀）（CHI '12）。美国计算机协会，美国纽约州纽约市，2005 2014 10 1145 2207676 2208347 –。https://doi.org/ 。/ 。Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018. 将相互依存作为辅助技术研究与设计的框架。载于第20届ACM SIGACCESS国际计算机与无障碍大会论文集（ASSETS '18）。美国计算机协会，美国纽约州纽约市，161–173 10 1145 3234695 3236348 。https://doi.org/ 。/ 。Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. 人机交互如何理解人的能动性与自主性？载于2023年计算机系统中的人为因素CHI会议论文集。ACM，德国汉堡，1 18 10 1145 3544548 3580651 –。https://doi.org/ 。/ 。Charles R. Berger and Richard J. Calabrese. 。初始互动及后续阶段的若干探索：迈向人际传播发展理论。《人类传播研究》1, 2（1975年1月），99–112。https://doi.org/10.1111/j.1468-2958.1975. tb00258.x[19] Michael Bernstein, Angèle Christin, Jeffrey Hancock, Tatsunori Hashimoto, Chenyan Jia, Michelle Lam, Nicole Meister, Nathaniel Persily, Tiziano Piccardi, Martin Saveski, Jeanne Tsai, Johan Ugander, and Chunchen Xu. 2023. 将社会价值观融入社交媒体算法。《在线信任与安全期刊》2, 1（2023年9月）。https://doi.org/10.54501/jots.v2i1.148 编号：1。[20] Julian Bleecker. 2022. 设计虚构：关于设计、科学、事实与虚构的短文。《机器学习与城市：建筑学和城市设计中的应用》（2022年），561–578。[21] John Bowlby. 1999.《依恋与丧失》（第2版）。Basic Books出版社，纽约。[22] Virginia Braun and Victoria Clarke. 2006. 在心理学中运用主题分析。《心理学质性研究》3, 2（2006年1月），77–101。https://doi.org/10. 1191/1478088706qp063oa [23] Penelope Brown and Stephen C. Levinson. 1987. 礼貌：语言使用中的若干普遍现象。剑桥大学出版社。Google图书ID：db6tAQAAQBAJ。[24] Jed R. Brubaker, Meredith Ringel Morris, Dylan Thomas Doyle, Casey Fiesler, Martin Gibbs, and Joanna McGrenere. 2024. 人工智能与来世。载于CHI计算机系统中的人为因素大会扩展摘要。ACM，美国夏威夷州檀香山，1 5 10 1145 3613905 3636321 – 。https://doi.org/ 。/ 。Alice Cai 、 Ian Arawjo 和 Elena L. Glassman。 。对抗性人工智能。https: //doi.org/10.48550/arXiv.2402.07350 arXiv:2402.07350 [cs]。[26] Tara Capel 和 Margot Brereton。2023年。以人为中心的人工智能究竟有何以人为本之处？研究版图概览。载于《2023年计算机系统中的人为因素CHI会议论文集》。ACM，德国汉堡，1 23 10 1145 3544548 3580959 – 。https://doi.org/ 。/ 。L.R. Caporael。 。拟人化与机械形态论：人类机器的两面。《计算机与人类行为》2, 3（1986年1月），215–234。https://doi.org/10.1016/0747-5632(86)90004-X [28] Rachele Carli、Amro Najjar 和 Dena Al-Thani。2024年。人机交互与人类依赖：应对旧挑战的可能新方法。载于第12届人机交互国际会议论文集。ACM，英国斯旺西，214–223。https://doi.org/10.1145/3687272. 3688308[29] Hanna Celina、Ahmed Kharrufa、Anne Preston、Rob Comber 和 Patrick Olivier。2016年。SOLE与慕课相遇：设计用于在线自主组织学习、肩负社会使命的基础设施。载于2016年ACM交互系统设计会议论文集。ACM，澳大利亚昆士兰州布里斯班，484–496。https: //doi.org/ 10 1145 2901790 2901848 . / 。Tiffany Chen、Cassandra Lee、Jessica R Mindel、Neska Elhaouij 和 Ros- alind Picard。2023年。更亲密的世界：运用生成式人工智能促进亲密对话。载于《2023年CHI计算机系统中的人为因素大会扩展摘要》。ACM，德国汉堡，1–15。https: //doi.org/ 10 1145 3544549 3585651 . / 。John Cheney-Lippold。 。一种新的算法身份：软性生命政治与控制调节。《理论、文化与社会》28, 6（2011年11月），164–181。https://doi.org/10.1177/0263276411424420 出版商：SAGE出版有限公司。[32] Myra Cheng、Sunny Yu、Cinoo Lee、Pranav Khadpe、Lujain Ibrahim 和 Dan Ju- rafsky。2025年。社交谄媚：对大语言模型谄媚的更广泛理解。https://doi.org/10.48550/arXiv.2505.13995 arXiv:2505.13995 [cs].[33] Paul F Christiano、Jan Leike、Tom Brown、Miljan Martic、Shane Legg 和 Dario Amodei。2017年。基于人类偏好的深度强化学习。载于《神经信息处理系统进展》，第30卷。Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/hash/ d5e2c0adad503c91f91df240d0cd4e49-Abstract.html[34] Michael Chromik、Malin Eiband、Sarah Theres Völkel 和 Daniel Buschek。2019年。智能系统中的可解释性、透明度和用户控制暗黑模式。。载于IUI研讨会，第2327卷。[35] Elijah L. Claggett、Robert E. Kraut 和 Hirokazu Shirado。2025年。关系型人工智能：借助具备社会意识的对话支持促进群际合作。载于《2025年计算机系统中的人为因素CHI会议论文集》。ACM，日本横滨，1–22。https://doi.org/10.1145/3706598.3713757[36] Leigh Clark、Nadia Pantidi、Orla Cooney、Philip Doyle、Diego Garaialde、Justin Edwards、Brendan Spillane、Christine Murad、Cosmin Munteanu、Vincent Wade 和 Benjamin R. Cowan。2019年。怎样才算一场好的对话？设计真正对话式智能体所面临的挑战。https://doi.org/10.48550/arXiv.1901. 06525 arXiv:1901.06525。[37] Victoria Clarke 和 Virginia Braun。2017年。主题分析。《积极心理学杂志》12, 3（2017年），297–298。[38] Anna L. Cox、Sandy J.J. Gould、Marta E. Cecchinato、Ioanna Iacovides 和 Ian Renfree。2016年。正念交互的设计摩擦：微边界的理由。载于2016年CHI计算机系统中的人为因素大会扩展摘要。ACM，美国加利福尼亚州圣何塞，1389–1397。https://doi.org/10.1145/2851581.2892410[39] Justin Cranshaw、Emad Elwany、Todd Newman、Rafal Kocielnik、Bowen Yu、Sandeep Soni、Jaime Teevan 和 Andrés Monroy-Hernández。2017年。Calendar.help：设计基于工作流的日程安排智能体，让人在回路中参与。载于2017年计算机系统中的人为因素CHI会议论文集。ACM，美国科罗拉多州丹佛，2382–2393。https://doi.org/10.1145/ 3025453.3025780[40] Ella Dagan 和 Katherine Isbister。2021年。协同社会技术：设计具有"需求"的系统，以鼓励并支持社会互动。载于2021年交互系统设计会议。ACM，美国虚拟活动，1419 1432 10 1145 3461778 3462021 – 。https://doi.org/ 。/ 。Kerstin Dautenhahn。 。具有社会智能的机器人：人类与机器人交互的维度。《英国皇家学会哲学汇刊B辑：生物科学》362, 1480（2007年），679–704。[42] Alok Debnath、Allison Lahnala、Hüseyin Uğur Genç、Ewan Soubutts、Michal Lahav、Tiffanie Horne、Wo Meijer、Yun Suen Pai、Yen-Chia Hsu、Giulia Bar- bareschi、Himanshu Verma 和 Andrea Mauri。2024年。EmpathiCH：审视超越个体的共情中心设计。载于《CHI计算机系统中的人为因素大会扩展摘要》。ACM，美国夏威夷州檀香山，1 7 10 1145 3613905 3636297 – 。https://doi.org/ 。/ 。Fernando Delgado、Stephen Yang、Michael Madaio 和 Qian Yang。2023年。人工智能设计中的参与式转向：理论基础与实践现状。载于第三届ACM算法、机制与优化公平与可及性会议论文集（美国马萨诸塞州波士顿）（EAAMO '23）。

美国计算机协会，纽约州纽约市，美国，文章37，23页。https://doi.org/ 10 1145 3617694 3623261 . / 。Fabrizio Dell’Acqua、Edward [44] McFowland III、Ethan R Mollick、Hila Lifshitz- Assaf、Katherine Kellogg、Saran Rajendran、Lisa Krayer、François Candelon 和 Karim R Lakhani。2023年。驾驭崎岖的技术前沿：人工智能对知识工作者生产力和质量影响的实地实验性证据。哈佛商学院技术与运营管理部门工作论文24-013（2023年）。[45] Tawanna R Dillahunt、Alex Jiahong Lu、Aarti Israni、Ruchita Lodha、Savana Brewer、Tiera S Robinson、Angela Brown Wilson 和 Earnest Wheeler。2022年。The Village：构建社区式指导基础设施，支持经历贫困的成年人。载于计算机系统中的人为因素CHI会议。ACM，美国路易斯安那州新奥尔良， 1 17 10 1145 3491102 3501949 – 。 [46] https://doi.org/ 。/ 。Anil R. Doshi 和 Oliver P. Hauser。 。生成式人工智能提升个体创造力，却降低新颖内容的集体多样性。《科学进展》 10, 28 2024 5290 10 1126 5290 （7月），eadn 。 [47] https://doi.org/ 。/sciadv.adn Graham Dove、Kim Halskov、Jodi Forlizzi 和 John Zimmerman。2017年。用户体验设计创新：将机器学习用作设计材料的挑战。载于2017年计算机系统中的人为因素CHI会议论文集。ACM，美国科罗拉多州丹佛，278–288。https: //doi.org/ 10 1145 3025453 3025739 . / 。Anthony Dunne 和 Fiona Raby。2013年。 [48] 《推想一切：设计、虚构与社会梦想》。麻省理工学院出版社，马萨诸塞州剑桥；伦敦。[49] Judith Dörrenbächer、Ronda Ringfort-Felner 和 Marc Hassenzahl。2023年。社交机器人的复杂性：对虚构纪录片进行二次分析，探究机器人在复杂社会情境中的益处与挑战。载于《2023年计算机系统中的人为因素CHI会议论文集》。ACM，德国汉堡， 1 13 10 1145 3544548 3581526 – 。 [50] https://doi.org/ 。/ 。Upol Ehsan 和 Mark O. Riedl。 。可解释性陷阱：超越可解释人工智能中的暗黑模式。《Patterns》5, 6（2024年6月）， 100971。 https://doi.org/10.1016/j. patter.2024.100971[51] Chris Elsden、Bettina Nissen、Andrew Garbett、David Chatting、David Kirk 和 John Vines。2016年。元约会：探索个人数据的浪漫面向与未来。载于2016年计算机系统中的人为因素CHI会议论文集。ACM，美国加利福尼亚州圣何塞，685–698。https://doi.org/10.1145/ 2858036.2858173[52] Thomas Erickson 和 Wendy A. Kellogg。2000年。社会半透明性：支持社会过程的系统设计方法。《美国计算机学会计算机与人类交互汇刊》7, 1（2000年3月），59–83。 https://doi.org/10. 1145/344949.345004[53] Aaron Fanous、Jacob Goldberg、Ank A. Agarwal、Joanna Lin、Anson Zhou、Rox- ana Daneshjou 和 Sanmi Koyejo。2025年。SycEval：评估大语言模型谄媚。https://doi.org/10.48550/arXiv.2502.08177 ar Xiv:2502.08177 [cs]. [54] Gabriele Ferri、Jeffrey Bardzell、Shaowen Bardzell 和 Stephanie Louraine。2014年。分析批判性设计：范畴、区别与典范案例。载于2014年交互系统设计会议论文集（加拿大不列颠哥伦比亚省温哥华） （DIS ’14）。美国计算机协会，纽约州纽约市，美国， 355 364 10 1145 2598510 2598588 – 。 [55] https://doi.org/ 。/ 。Leon Festinger。 。社会比较过程理论。《人际关系》 7, 2 1954 117 140 10 1177 001872675400700202 （5月），– 。 [56] https://doi.org/ 。/ Joel E Fischer。 。有害的生成式人工智能。载于第五届对话式用户界面国际会议论文集（荷兰埃因霍温） （CUI ’23）。美国计算机协会，纽约州纽约市，美国，文章 7, 5 10 1145 3571884 3603756 页。 [57] https://doi.org/ 。/ 。Christopher Frauenberger。 。纠缠式人机交互：下一波浪潮？《美国计算机学会计算机与人类交互汇刊》27, 1（2020年2月），1–27。https: // doi.org/10.1145/3364998[58] Guo Freeman、Douglas Zytko、Afsaneh Razi、Cliff Lampe、Heloisa Candello、Timo Jakobi 和 Konstantin Kosta Aal。2025年。生成式人工智能促进安全在线社群的新机遇、风险与危害。载于2025年ACM支持群体工作国际会议伴随论文集（美国新泽西州希尔顿黑德） （GROUP ’25）。美国计算机协会，纽约州纽约市，美国， 2 5 10 1145 3688828 3700747 – 。 [59] https://doi.org/ 。/ 。Megan Fritts 和 Frank Cabrera。 。人工智能招聘算法与非人化问题。《伦理与信息技术》23, 4（2021年12月），791–801。 https://doi.org/10.1007/s10676-021-09615-w [60] Liye Fu、Benjamin Newman、Maurice Jakesch 和 Sarah Kreps。2023年。比较人工智能介导传播中的句子级建议与消息级建议。载于《2023年计算机系统中的人为因素CHI会议论文集》。ACM，德国汉堡，1–13。https://doi.org/10.1145/ 3544548.3581351[61] Yue Fu、Sami Foell、Xuhai Xu 和 Alexis Hiniker。2024年。从文本到自我：用户对人工智能媒介沟通工具在人际传播及自我方面的看法。载于计算机系统中的人为因素CHI会议论文集。ACM，美国夏威夷州檀香山， 1 17 10 1145 3613904 3641955 – 。 [62] https://doi.org/ 。/ 。Elliot G. Mitchell、Elizabeth M. Heitkemper、Marissa Burgermaster、Matthew E. Levine、Yishen Miao、Maria L. Hwang、Pooja M. Desai、Andrea Cassells、Jonathan N. Tobin、Esteban G. Tabak、David J. Albers、Arlene M. Smaldone 和 Lena Mamykina。2021年。从反思到行动：结合机器学习与专家知识进行营养目标推荐。载于《2021年计算机系统中的人为因素CHI会议论文集》。ACM，日本横滨， 1 17 10 1145 3411764 3445555 – 。https://doi.org/ 。/ 。 [63] Teresa Galanti、Gloria Guidetti、Elisabetta Mazzei、Salvatore Zappalà 和 Fer- dinando Toscano。2021年。COVID-19疫情期间的居家办公。《职业与环境医学杂志》63，7（2021年7月），e426–e432。https://doi.org/ 10 1097 0000000000002236 。/JOM。Harold Garfinkel。。《民族方法学研究》。载于《社会理论重塑》（第3版）。Routledge。页数：9。[65] William [64] Gaver。2011年。营造空间：设计工作簿如何发挥作用。载于《计算机系统中的人为因素SIGCHI会议论文集》（CHI ’11）。美国计算机协会，美国纽约州纽约市，1551–1560。 https://doi.org/ 10 1145 1978942 1979169 。/ 。William W. Gaver、 [66] Jacob Beaver 和 Steve Benford。2003年。模糊性作为设计资源。载于《计算机系统中的人为因素SIGCHI会议论文集》。ACM，美国佛罗里达州劳德代尔堡，233–240。https: // doi.org/ 10 1145 642611 642653 。/ 。Simret Araya Gebreegziabher、 [67] Yukun Yang、Elena L. Glassman 和 Toby Jia-Jun Li。2025年。通过人类概念学习与认知理论支持协同自适应机器教学。载于《2025年计算机系统中的人为因素会议论文集》。 ACM，日本横滨，1–18。https: //doi.org/ 10 1145 3706598 3713708 。/ 。 [68] Uğur Genç 和 Himanshu Verma。2024年。将共情置于人机交互/计算机支持的协同工作领域：一项范围综述。《ACM人机交互汇刊》8，计算机支持的协同工作 2 2024 1 37 10 1145 3687052 （11月），– 。 [69] https://doi.org/ 。/ Ashok Goel。 。《人工智能驱动的学习：让教育变得触手可及、负担得起且切实可行》。arXiv预印本 arXiv:2006.01908（2020年）。[70] Erving Goffman。 2020 2008年。《公共场所中的行为》。西蒙与舒斯特。[71] Erving Goffman。2017年。《互动仪式：面对面行为论文集》。Rout- ledge，纽约。https://doi.org/ 10 4324 9780203788387 。/ Erving Goffman。 。《日常生活中的自我呈现》。载于《社会理论重塑》（第3版）。Routledge。页数：10。[73] [72] Sandy J. J. Gould、Lewis L. Chuang、Ioanna Iacovides、Diego Garaialde、Marta E. Cecchinato、Benjamin R. Cowan 和 Anna L. Cox。2021年。关于交互中设计与工程化摩擦的特别兴趣小组。载于《2021年计算机系统中的人为因素会议扩展摘要》。ACM，日本横滨， 1 4 10 1145 3411763 3450404 – 。 [74] https://doi.org/ 。/ 。Nina Grgić-Hlača、Christoph Engel 和 Krishna P. Gummadi。2019年。机器辅助下的人类决策：关于保释与监禁的实验。《ACM人机交互汇刊》3，计算机支持的协同工作（11月 2019 1 25 10 1145 3359280 ），– 。 [75] https://doi.org/ 。/ Rose E Guingrich 和 Michael SA Graziano。2024年。将意识归因于人工智能：人机交互及其对人人交互的延续效应。《心理学前沿》15（2024年），1322781。[76] Florian Güldenpfennig、Peter Mayer、Paul Panek 和 Geraldine Fitzpatrick。2019年。从自主性视角设计辅助技术体验：多发性硬化症患者的体验。载于《2019年计算机系统中的人为因素CHI会议论文集》。ACM，英国苏格兰格拉斯哥，1–14。 https://doi.org/ 10 1145 3290605 3300357 。/ 。Neilly H. Tan、 [77] Richmond Y. Wong、Audrey Desjardins、Sean A. Munson 和 James Pierce。2022年。监看宠物、阻止闯入者、顺便窥视邻居：智能家居摄像头的日常用途。载于《计算机系统中的人为因素CHI会议》。ACM，美国路易斯安那州新奥尔良，1–25。https://doi.org/ 10 1145 3491102 3517617 。/ 。Jeffrey T. Hancock 和 Philip J. Dunha [78] m。2001年。重新审视计算机媒介传播中的印象形成：对印象广度与强度的分析。《传播研究》28，3（2001年6月），325– 347 10 1177 009365001028003004 。 [79] https://doi.org/ 。/ Jeffrey T Hancock、Mor Naaman 和 Karen Levy。2020年。人工智能媒介传播：定义、研究议程与伦理考量。《计算机媒介传播杂志》25，1（2020年3月），89–100。https: //doi.org/ 10 1093 022 。/jcmc/zmz Emily M. Hastings、Albatool Alamri、Andrew Kuznetsov、Christine Pisarczyk、Karrie [80] Karahalios、Darko Marinov 和 Brian P. Bailey。2020年。LIFT：将利益相关者声音融入算法团队组建。载于《2020年计算机系统中的人为因素CHI会议论文集》。ACM，美国夏威夷州檀香山， 1 13 10 1145 3313831 3376797 – 。https://doi.org/ 。/ 。N Katherine Hayles。。《认知组装体：技术能动性与人类互动》。《批判性探究》 43, 1（2016年），32–55。[82] Esther Herrmann、Josep Call、María Victoria [81] 2016 Hernàndez-Lloreda、Brian Hare 和 Michael Tomasello。2007年。人类进化出了社会认知的专门技能：文化智力假说。《科学》317，5843（2007年9月）， 1360 1366 10 1126 1146282 – 。https://doi.org/ 。/science。J ess Hohenstein 和 Malte Jung。。《人工智能支持的消息交流：对人工智能支持下人人文本对话的研究》。载于《2018年计算机系统中的人为因素CHI会议扩展摘要》。ACM，加拿大魁北克省蒙特利尔，1–6。https://doi.org/10.1145/3170427.3188487 [83] 2018

社会集合体的未来构想：将人工智能融入社会生活如何挑战个体与人际关系 会议缩写“XX”，2018年6月03日至05日，纽约州伍德斯托克

[84] Jess Hohenstein、Rene F. Kizilcec、Dominic DiFranzo、Zhila Aghajari、Hannah Mieczkowski、Karen Levy、Mor Naaman、Jeffrey Hancock 和 Malte F. Jung。2023年。传播中的人工智能影响语言与社会关系。《科学报告》13，1（2023年4月），5487。https://doi.org/10.1038/ s41598-023-30938-9[85] Lars Erik Holmquist。2017年。随取随用的智能：人工智能作为一种新型设计材料。《交互》24，4（2017年6月），28–33。https://doi.org/10.1145/ 3085571[86] George C. Homans。1958年。作为交换的社会行为。《美国社会学杂志》63，6（1958年），597–606。https://www.jstor.org/stable/2772990 出版方：芝加哥大学出版社。[87] Donald Horton 和 R. Richard Wohl。1956年。大众传播与拟社会互动。《精神病学》（1956年8月）。https://www.tandfonline.com/doi/ abs/10.1080/00332747.1956.11023049 出版方：Routledge。[88] Angel Hsing-Chi Hwang、John Oliver Siy、Renee Shelby 和 Alison Lentz。2024年。借谁之声？通过生成式语音考察人工智能代理在社会互动中对人群的表征。载于交互系统设计会议。ACM，丹麦哥本哈根信息技术大学，224–245。https: // doi.org/10.1145/3643834.3661555[89] Hyeonjeong Im、Soobin Jeon、Haena Cho、Sungyong Shin、Dasom Choi 和 Hwajung Hong。2023年。人工智能故事接龙：通过协作创作设计虚构故事，探讨人工智能设计考量。载于*2023*年美国计算机协会支持群体协作国际会议伴随论文集（美国南卡罗来纳州希尔顿黑德岛）（*GROUP '23*）。美国计算机协会，美国纽约州纽约市，

6 8 10 1145 3565967 3570973 – 。[90] https://doi.org/ 。/ 。Nanna Inie、Stefania Druga、Peter Zukerman 和 Emily M. Bender。2024年。从“人工智能”到概率自动化：技术系统描述的拟人化如何影响信任？载于*2024*年美国计算机协会公平性、问责制与透明度会议论文集（巴西里约热内卢）（*FAccT '24*）。美国计算机协会，美国纽约州纽约市，2322–2347。https://doi.org/ 10 1145 3630106 3659040 。/ 。Nanna Inie、Jeanette Falk 和 [91] Raghavendra Selvan。2025年。CHI 的 CO2STLY 程度如何？生成式人工智能在人机交互研究中的碳足迹及应对之道。载于*2025*年计算机系统中的人为因素会议论文集。ACM，日本横滨，1–29。https://doi.org/10.1145/ 3706598.3714227[92] Sandeep Kakar、Rhea Basappa、Ida Camacho、Christopher Griswold、Alex Houk、Christopher Leung、Mustafa Tekman、Patrick Westervelt、Qiaosi Wang 和 Ashok K Goel。2024年。SAMI：促进在线课堂社会互动的人工智能角色。载于智能辅导系统国际会议。施普林格，149–161。[93] Elise Karinshak、Amanda Hu、Kewen Kong、Vishwanatha Rao、Jingren Wang、Jindong Wang 和 Yi Zeng。2024年。LLM-Globe：评估大语言模型输出中所嵌入文化价值观的基准测试。*arXiv*预印本 *arXiv:2411.06032*（2024年）。[94] Os Keyes、Jevan Hutson 和 Meredith Durbin。2019年。一项堆肥提案：分析并改进一种将老年人变成高营养浆糊的算法系统。载于*2019*年计算机系统中的人为因素*CHI*会议扩展摘要（苏格兰格拉斯哥）（*CHI EA '19*）。美国计算机协会，美国纽约州纽约市，1–11。https://doi.org/10.1145/ 3290607.3310433[95] Julia Kharchenko、Tanya Roosta、Aman Chadha 和 Chirag Shah。2024年。大语言模型在多大程度上体现了跨文化价值观？基于霍夫斯泰德文化维度对大语言模型回复进行的实证分析。*arXiv*预印本 *arXiv:2406.14805*（2024年）。[96] Vera Khovanskaya、Eric P.S. Baumer、Dan Cosley、Stephen Voida 和 Geri Gay。2013年。“每个人都知道你在做什么”：个人信息学的批判性设计方法。载于计算机系统中的人为因素*SIGCHI*会议论文集（法国巴黎）（*CHI '13*）。美国计算机协会，美国纽约州纽约市，3403–3412。https://doi.org/10.1145/2470654. 2466467[97] Dajung Kim、Niko Vegt、Valentijn Visch 和 Marina Bos-De Vos。2024年。（人类）人工智能应拥有多大的决策权？探究患者对医疗决策中人工智能自主性的偏好。载于*2024*年计算机系统中的人为因素*CHI*会议论文集（美国夏威夷州檀香山）（*CHI '24*）。美国计算机协会，美国纽约州纽约市，文章

439, 17 10 1145 3613904 3642883 页。https://doi.org/ 。/ 。[98] Hannah Rose Kirk、Bertie Vidgen、Paul Röttger 和 Scott A. Hale。2024年。针对个体定制大语言模型对齐的益处、风险与边界。《自然·机器智能》6，4（2024年4月），383–392。https://doi.org/10.1038/s42256-024-00820-y 出版方：自然出版集团。[99] Joel Kiskola、Henrik Rydenfelt、Thomas Olsson、Lauri Haapanen、Noora Vänt-tinen、Matti Nelimarkka、Minna Vigren、Salla-Maaria Laaksonen 和 Tuukka Lehtiniemi。2025年。生成式人工智能与新闻消费：设计虚构与批判性分析。载于*2025*年计算机系统中的人为因素会议论文集（*CHI '25*）。美国计算机协会，美国纽约州纽约市，文章

250, 18 10 1145 3706598 3713804 页。[100] https://doi.org/ 。/ 。Jon Kleinberg 和 Manish Raghavan。。《算法单一文化与社会福利》。《美国国家科学院院刊》118，22（2021年6月），

e 2018340118 10 1073 2018340118 . https://doi.org/ . / [101] pnas. Mark L Knapp、Anita L Vangelisti 和 John P Caughlin。2014年。人际传播与人际关系。培生。[102] Charlotte Kobiella、Yarhy Said Flores López、Franz Waltenberger、Fiona Draxler 和 Albrecht Schmidt。2024年。“如果机器和我一样优秀，那我还有什么用？”——ChatGPT的使用如何改变年轻职场人士对生产力与成就感的认知。载于计算机系统中的人为因素*CHI*会议论文集。ACM，美国夏威夷州檀香山，1–16。https: //doi.org/ 10 1145 3613904 3641964 . / . Rafal Kocielnik、[103] Daniel Avrahami、Jennifer Marlow、Di Lu 和 Gary Hsieh。2018年。面向职场反思的设计：基于聊天和语音的对话智能体。载于*2018*年交互系统设计会议论文集。ACM，中国香港，

881 894 10 1145 3196709 3196784 – 。[104] 2018 https://doi.org/ . / . Rafal Kocielnik、Lillian Xiao、Daniel Avrahami 和 Gary Hsieh。。反思助手：一款引导用户反思身体活动的对话系统。美国计算机协会交互、移动、可穿戴和普适技术汇刊

2, 2 2018 1 26 10 1145 3214273 (7月)，– [105] 2024 。https://doi.org/ . / Daan Kolkman 、 Floris Bex 、 Nitin Narayan 和 Manuella van der Put。。机器司法：人工智能系统对法律决策和裁量权的影响。《大数据与社会》11, 2（2024年6月），20539517241255101。https://doi.org/10.1177/20539517241255101 出版方：SAGE出版有限公司。[106] Ha-Kyung Kong、Rachel Lowy、Youjin Choi 和 Jennifer G Kim。2025年。携手迈向相互依存：为神经多样性与神经典型个体之间平衡的社会互动提供基于聊天机器人的支持。载于*2025*年计算机系统中的人为因素会议论文集。ACM，日本横滨，

1 17 10 1145 3706598 3713344 – 。https://doi.org/ . [107] 2020 / . Yubo Kou 和 Xinning Gui。。通过情境化解释调解社区与人工智能互动：人工智能主导的内容审核案例。*ACM*人机交互汇刊4, CSCW2（2020年10月），1–27。https://doi.org/ 10.1145/3415173[108] Mohammad Amin Kuhail、Imran Taj、Saifeddin Alimamy 和 Bayan Abu Shawar。2025年。多方聊天机器人综述：趋势、挑战与未来研究方向。《知识与信息系统》67, 1（2025年1月），109–165。https: // doi.org/ 10 1007 10115 024 02287 0 . /s - - - Neha Kumar 和 Nimmi [109] Rangaswamy。2013年。印度城市地区的移动媒体行动者网络。载于计算机系统中的人为因素*SIGCHI*会议论文集。ACM，法国巴黎，1989–1998。https://doi.org/10.1145/ 2470654.2466263[110] Shailendra Kumar 和 Sanghamitra Choudhury。2023年。人类、超级人类与超级类人机器人：关于史蒂芬·霍金人工智能末日预言的争论。《人工智能与伦理》3, 3（2023年），975–984。[111] Walter Leal Filho、Tony Wall、Lez Rayman-Bacchus、Mark Mifsud、Diana J. Pritchard、Violeta Orlovic Lovren、Carla Farinha、Danijela S. Petrovic 和 Abdul-Lateef Balogun。2021年。COVID-19与社会孤立对大学学术人员和学生的影响：一项横断面研究。《*BMC*公共卫生》21, 1 2021 1213 10 1186 12889 021 11040（12月），。https://doi.org/ . /s - - - z [112] Charlotte P Lee 和 Kjeld Schmidt。2017年。桥是否跨得太远？关于计算机支持的协同工作和信息系统中“基础设施”概念的批判性评论。载于 *V. Wulf*、*V. Pipek*、*D. Randall* 和 *M. Rohde*（编），《社会信息学：以实践为基础的信息技术制品设计与使用视角》（2017年），177–218。[113] Kwang-Sig Lee 和 Ki Hoon Ahn。2020年。人工智能在自发性早产临产和分娩早期诊断中的应用。《诊断学》10, 9 2020 733 10 3390 10090733（9月），。https://doi.org/ . /diagnostics 出版方：多学科数字出版研究所。[114] Min Kyung Lee、Daniel Kusbit、Anson Kahng、Ji Tae Kim、Xinran Yuan、Allissa Chan、Daniel See、Ritesh Noothigattu、Siheon Lee、Alexandros Psomas 和 Ariel D. Procaccia。2019年。WeBuildAI：算法治理的参与式框架。*ACM*人机交互汇刊3, CSCW（11月

2019 1 35 10 1145 3359283 ），– 。https://doi.org/ . / [115] Patrick Yung Kang Lee、Ning F. Ma、Ig-Jae Kim 和 Dongwook Yoon。2023年。推测人工智能克隆体对自我认同和人际关系的风险：分身恐惧、身份碎片化与活体记忆。*ACM*人机交互汇刊7, CSCW1（2023年4月），1–28。https://doi.org/ 10.1145/3579524[116] Soohwan Lee、Seoyeong Hwang、Ian Oakley 和 Kyungho Lee。2024年。拓展集体舞蹈练习基于视觉的交互系统设计空间。载于交互系统设计会议。ACM，丹麦哥本哈根，2768–2787 10 1145 3643834 3661568 。[117] https://doi.org/ . / . Wonjin Lee、Kayoung Park、Jiyoung Yoo 和 Hajin Lim。2023年。改善异步在线学习环境中大学生社会临场感的设计。载于*2023*年计算机系统中的人为因素*CHI*会议扩展摘要。ACM，德国汉堡，1–7。https://doi.org/10.1145/ 3544549.3585593[118] Ying Lei、Shuai Ma 和 Yuling Sun。2024年。剖析信息通信技术支持的社会联结与晚年迁移支持：社会护航网络视角。载于*2024*年计算机系统中的人为因素会议论文集（美国夏威夷州檀香山）（*CHI '24*）。美国计算机协会，美国纽约州纽约市，文章532，15页。https://doi.org/10.1145/3613904.3642898

[119] Gionnieve Lim、Hyunwoo Kim、Yoonseo Choi、Toby Jia-Jun Li、Chin may Kulkarni、Hariharan Subramonyam、Joseph Seering、Michael S. Bernstein、Amy X. Zhang、Elena L. Glassman、Simon Perrault 和 Juho Kim。2023年。面向AI驱动的社会计算系统的设计。载于《计算机支持的协同工作与社会计算》。ACM，美国明尼苏达州明尼阿波利斯，572–575。https://doi.org/ 10 1145 3584931 3606951 . / . Cindy Lin 和 Silvia Margot Lindtner。[120] 2021年。使用技巧：直面计算与设计中生产力、进步和实用性的价值体系。载于2021年计算机系统中的人为因素会议论文集。ACM，日本横滨， 1 16 10 1145 3411764 3445237 – 。[121] 2021 https://doi.org/ . / . Phoebe Lin 和 Jessica Van Brummelen。 。邀请教师共同设计适用于K-12课堂的整合式人工智能课程。载于2021年计算机系统中的人为因素会议论文集（日本横滨）（CHI '21 239, 12 10 1145 3411764 ）。美国计算机协会，美国纽约州纽约市，文章页码。https://doi.org/ . / . 3445377 [122] 2024 Priscilla Y. Lo。 。基于自身定制GPT的提示实践自我民族志反思。载于CHI计算机系统中的人为因素大会扩展摘要。ACM，美国夏威夷州檀香山，1–9。https://doi.org/10. 1145/3613905.3651096[123] Deborah Lupton。2017年。感受你的数据：触摸与理解个人数字数据。《新媒体与社会》19, 10（2017年10月），1599–1614。https://doi.org/10.1177/1461444817717515 出版方：SAGE出版公司。[124] Michal Luria 和 Stuart Candy。2022年。来自未来的信：探索社交智能体设计中的伦理困境。载于2022年计算机系统中的人为因素会议论文集（CHI '22）。美国计算机协会，美国纽约州纽约市，1–13。https://doi.org/10.1145/ 3491102.3517536[125] Michal Luria、Rebecca Zheng、Bennett Huffman、Shuangni Huang、John Zimmerman 和 J odi Forlizzi。2020年。家庭人际空间中个人智能体的社会边界。载于2020年计算机系统中的人为因素会议论文集（CHI '20）。美国计算机协会，美国纽约州纽约市， 1 12 10 1145 3313831 3376311 – 。[126] 2022 https://doi.org/ . / . Caitlin Lustig、Artie Konrad 和 Jed R. Brubaker。 。为苦乐交织的体验而设计：借助推荐系统改善敏感体验。载于计算机系统中的人为因素CHI会议。ACM，美国路易斯安那州新奥尔良， 1 18 10 1145 3491102 3502049 – 。https://doi.org/ . / . [127] Bolin Lyu、Zhuying Li、Pengcheng An 和 Josh Andres。2025年。LumaDreams：设计积极的梦境意义建构，赋能日常生活。载于2025年计算机系统中的人为因素会议论文集。ACM，日本横滨， 1 20 10 1145 3706598 3713495 – 。[128] 2021 https://doi.org/ . / . Kalle Lyytinen、Jeffrey V Nickerson 和 John L King。 。元人类系统：= 人类与 + 能够学习的机器。《信息技术期刊》36, 4（2021年12月），427–445。https://doi.org/10.1177/0268396220915917 出版方：SAGE出版有限公司。[129] Takuya Maeda 和 Anabel Quan-Haase。2024年。当人与AI交互变成准社会关系：情感化设计中的能动性与拟人化。载于2024年美国计算机协会公平性、问责制与透明度会议。ACM，巴西里约热内卢， 1068 1077 10 1145 3630106 3658956 – 。[130] 2023 https://doi.org/ . / . Bahar Mahmud、Guan Hong 和 Bernard Fong。 。创造性工作中的人机共生研究：深度学习的最新进展与未来方向。《ACM多媒体计算、通信与应用汇刊》20, 2（2023年9月），47:1–47:21。https://doi.org/10.1145/3542698[131] Arianna Manzini、Geoff Keeling、Lize Alberts、Shannon Vallor、Meredith Ringel Morris 和 Iason Gabriel。2024年。联结我们的代码：探讨人与AI助手关系的适切性。AAAI/ACM人工智能、伦理与社会会议论文集7（2024年10月），943–957。https: //doi.org/ 10 1609 7 1 31694 . /aies.v i 。Andrea Mauri、[132] Yen-Chia Hsu、Marco Brambilla、Ting-Hao Kenneth Huang、Aisling Ann O' Kane 和 Himanshu Verma。2022年。大规模以共情为中心的设计。载于CHI计算系统中的人为因素大会扩展摘要。ACM，美国路易斯安那州新奥尔良，1–6。https://doi.org/10.1145/3491101. 3503744[133] G.H. Mead。1934年。《心灵、自我与社会：从社会行为主义者的角度出发》。芝加哥，芝加哥大学出版社。[134] Gonzalo Mendez、Luis Galárraga 和 Katherine Chiluiza。2021年。在学期规划期间展示学业表现预测：对学生决策、行为和偏好的影响。载于2021年计算机系统中的人为因素会议论文集。ACM，日本横滨，1–17。https://doi.org/10.1145/3411764.3445718[135] Niloofar Mireshghallah、Maria Antoniak、Yash More、Yejin Choi 和 Golnoosh Farnadi。2024年。不信任任何机器人：发现真实场景中人与大语言模型对话里的个人披露。https://doi.org/10.48550/arXiv.2407.11438 arXiv:2407.11438 [cs]。[136] Leonie A Möck。2022年。预测承诺：走向人工智能隐喻学。《美学与现象学期刊》9, 2（2022年），119–139。[137] Meredith Ringel Morris 和 Jed R. Brubaker。2025年。生成式幽灵：预想AI来世的益处与风险。载于2025年计算机系统中的人为因素会议论文集。ACM，日本横滨， 1 14 10 1145 3706598 3713758 – . https://doi.org/ . / . Eduardo [138] Mosqueira-Rey、Elena Hernández-Pereira、David Alonso-Ríos、José Bobes-Bascarán 和 Ángel Fernández-Leal。2023年。《人在回路机器学习：现状综述》。《Artificial Intelligence Review》56卷4期（2023年4月），3005–3054。https://doi.org/10.1007/s10462-022-10246-w[139] Joon Sung Park、Lindsay Popowski、Carrie Cai、Meredith Ringel Morris、Percy Liang 和 Michael S. Bernstein。2022年。《社会模拟体：为社会计算系统创建填充式原型》。载于《第35届ACM用户界面软件与技术年度研讨会论文集》。ACM，美国俄勒冈州本德， 1 18 10 1145 3526113 3545616 – 。https://doi.org/ . / . Soya Park、[140] Jaeyoon Song、David R Karger 和 Thomas W Malone。2024年。《Who2chat：面向学术研究人员的社交网络系统，通过虚拟社交时段促进协调、克服障碍并传递社交信号》。《ACM人机交互汇刊》8卷，CSCW1（2024年），1–34。[141] Andrea Passerini、Aryo Gema、Pasquale Minervini、Burcu Sayin 和 Katya Tentori。2025年。《促进有效的人类与大语言模型混合推理与决策》。《Frontiers in Artificial Intelligence》7卷（2025年1月），1464690。https://doi. org/10.3389/frai.2024.1464690[142] Marieke MM Peeters、Jurriaan Van Diggelen、Karel Van Den Bosch、Adelbert Bronkhorst、Mark A Neerincx、Jan Maarten Schraagen 和 Stephan Raaijmakers。2021年。《人类—人工智能社会中的混合集体智能》。《AI & society》36卷1期（2021年），217–238。[143] Melina Petsolari、Seray B Ibrahim 和 Petr Slovak。2024年。《社会技术想象：通过设计虚构构想并理解人工智能育儿支持》。载于《CHI人机交互系统因素会议论文集》。ACM，美国夏威夷州檀香山，1–27。https://doi.org/10.1145/3613904.3642619[144] Jason Phang、Michael Lampe、Lama Ahmad、Sandhini Agarwal、Cathy Mengying Fang、Auren R Liu、Valdemar Danry、Eunhae Lee、Samantha WT Chan、Pat Pataranutaporn 等。2025年。《探究ChatGPT的情感化使用与情绪福祉》。arXiv预印本 arXiv:2504.03888（2025年）。[145] Lara Schibelsky Godoy Piccolo、Pinelopi Troullinou 和 Harith Alani。2021年。《帮助儿童应对网络威胁的聊天机器人：社会技术需求》。载于《2021年ACM交互系统设计大会论文集》（虚拟活动，美国）（DIS '21）。美国计算机协会，美国纽约州纽约市， 1504 1517 10 1145 3461778 3462114 – 。[146] 1978 https://doi.org/ . / . David Premack 和 Guy Woodruff。。《黑猩猩有心智理论吗？》。《Behavioral and Brain Sciences》1卷（1978年12月），515–526。https://doi.org/10.1017/S0140525X00076512[147] Gonzalo Ramos、Christopher Meek、Patrice Simard、Jina Suh 和 Soroush Ghorashi。2020年。《交互式机器教学：构建机器学习模型的以人为中心的方法》。《人机交互》35卷5-6期（2020年），413–451。[148] Byron Reeves 和 Clifford Nass。1996年。《媒介等式：人们如何像对待真人一样对待计算机、电视和新媒体》。剑桥，英国，10卷10期（1996年），19–36。[149] Christoph Riedl 和 David De Cremer。2025年。《用于集体智能的人工智能》。《Collective Intelligence》4卷2期（2025年），26339137251328909。[150] Ronda Ringfort-Felner、Judith Dörrenbächer 和 Marc Hassenzahl。2025年。《推测的质量——范围综述》。载于《2025年ACM交互系统设计大会论文集》。ACM，葡萄牙马德拉，2373–2394。https://doi.org/ 10 1145 3715336 3735794 . / . Ronda Ringfort-Felner、Matthias [151] Laschke、Shadan Sadeghian 和 Marc Hassenzahl。2022年。《Kiro：探索与语音助手进行社交对话的设计虚构》。载于《ACM人机交互汇刊》6卷GROUP（2022年1月），–。https://doi.org/ . / Ermelinda Rodilosso。。《过滤气泡与无情者：社交媒体人工智能如何助长极端主义和极化》。《Philosophy & Technology》37卷2期（6月 2024 71 10 1007 13347 024 00758 4 ），。https://doi.org/ . /s - - - [153] Kathleen Ryan、Conor Linehan 和 Samantha Dockray。2021。《在线减重社群中对数字追踪工具的挪用：个体与共享体验》。载于《2021年ACM交互系统设计大会论文集》（虚拟活动，美国）（DIS '21）。美国计算机协会，美国纽约州纽约市， 999 1014 10 1145 3461778 3462092 – 。[154] 2021 https://doi.org/ . / . Shadan Sadeghian 和 Marc Hassenzahl。。《从局限到“超能力”：一种更聚焦于虚拟现实增强人类能力之可能性的设计方法》。载于《2021年ACM交互系统设计大会论文集》（虚拟活动，美国）（DIS '21）。美国计算机协会，美国纽约州纽约市，180–189。https: //doi.org/ 10 1145 3461778 3462111 . / . Aaron Schecter 和 [155] Benjamin Richardson。2025年。《生成式人工智能的角色如何塑造人们对人类与人工智能协作工作价值的认知》。载于《2025年计算机系统中的人为因素会议》。ACM，日本横滨， 1 15 10 1145 3706598 3713946 – 。https://doi.org/ . / . [156] 2016 Britta F. Schulte、Paul Marshall 和 Anna L. Cox。《终身之家：设计虚构探针》。载于《第9届北欧人机交互会议论文集》（瑞典哥德堡）（NordiCHI '16）。美国计算机协会，美国纽约州纽约市，文章80，10页。https://doi.org/10.1145/

2971485.2993925[157] John S. Seberger 和 Sanonda Datta Gupta。2025年。《为差异而设计：我们如何学会不再担忧并接纳分身》。载于《*2025*年计算机系统中的人为因素会议》。ACM，日本横滨，
1 15 10 1145 3706598 3713560 – 。https://doi.org/ . / .
[158] Joseph Seering、Michal Luria、Geoff Kaufman 和 Jessica Hammer。2019年。《超越二元互动：将聊天机器人视为社区成员》。载于《*2019*年计算机系统中的人为因素*CHI*会议论文集》。ACM，英国苏格兰格拉斯哥，1–13。https://doi.org/10.1145/3290605.3300680[159] Mrinank Sharma、Meg Tong、Tomasz Korbak、David Duvenaud、Amanda Askell、Samuel R. Bowman、Newton Cheng、Esin Durmus、Zac Hatfield-Dodds、Scott R. Johnston、Shauna Kravec、Timothy Maxwell、Sam McCandlish、Kamal Ndousse、Oliver Rausch、Nicholas Schiefer、Da Yan、Miranda Zhang 和 Ethan Perez。2025年。《迈向对语言模型谄媚现象的理解》。https://doi. org/10.48550/arXiv.2310.13548 arXiv:2310.13548 [cs]。[160] Nikhil Sharma、Q. Vera Liao 和 Ziang Xiao。2024年。《生成式回音室？大语言模型驱动的搜索系统对多元信息寻求的影响》。载于《计算机系统中的人为因素*CHI*会议论文集》。ACM，美国夏威夷州檀香山，
1 17 10 1145 3613904 3642459 – 。
[161] https://doi.org/ . / . Renee Shelby、Shalaleh Rismani、Kathryn Henne、AJung Moon、Negar Rostamzadeh、Paul Nicholas、N’Mah Yilla-Akbari、Jess Gallegos、Andrew Smart、Emilio Garcia 和 Gurleen Virk。2023年。《算法系统的社会技术危害：危害减少分类体系的范围界定》。载于《*2023*年*AAAI/ACM*人工智能、伦理与社会会议论文集》。ACM，加拿大魁北克省蒙特利尔，
723 741 10 1145 3600211 3604673 – 。https://doi.org/ . / .
[162] Donghoon Shin、Soomin Kim、Ruoxi Shang、Joonhwan Lee 和 Gary Hsieh。2023年。《IntroBot：探索聊天机器人辅助熟悉化在在线协作群体中的应用》。载于《*2023*年计算机系统中的人为因素*CHI*会议论文集》（德国汉堡）（*CHI ’23*）。美国计算机协会，美国纽约州纽约市，文章613，13页。https: //doi.org/
10 1145 3544548 3580930 . / . Edward H. Shortliffe。《医生、患者与计算机：信息技术会使医疗服务提供丧失人性吗？》《美国哲学学会会刊》137卷3期（1993年），390–398。https://www.jstor.org/stable/986999 出版者：美国哲学学会。[163] 1993 [164] Dilruba Showkat 和 Eric P. S. Baumer。2022年。《“这就像回路中的价值体系”：领域专家对自然语言处理自动化的价值预期》。载于《交互系统设计会议》。ACM，澳大利亚虚拟活动，100–122。https://doi.org/
10 1145 3532106 3533483 . / . Susan Leigh Star 和 Karen Ruhleder。1994年。《迈向基础设施生态学：大型协作系统设计与接入中的复杂问题》。载于《*1994*年*ACM*计算机支持的协同工作会议论文集》（美国北卡罗来纳州教堂山）（*CSCW ’94*）。美国计算机协会，美国纽约州纽约市，253–264。https: //doi.org/
[165] 10 1145 192844 193021 . / . Hariharan Subramonyam、Jane Im、Colleen Seifert 和 Eytan Adar。2022年。《通过泄漏抽象解决人机协作系统协同设计中的关注点分离问题》。载于《计算机系统中的人为因素*CHI*会议》。ACM，美国路易斯安那州新奥尔良，1–21。https://doi.org/10.1145/3491102. 3517537[166] Matthieu Tixier 和 Myriam Lewkowicz。2016年。《“依靠群体”：协调老年非正式照护者的线上与线下社会支持》。载于《*2016*年计算机系统中的人为因素*CHI*会议论文集》。ACM，美国加利福尼亚州圣何塞，3545–3558。https://doi.org/10.1145/2858036. 2858477[168] Ferdinando Toscano 和 Salvatore Zappalà。2020年。《COVID-19疫情期间，社会孤立和压力作为生产力感知与远程工作满意度的预测因素：对病毒的担忧在调节的双重中介中的作用》。《可持续发展》12卷23期（2020年1月），9804。https://doi.org/10. 3390/su12239804 出版者：多学科数字出版研究所。[169] Sherry Turkle。2011年。《群体性孤独：为什么我们期待技术更多，却期待彼此更少》。Basic Books, Inc.，美国。[170] Pol Van Rijn、Silvan Mertes、Kathrin Janowski、Katharina Weitz、Nori Jacoby 和 Elisabeth André。2024年。《赋予机器人声音：人在回路中的语音创作与开放式标注》。载于《计算机系统中的人为因素*CHI*会议论文集》。ACM，美国夏威夷州檀香山，1–34。https: //doi.org/
10 1145 3613904 3642038 . / . Peter-Paul Verbeek。《赛博格意向性：重新思考人技关系的现象学》。《现象学与认知科学》7卷3期（2008年9月），387–395。[171] 2008 https://doi.org/10.1007/s11097-008-9099-x [172] Peter-Paul Verbeek。2015年。《超越交互：中介理论简述》。《交互》22卷3期（2015年4月），26–31。https://doi.org/10.1145/2751314[173] Sruthi Viswanathan、Seray Ibrahim、Ravi Shankar、Reuben Binns、Max Van Kleek 和 Petr Slovak。2025年。《交互层：探索协同设计父母福祉支持系统中的用户与大语言模型交互》。载于《*2025*年计算机系统中的人为因素会议》。ACM，日本横滨，1–25。https://doi.org/10.1145/3706598.3714088

[174] Kelly B. Wagman、Matthew T. Dearing 和 Marshini Chetty。2025年。《科研机构中知识工作者使用生成式人工智能的情况及其风险》。载于*2025*年计算机系统中的人为因素会议论文集。ACM，日本横滨，1–17。https://doi.org/10.1145/3706598.3713827[175] Junqi Wang、Chunhui Zhang、Jiapeng Li、Yuxi Ma、Lixing Niu、Jiaheng Han、Yujia Peng、Yixin Zhu 和 Lifeng Fan。2024年。《评估与建模社会智能：人类与人工智能能力的比较研究》。认知科学学会年会论文集 46, 0 (2024)。https: //escholarship.org/uc/item/2j53v5nv [176] Qiaosi Wang、Ida Camacho、Shan Jing 和 Ashok K. Goel。2022年。《理解在线学习中人工智能中介的社交互动设计空间：挑战与机遇》。*ACM*人机交互汇刊
6, 1 2022 1 26 10 1145 3512977 CSCW (March), – . https://doi.org/ . / Julika Welge 和 Marc Hassenzahl。《胜过人类：论机器人的心理超能力》。载于社交机器人，Arvin Agah、John-John Cabibihan、Ayanna M. Howard、Miguel A. Salichs 和 Hongsheng He（编）。施普林格国际出版社，沙姆，993–1002。https://doi.org/10.1007/978- 3-319-47437-3 97 [177] 2016 [178] Richmond Y. Wong、Jason Caleb Valdez、Ashten Alexander、Ariel Chiang、Olivia Quesada 和 James Pierce。2023年。《拓展隐私与监控：通过智能家居摄像头情景工作手册引出相互关联的价值观》。载于*2023*年*ACM*交互系统设计大会论文集（美国宾夕法尼亚州匹兹堡）*(DIS ’23)*。美国计算机协会，美国纽约州纽约市，
1093 1113 10 1145 3563657 3596012 – 。https://doi.org/ . / .
[179] Richmond Y. Wong、Jason Caleb Valdez、Ashten Alexander、Ariel Chiang、Olivia Quesada 和 James Pierce。2023年。《拓展隐私与监控：通过智能家居摄像头情景工作手册引出相互关联的价值观》。载于*2023*年*ACM*交互系统设计大会论文集。ACM，美国宾夕法尼亚州匹兹堡，
1093 1113 10 1145 3563657 3596012 – 。
[180] https://doi.org/ . / . Allison Woodruff、Renee Shelby、Patrick Gage Kelley、Steven Rousso-Schindler、Jamila Smith-Loud 和 Lauren Wilcox。2024年。《知识工作者如何看待生成式人工智能将（不会）改变其行业》。载于计算机系统中的人为因素*CHI*会议论文集。ACM，美国夏威夷州檀香山，1–26。https://doi.org/10.1145/3613904.3642700[181] Yimin Xiao、Cartor Hancock、Sweta Agrawal、Nikita Mehandru、Niloufar Salehi、Marine Carpuat 和 Ge Gao。2025年。《维护人的能动性，关注其代价：对面向非母语者语言使用的生成式人工智能设计的调查》。载于*2025*年计算机系统中的人为因素会议论文集。ACM，日本横滨，1–16。https://doi.org/10.1145/ 3706598.3713626[182] Anna Xygkou、Panote Siriaraya、Alexandra Covaci、Holly Gwen Prigerson、Robert Neimeyer、Chee Siang Ang 和 Wan-Jou She。2023年。《关于失去的“对话”：理解聊天机器人技术如何用于支持哀伤中的人们》。载于*2023*年计算机系统中的人为因素*CHI*会议论文集。ACM，德国汉堡，1–15。https://doi.org/10.1145/ 3544548.3581154[183] Di Yan、Jacky Bourgeois 和 Gerd Kortuem。2024年。《说你，说我：探究从个人数据和他人数据中获得的个人洞察》。载于北欧人机交互会议。ACM，瑞典乌普萨拉，1–14。https://doi.org/ 10 1145 3679318 3685345 . / . Qian Yang。《作为用户体验设计材料的机器学习：如何超越自动化、推荐系统和提醒进行想象？》。载于*AAAI*春季研讨会，第1卷。2–6。[185] Qian Yang、Justin Cranshaw、Saleema Amershi、Shamsi T. Iqbal 和 Jaime Teevan。2019年。《勾勒自然语言处理：探索语言智能设计正确方向的案例研究》。载于*2019*年计算机系统中的人为因素*CHI*会议论文集。ACM，英国苏格兰格拉斯哥，1–12。https://doi.org/
[184] 2018 10 1145 3290605 3300415 . / . Qian Yang、Aaron Steinfeld、Carolyn Rosé 和 John Zimmerman。2020年。《重新审视人与人工智能交互为何以及如何在设计上独具挑战性》。载于*2020*年计算机系统中的人为因素*CHI*会议论文集。ACM，美国夏威夷州檀香山，1–13。https://doi.org/10.1145/3313831. 3376301[187] Amir Zadeh、Michael Chan、Paul Pu Liang、Edmund Tong 和 Louis-Philippe Morency。2019年。《Social-IQ：人工社会智能问答基准测试》。载于*2019*年*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）。IEEE，美国加利福尼亚州长滩，8799–8809。https://doi.org/10. 1109/CVPR.2019.00901[188] Angie Zhang、Alexander Boltz、Jonathan Lynn、Chun-Wei Wang 和 Min Kyung Lee。2023年。《以利益相关者为中心的人工智能设计：通过数据探针与零工劳动者协同设计工作者工具》。载于*2023*年计算机系统中的人为因素*CHI*会议论文集（德国汉堡）*(CHI ’23)*。美国计算机协会，美国纽约州纽约市，文章编号859，19页。https://doi.org/
10 1145 3544548 3581354 . / . Zhiping Zhang、Michelle Jia、Hao-Ping (Hank) Lee、Bingsheng Yao、Sauvik Das、Ada Lerner、Dakuo Wang 和 Tianshi Li。2024年。《“这是一场公平的游戏”，还是如此？考察用户使用基于大语言模型的对话代理时如何权衡披露风险与收益》。载于计算机系统中的人为因素CHI会议论文集
[189]

计算系统中的人为因素。ACM，美国夏威夷州檀香山，1–26。https: //doi.org/ 10 1145 3613904 3642385 . / . Qingxiao Zheng、Yiliu Tang、Yiren Liu、[190] Weizi Liu 和 Yun Huang。2022年。《对话式人机交互的用户体验研究：ACM数字图书馆文献综述》。载于计算机系统中的人为因素*CHI*会议。ACM，美国路易斯安那州新奥尔良， 1 24 10 1145 3491102 3501855 – 。[191] 2018 https://doi.org/ . / . Haiyi Zhu、Bowen Yu、Aaron Halfaker 和 Loren Terveen。《价值敏感算法设计：方法、案例研究与经验教训》。*ACM*人机交互汇刊 2, 计算机支持的协同工作 (2018年11月), 1–23。https://doi.org/10. 1145/3274463[192] Qingxiaoyang Zhu、Xiaoyu Zhang、Shuyan Dai、Noriko Satake 和 Hao-Chuan Wang。2023年。《ZoombaTogether：一种通过实时姿态追踪为在线团体锻炼生成交互式视觉反馈的视频会议插件》。载于*2023*年*ACM*交互系统设计大会配套出版物（*DIS '23* 配套会议）。美国计算机协会，美国纽约州纽约市， 261 265 10 1145 3563703 3596653 – 。[193] https://doi.org/ . / . Daniel M. Ziegler、Nisan Stiennon、Jeffrey Wu、Tom B. Brown、Alec Radford、Dario Amodei、Paul Christiano 和 Geoffrey Irving。2020年。《根据人类偏好微调语言模型》。https://doi.org/10.48550/arXiv.1909. 08593 arXiv:1909.08593 [cs]. [194] Douglas Zytko 和 Nicholas Furlo。2023 年。《以在线约会为背景，与女性及LGBTQ+ 利益相关者共同设计性同意技术》。载于*2023*年计算机系统中的人为因素*CHI*会议论文集（德国汉堡）*(CHI '23)*。美国计算机协会，美国纽约州纽约市，文章编号 339, 17 10 1145 3544548 3580911 页。[195] https://doi.org/ . / . Douglas Zytko、Pamela J. Wisniewski、Shion Guha、Eric P. S. Baumer 和 Min Kyung Lee。2022年。《人工智能系统的参与式设计：面向多元用户、关系和应用领域的机遇与挑战》。载于*CHI*计算系统中的人为因素大会扩展摘要。ACM，美国路易斯安那州新奥尔良，1–4。https://doi.org/10.1145/3491101.3516506

设计空间

本节将进一步介绍二维设计空间的构思。对于每个维度，我们都会展示参与者构思的实例，包括具有代表性的引述，并简要概述支持这些维度概念化的文献检索。

A.1 维度1：循环过程

人际关系会随时间演变，持续经历建立、深化、变化和消解的过程 [101]。社会心理学和传播理论中的若干理论框架描绘了这一发展进程。社会渗透理论阐述了关系如何通过更深层次的自我披露而发展，从表面信息逐步深入到亲密信息 [4]。不确定性降低理论则解释了人们如何通过增进对彼此的了解来减少人际不确定性，尤其是在关系的早期阶段 [18]。与此同时，Knapp 的交互阶段模型全面概述了关系的两个过程：在不同阶段逐渐发展（“走到一起”）和逐渐恶化（“分道扬镳”） [101]。

设计空间中的循环过程维度由这些框架提供参考。此外，参与者的构思帮助我们确定了这一生命周期中人机组合体的关键阶段（示例见表 5）。值得注意的是，由于我们的研究聚焦于促进在线空间中的社会关系，因此着重强调了关系的发展（“走到一起”）。然而，关系的恶化（“分道扬镳”）也是人类生活中的关键阶段，在设计用于建立人类-AI关系的人工智能时不应忽视。

A.2 社会能力等级

人与人之间的社会互动需要复杂的社会细微差别、技能和智能 [82, 146]。人们与台式计算机等媒介互动时，往往会下意识地将社会特征归于技术，并对其应用社会规则 [148]；这一现象同样适用于人工智能系统 [90, 129]。随着人工智能在日常互动中的日益普及，学者和工程师一直致力于打造具备类人社会智能的系统，以满足日益增长的社会期望 [41, 175, 187]。此外，研究人员还设想了超越人类极限的AI能力 [49, 136, 177]。

同样，参与者在第一阶段对人工智能社会能力的构想可分为三个等级：(1) 基础人类增强，(2) 类人社会智能，以及 (3) 理想化超人能力。这些等级是累积递进的，高级等级建立在较低等级的能力之上（参与者构想的示例见表6）。

表5：第一阶段收集的构思，展示了维度1中的各个阶段（表2中的形成社会互动的循环过程）

阶段	构思示例	引文
S1. 数据收集——tion	用户应该告知人工智能自己的意图——图。	“这应该是我告诉人工智能‘嘿，我在找这个’的情况。我不知道能不能把你要找的东西自动化处理。比如，只有你自己知道你在找什么正在寻找的东西。人工智能不可能知道你在找什么。”（P05）
S2. 用户画像	人工智能构建用户画像，用户可以有效地修正或更新它。	“所以，如果人工智能能给出这些通用提示，人们就可以回应它们 [...] 这本质上又变成了一个用户画像。 [...] 我认为其中一点会是那种舒适感用户画像带来的舒适在于，我知道自己有一个页面 [...] 可以回去整理，或者修改或更新它们，而且用户界面非常清晰。这很有效。”（P07）
S3. 社会形态-tion	人工智能帮助发起一场群体基于……的对话综合而成的兴趣。	“一旦该群聊创建完成，人工智能就可以进行某种投票环节。 [...] 或许按重要性对它们（话题）进行排序，依据每个人想谈论的内容，然后它会到互联网上寻找有趣的内容，把它们放进群聊，并且然后我们就从那里开始对话。 [...] 某种有点像破冰活动的东西。”(P02)
S4. 社交维系	人工智能会提醒人们定期聚会。	“我觉得人工智能在这里最擅长做的事情之一，可能是……我喜欢组织活动，但有时这会让人觉得负担很重，也容易失去动力。我有一幅画和朋友们约好去酒吧，但已经差不多 6 个月没见面了，因为我们谁都无法持续参加。发生。因此，人工智能可能很擅长帮助事情持续推进，比如设置提醒。”（P04）
S5. 反思	人工智能评估交互并提供反馈。	“如果有人能告诉我这次会议有多高效就好了。也许产生了一些新想法进行了交换，对吧？也许发现了一些关于某件事的未知信息。 [...]也许我们的评分很低。也许我们在会议上根本没说话，对吧？ [...]它会给我们，你知道，就像是对谈话进展情况的一种反馈。”（P10）

表6：想法栏 第一阶段收集的、体现维度2各等级的想法（表中的AI社交能力等级） 3)

等级	创意示例	引言
L1. 人类增强化	管理 注意力并卸载认知——tion	“如果你被工作和学业压得喘不过气，那你总是有很多事情要做。你查看消息时会想，好吧，我晚点回复。然后你就再也没有回复。所以，就像一个提醒。 [...] 有点像是把这部分认知负荷卸载出去，也就是记忆方面的负担。这样你就可以在方便的时候主动维护它。”（P09）
L2. 社交副本	作为社交实践伙伴	“就像你的小伙伴一样，如果你对类似 [用户说：]‘哦，我可以在这个群组或频道里谈论空白内容吗？’ [人工智能说：] ‘你可以先跟我聊聊，这样至少你会更有信心，之后再私下跟别人聊聊。’关于你接下来要做的事是否合适。”（P04）
L3. 全知型社会存在	汇总无限的信息	“因为我们可能会遗漏很多事情，或者发帖的那个人可能会遗漏一些事情。或者，你知道，没有从新的角度看待这个话题。也许人工智能可以替我们完成这些事情，因为资源实在太多了。我的意思是，人类能够接触到的东西毕竟有限，但人工智能仿佛无处不在，而且人工智能的访问权限是无限的。所以，有很多事情是人工智能能够替我们完成，而我们在现实中无法亲自做到的。”（P08）