LRH：AI介导的通信

对人工智能媒介传播的担忧，源于对**自身与他人使用方式**的不同预期

Zoe A. Purcell[1*]、Mengchen Dong[2], 、Anne-Marie Nussberger[2], 、Nils Köbis[2], 和 Maurice Jakesch[3]

1. 图卢兹人工与自然智能研究所（ANITI），图卢兹大学，法国图卢兹

2. 人类与机器中心，马克斯·普朗克人类发展研究所，德国柏林。

3. 康奈尔科技学院，康奈尔大学，美国纽约

*通讯作者：zoe.purcell@iast.fr

贡献声明：MJ 和 NK 提出了最初的想法并开展了研究1。MJ 为研究1和研究2制作了刺激材料，并对研究1结果进行了初步分析。ZP 是首席研究员，负责研究2的设计，并在与所有作者协商后，主导了研究2与研究1的概念整合。ZP 在与所有作者协商后开展了研究2。MD 负责研究2的伦理审批和预注册。ZP 进行了效能分析、开展了分析、设计了图表，并在与 MD 和 AM 协商后起草了稿件。所有作者均提出了批判性意见，并协助完善研究设计、分析和稿件。

人工智能媒介传播

摘要

人工智能介导的传播技术（AICTs）迅速发展——这些数字工具利用人工智能增强我们的人际信息——引发了人们对人际信任未来的担忧，并促使人们紧急讨论披露和采用问题。我们通过评估人们对自我和他人使用公开及秘密人工智能传播技术的可接受性和使用情况的看法，为这一讨论作出贡献。在两项采用代表性样本的研究中（英国：N=477，美国：N=765），我们发现：（a）秘密（即未披露的）人工智能传播技术使用被认为不如公开使用可接受；（b）人们高估了他人对人工智能传播技术的使用；（c）人们预期他人会不负责任地使用人工智能传播技术。因此，我们对以下可能性表示担忧：对他人的错误认知和不同预期，可能会推动人们对人工智能媒介传播形成自我实现的悲观预期。

通俗摘要

人工智能介导的传播技术（人工智能传播技术）可以实时增强我们的信息——让人们的文本更有说服力、语音听起来更愉悦、图像看起来更具吸引力。然而，随着这些技术不断进步（例如，GPT-4 等大语言模型的出现），人们越来越难以区分原始传播内容与增强后的内容。功能强大的人工智能传播技术不断涌现，引发了人们的担忧：它们可能导致欺骗行为大范围蔓延，并使人际沟通中的信任大规模弱化。我们通过考察公开和秘密使用人工智能传播技术的可接受性，以及人们对自我与他人使用这类技术的预期，评估这些担忧是否成立。研究证实了悲观假设：秘密人工智能传播技术的可接受性低于公开人工智能沟通技术；人们高估了他人使用人工智能传播技术的程度，并认为他人会不负责任地使用这类技术。我们探讨了错误认知以及人们对他人使用人工智能传播技术的不同预期，如何可能促成自我实现的反乌托邦前景

，尤其是当个人和监管机构据此采取行动时。

关键词

人工智能（AI）；人工智能媒介传播（AIMC）；计算机媒介传播（CMC）；人机交互（HAII）；信任；技术风险评估。

人工智能媒介沟通

对人工智能媒介传播的担忧，源于人们对自身使用与他人使用的不同预期

人类是社会性存在——我们的生活和身份由人际关系定义。人际关系依赖于沟通、对真实性的假设以及人际信任（Grueter & White，2014年；Hruschka，2010年）。人工智能（AI），尤其是以人工智能介导的传播技术（AICTs）的形式，正彻底改变我们的沟通方式（Hancock 等人，2020年；Sundar，2020年；Sundar & Lee，2022年）。人工智能传播技术有望提升我们的沟通效率。然而，它们也可能增加欺骗，威胁我们对他人真实性的判断，并助长不信任（Jago，2019年；Jakesch、Hancock等人，2023年）。这些因素可能构成一幅关于人工智能传播技术及人机交互未来的反乌托邦图景。然而，这些担忧基于一些假设：人们认为秘密使用人工智能传播技术存在问题，认为他人会使用这些技术，并认为他人会不负责任地使用它们。我们将审视这些假设，希望找到摆脱这种悲观图景的出路。

如今，人工智能已应用于许多传播工具，例如 OpenAI 的 ChatGPT 等聊天机器人，以及亚马逊的 Alexa 等语音助手。尽管这些以人工智能为代理的传播工具是社会领域的重要参与者，但我们关注的是更复杂的人工智能媒介传播工具。人工智能媒介传播是指人与人之间的传播，其中计算代理会修改、扩充或生成信息，以实现传播目标或人际目标（Hancock 等人，2020年）。人工智能传播技术包括 Grammarly 和 Wordtune 等专用人工智能文本编辑器，也包括依托更通用人工智能的工具，例如 GPT-4 中的大语言模型。人工智能传播技术所生成的传播内容中，人类发送者与人工智能贡献者各自生成的部分之间的界限日益模糊。随着人们开始探讨这项复杂技术的伦理问题及其影响，我们必须

了解将引导这些讨论及后续监管的认知与个人观点。

人工智能传播技术本质上容易引发分歧，因为它们可能加剧欺骗和不信任（Jakesch 等人，2019年；Köbis 等人，2021年）。因此，它们引发了关于新兴人工智能语言技术的伦理问题及其影响的热烈讨论（Chesney 和 Citron，2019年；Hancock 等人，2020年；Jakesch、Bhat 等人，2023年；Ruggeri，2023年）。监管机构正忙于界定、劝阻并限制人工智能生成工具的不当使用。例如，纽约市学区是最早禁止 ChatGPT 的机构之一（Yang，2023年）；意大利实施了全面禁令（Browne，2023年）；欧洲刑警组织则对其潜在的犯罪用途表示严重关切（Chee，2023年）。与此同时，开发者正在寻求技术解决方案，例如使用 ZeroGPT 等 ChatGPT 检测算法。然而，人工智能传播技术日益普及，将使所有权、真实性和能动性之间的界限愈发模糊，并引发有关人工智能传播技术的认知、采用及其后果等根本性问题；这些问题不仅关乎政策制定者，也关乎科学专家和普通用户（Hsu，2023年；Munch 等人，2023年）。尽管大量资金投入了这些工具的开发，但人们对其如何影响个人观念、人际信任及社会层面的影响关注甚少。

本文探讨随着人工智能传播技术不断发展，哪些因素将影响个人对这些技术的看法及其采用。我们研究人们对可接受性的关键认知，以及对人工智能通信技术的使用的预期；还考察诸如透明度等实际前因——即是否公开或秘密地使用人工智能传播技术——以及用户这一因素，即我们讨论的是自我还是他人对人工智能传播技术的使用。我们聚焦于自我—他人区分，因为人们对（1）哪些行为可以接受、哪些不可接受（即命令性规范），以及（2）大多数他人通常会做什么（即描述性规范）的看法，都是预测个人行为的关键因素，尤其是那些在道德上存疑的行为（Bicchieri 和 Xiao，2009）。

此外，人们对自我和他人并不采用完全相同的道德标准（Valdesolo 与 DeSteno，2007；Weiss 等人，2018），对公开场合与私人场合也并非如此（Dana 等人，2007）。

在这方面，我们探讨了三个问题。人们是否认为秘密人工智能传播技术不如公开人工智能沟通技术可接受（主张1）？人们是否认为他人使用人工智能传播技术的频率会高于自己使用的频率（主张2）？人们是否认为他人使用人工智能传播技术是不负责任的（主张3）？除了这些问题，我们还探讨了包括对人工智能传播技术的看法以及对人工智能沟通工具的熟悉程度在内的个体因素所发挥的作用。我们的研究发现对个体、人际关系和社会层面均具有启示意义。研究结果揭示，个体是否对人工智能传播技术的未来持悲观态度，以及人工智能传播技术是否会影响人际信任。在社会层面，我们的研究发现有助于了解公众对人工智能传播技术的立场，以及这一立场将如何影响与更广泛的人机关系相关的技术和政策发展。

研究1

方法

参与者与方法。研究1考察了主张1，即工具使用的透明度（隐秘或公开）是否会影响其可接受性；以及主张2，即使用预期是否会受到相关用户（自我或他人）的影响。研究向一组具有代表性的英国参与者（N = 477）展示了隐秘和公开的人工智能传播技术，参与者年龄介于18至75岁之间（M = 48.48，SD = 15.83；女性 = 266[1]），并要求他们说明对自身与他人使用人工智能传播技术的预期。所有参与者均观看了六个关于人工智能传播技术的示例：三种媒介（视频、语音、文本），每种媒介各包含两种增强强度（弱增强和强增强；示例见图1和补充材料）。对于每一种人工智能传播技术，每位参与者都被要求回答四个问题。

[1] 其他性别：男性 = 200、非二元性别 = 4、不愿透露 = 3、未作答 =4。

问题，其中两个关于可接受性：1) 你认为某人公开使用这项技术的可接受程度如何？2) 你认为某人秘密使用这项技术的可接受程度如何？[1= 非常不可接受至5 = 非常可接受]，；另外两个关于使用预期：1) 他人使用此类技术的可能性有多大？以及 2) 你使用此类技术的可能性有多大？[1 = 非常不可能至5 = 非常可能]。



图1. 研究1中使用的刺激材料示例（媒介=视频，强度=强）。更多示例请参阅OSF。

结果

可接受性。与主张1一致，我们发现，人们对秘密使用人工智能传播技术的接受程度低于公开使用（图2（A））。我们采用线性混合模型，根据人工智能传播技术的透明度预测“可接受性”，并将参与者、媒介和强度纳入模型，作为随机截距效应。模型显示，透明度的主效应显著（B = - 0.69，95% 置信区间 [-0.74，-0.64], t(5718) = -26.53，p < .001）。因此，相较于公开使用人工智能传播技术（eM = 3.41，SE = .25），人们对秘密使用（eM = 2.72，SE = .25）的接受程度较低。

使用预期。与主张2一致，我们观察到，人们认为他人使用人工智能传播技术的程度会高于自己（图2（B））。我们采用线性混合模型，根据用户（自我或他人）预测预期使用情况，并将参与者、媒介和强度作为随机截距纳入模型。模型显示，用户的主效应显著（B = 0.93，95% 置信区间 [0.88，0.98], t(5718) = 33.98，p < .001）。因此，人们预期他人使用人工智能传播技术的程度（eM =3.73，SE = .19）会高于自己（eM = 2.80，SE = .19）。

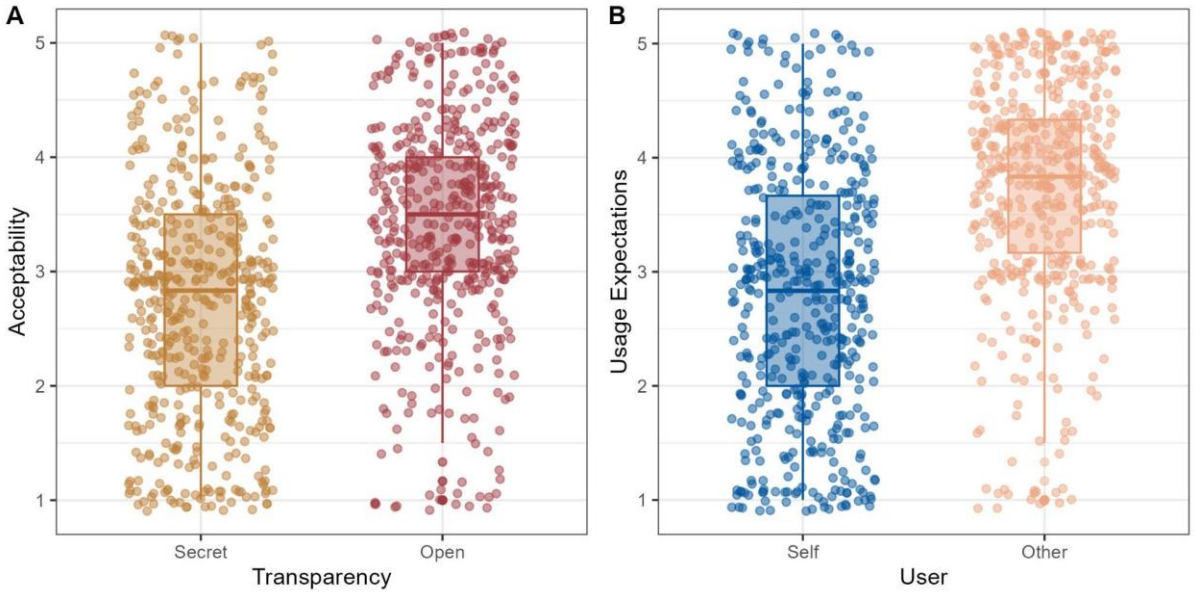

图2。研究1的结果显示，参与者对公开使用人工智能传播技术的接受度高于秘密使用（面板A），并且参与者认为他人使用人工智能传播技术的程度会高于自己（面板B）。数据点在垂直方向上随机抖动了0.1。

研究1发现，支持主张1的是，参与者对公开使用人工智能传播技术的接受度高于秘密使用；支持主张2的是，参与者认为他人使用人工智能传播技术的可能性高于自己使用的可能性。在研究1中，我们观察了当透明度/用户变得显著时参与者的反应。具体而言，对于每种媒介（文本、语音或视频），我们都询问了每位参与者公开和秘密使用该媒介的可接受程度、他们预期自己使用这些工具的程度，以及他们预期他人使用这些工具的程度。因此，每位参与者都能够直接对比自己在这方面的回答

关于透明度和用户。此外，除了检验主张3之外，研究2还考察了主张1和主张2的适用边界及二者之间的关系，包括当透明度和用户不再作为显著因素呈现时，这些反应模式是否依然成立。

研究2

研究2的所有方法和分析均已在OSF上预注册。

方法

参与者与方法。研究2通过考察人们对自身与他人使用人工智能传播技术的可接受性和使用预期之间的关系，评估人们是否预期他人会负责任地使用人工智能传播技术（主张3）。此外，研究2通过在被试间操纵透明度和用户，使我们能够考察主张1和主张2的边界条件。

我们调查了美国人口的一个代表性样本，按年龄、性别和族裔分层（N = 765），年龄为 18 至 93 岁（M = 45.47，SD = 15.75；女性 = 375[2]）。我们采用了 2（透明度：公开 vs. 秘密；被试间）× 2（用户：自我 vs. 他人；被试间）× 3（媒介：文本 vs. 语音 vs. 视频；被试内）的设计。参与者被分配到四种条件之一，看到的人工智能传播技术示例要么是公开的，要么是秘密的，从而操纵透明度。这些技术供参与者本人或他人使用，从而操纵用户。每位参与者都看到了三种人工智能传播技术示例：文本、语音和视频。为控制比较数量，我们仅使用了“强”增强型人工智能传播技术（示例见图 1 和图 3 及补充信息）。

[2] 其他性别：男性 = 379、非二元性别 = 8、不愿透露 = 3。



图3. 研究2中使用的刺激材料示例（透明度=公开，媒介=视频）。

有关透明度=秘密条件下等效刺激材料的示例，请参见图1。

针对三种人工智能传播技术，参与者被询问其使用预期[0 =非常不可能至100 = 非常可能] 以及可接受性[0 = 非常不可接受至100 = 非常可接受]。例如，处于“公开-自我”条件下的参与者会看到三种人工智能传播技术，这些技术 a) 会声明（以视觉或听觉方式）已经或正在使用人工智能辅助工具；随后参与者会被询问 b) 自己是否会使用该人工智能传播技术，以及 c) 这样做的可接受程度。而处于“隐瞒-他人”条件下的参与者会看到三种人工智能传播技术，这些技术 a) 不会声明是否使用了人工智能辅助工具；随后参与者会被询问 b) 他人是否会使用该人工智能传播技术，以及 c) 这样做的可接受程度。

针对每种媒介，我们探讨了若干可能相关的个体因素：1）受访者对人工智能传播技术的熟悉程度（例如：“您对在视频通话期间改变外貌的人工智能视频编辑工具有多熟悉？[0=完全不熟悉至100 =非常熟悉]”）；2）参与者是否认为使用人工智能传播技术会导致信息丢失；

（例如：“如果某人在视频通话期间使用人工智能视频编辑工具改变外貌，会损失多少信息？[0= 完全没有至100 = 损失很多]”）；3）参与者是否认为他人使用或不使用人工智能传播技术体现了其角色特征（例如：“他人是否使用人工智能视频编辑工具，会让我了解一些关于此人角色特征的信息。[0=非常不同意至100 = 非常同意]”）。我们在附录中报告了个体因素的结果，以及这些因素与主要主张之间的关系。

结果

以下部分将讨论：1）可接受性的评分以及主张1的边界；2）使用预期的评分以及主张2的边界；最后，3）主张3——对不负责任使用的预期，以及可接受性评分与使用预期之间的关系。

可接受性。关于主张1，在这些不显著的条件下，透明度对可接受性有较小且仅达到边缘显著水平的影响。然而，参与者评估的是自己使用还是他人使用的可接受性，会影响对人工智能传播技术可接受性的评分：与考虑自己使用相比，人们在考虑他人使用时认为人工智能传播技术更可接受。

我们使用线性混合模型，以用户（自我、他人）、透明度（公开、秘密）及二者的交互作用作为预测变量，并将参与者和媒介纳入模型，作为随机截距效应。在预测可接受性的模型中，用户的主效应显著（B = 4.79，95% 置信区间 [1.31，8.26]，t(2288) = 2.70，p = .007）。与考虑他人的使用情况（eM =57.64，SE = 4.94）相比，人们在考虑自己使用时的接受程度更低（eM = 52.85，SE = 4.94）。关于主张1，透明度的主效应处于边缘显著水平（B =3.25，95% 置信区间 [-0.23，6.73]，t(2288) = 1.83，p = .067），两者之间仅存在轻微差异。

他们对公开人工智能传播技术（eM = 56.87，SE = 4.94）的可接受性高于对秘密人工智能传播技术（eM = 53.62，SE= 4.94）的可接受性。透明度与用户对可接受性的交互作用不显著（B = - 1.74，95% 置信区间 [-8.69，5.22], t(2288) = -0.49，p = .624）。

使用预期。 与主张2一致，人们认为他人使用人工智能传播技术的频率会高于自己，即使在这些不显著条件下也是如此。然而，透明度并未影响使用预期。

我们使用线性混合模型，根据用户（自我、他人）、透明度（公开、秘密）以及用户与透明度的交互作用预测“使用”，并将参与者和媒介纳入为随机截距效应。与主张2一致，用户的主效应显著（B =30.97，95%置信区间 [27.58，34.36], t(2288) = 17.92，p < .001）。人们预期他人（eM =71.41，SE = 4.10）使用人工智能传播技术的程度会高于自己（eM = 40.44，SE =4.10）。然而，透明度的主效应不显著（B = 0.47，95%置信区间 [-2.92，3.86], t(2288) = 0.27，p = .787）；人们对公开的人工智能传播技术（eM = 56.16，SE = 4.10）和秘密人工智能传播技术（eM = 55.70，SE = 4.10）的使用预期没有差异。透明度与用户之间的交互作用也不显著（B = -0.85，95% 置信区间 [-7.63，5.92], t(2288) = -0.25，p = .805）。

使用预期与可接受性。如图4所示，我们发现主张3得到了支持：对于考虑自身使用的参与者而言，可接受性与使用之间的关联比考虑他人使用时更强。

为了分析这一发现，我们使用线性混合效应模型，根据用户、透明度和评价（可接受性、使用预期）以及它们之间的二重和三重交互作用预测评分。参与者和媒介作为随机截距效应纳入模型。与主张3一致，我们发现用户与评价之间存在显著交互作用（B = 26.18，95% CI [23.62，28.75], t(4579) = 20.00，p < .001）。有趣的是，我们还观察到

透明度与评价之间的交互作用较小但显著（B = -2.78，95%置信区间[-5.35，-0.22], t(4579) = -2.13，p = 0.034）。这些效应未受到三重交互作用的限定（p = .736；模型中的他人效应见附录）。

我们通过分别考察参与者在考虑自身人工智能通信技术的使用（用户=自我）和考虑他人人工智能通信技术的使用（用户=他人）时，可接受性与使用预期之间的关系，探讨了用户与评价之间的交互作用。我们使用线性混合模型，根据可接受性评分预测使用预期，并将参与者和媒介作为随机效应。对于“自我”人工智能通信技术的使用，可接受性与使用预期之间的关系（B = 0.78，95% 置信区间 [0.73，0.82], t(1126) =35.82，p < .001）明显强于“他人”人工智能通信技术的使用（B = 0.34，95% 置信区间 [0.30，0.38], t(1159) = 15.56，p< .001）。

同样，为了探究透明度与评价之间的交互作用，我们分别考察了考虑公开人工智能沟通技术的参与者和考虑秘密人工智能传播技术的参与者中，可接受性与使用预期之间的关系。我们采用线性混合模型，以可接受性评分预测使用预期，并将参与者和媒介设为随机效应。然而，公开人工智能沟通技术（B = 0.59，95% 置信区间 [0.54, 0.63], t(1129) = 23.81，p < .001）与秘密人工智能传播技术（B = 0.58，95% 置信区间 [0.53，0.63], t(1156) = 23.26，p < .001）中，可接受性与使用预期之间的关系相似。

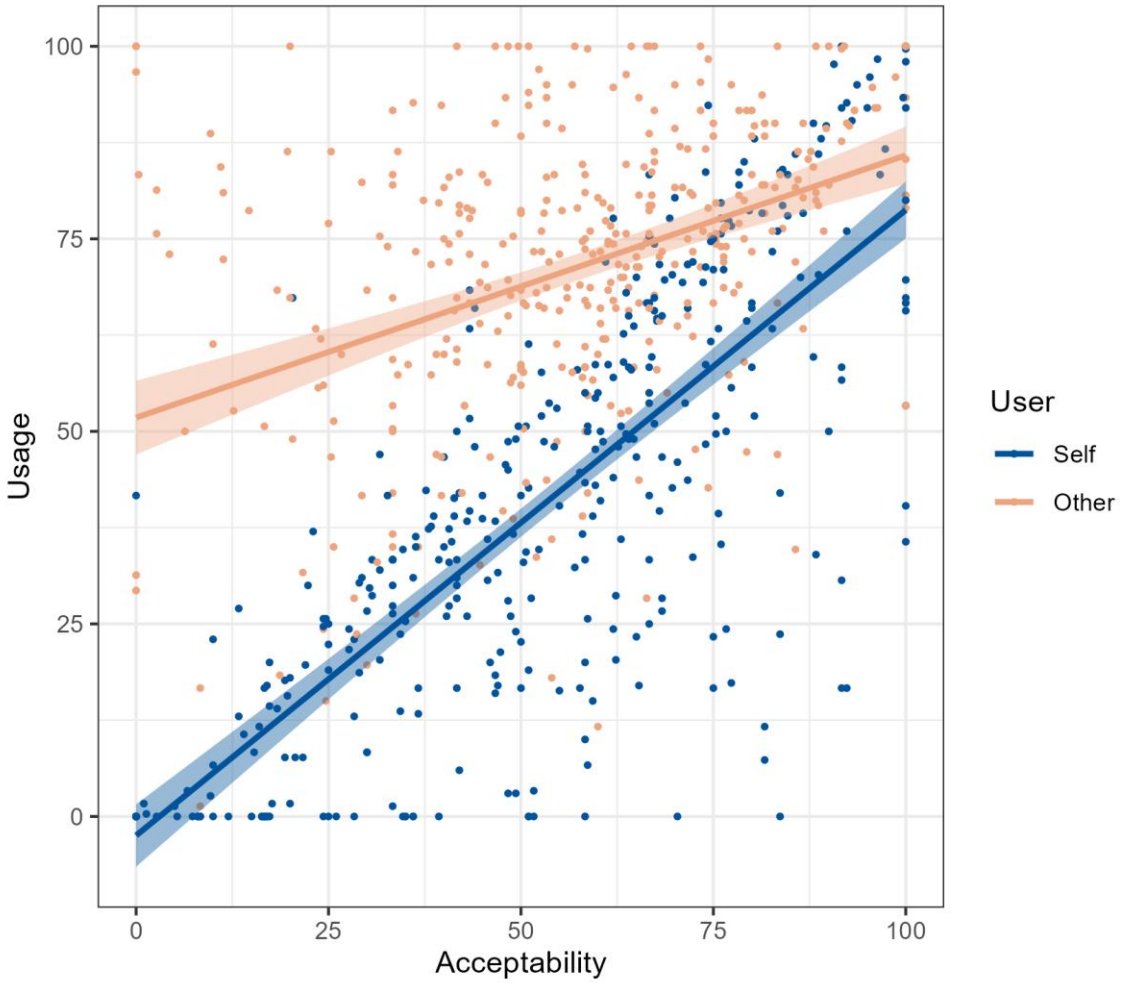

图4。 研究2的结果表明，可接受性与使用预期之间的关系，在人们考虑自己使用时，比考虑他人使用时更强。

研究2发现，关于主张1，在非显著性条件下，透明度对可接受性的影响仅达到边际显著水平；对公开人工智能沟通技术进行评价的参与者，比对秘密人工智能传播技术进行评价的参与者仅表现出略高的接受程度。关于主张2，参与者认为，他人使用人工智能传播技术的可能性高于自己使用的可能性，即使在非显著性条件下也是如此。研究2还发现，相较于自身使用人工智能传播技术，他人使用人工智能传播技术被认为更具可接受性，但透明度并未影响使用预期。

研究2明确支持主张3：人们在考虑他人的使用时，对何为可接受的判断与对使用情况的预期之间的吻合度，低于考虑自己使用时的吻合度。人们认为自己的使用会是负责任的，也就是说，会符合他们认定的可接受标准；但他们并不认为他人的使用也会同样负责任。

讨论

人工智能传播工具的开发与采用正呈指数级增长。一个典型例子是，Chat-GPT仅用五天就获得了一百万名用户（Buchholz，2023年）。尽管此类工具迅速普及，人们对这一发展持有怎样的看法仍鲜为人知，因此我们也不清楚这一发展可能会如何演进。为此，我们研究了人们对人工智能通信技术的使用的可接受性及其预期，并探讨了两个关键的情境因素：人工智能传播技术透明度（公开使用还是秘密使用），以及相关用户（自我还是他人）。有几种尚未检验的可能性或会挑战人们担忧人工智能媒介传播发展的这一假设：人们可能并不担心秘密使用人工智能传播技术；可能不认为他人会使用人工智能传播技术；即便认为他人会使用，也可能相信他人会负责任地使用。对于这三种说法，我们的研究结果均支持悲观假设。

人们认为，秘密人工智能传播技术——即不会告知接收者人工智能参与其中的技术——不如公开人工智能沟通技术——即会告知接收者人工智能参与其中的技术——可以接受。在研究1中，当有关公开使用和秘密使用的问题同时呈现时，这一效应十分显著。有趣的是，在透明度不那么突出的研究2中，这一效应较弱，且仅达到边缘显著。在研究2中，我们采用了被试间操纵来呈现公开使用或秘密使用，但并未明确提醒参与者人工智能传播技术存在这些运行模式。这些发现对于理解以下观点十分重要：

原则上，人们在意人工智能通信技术使用中的透明度。然而，人工智能使用的可接受性取决于技术如何引入，以及透明度措施如何传达。这一认识对未来的人工智能可接受性研究至关重要。此外，就人工智能传播技术的采用和消费者偏好而言，这些发现意味着，在人工智能传播技术产品的伦理设计中，应认真考虑透明度。不过，诸如人们是否了解不同设计选项之类的情境因素，也会对用户的决策至关重要。随着认知的加深，用户可能会逐渐放弃不符合其伦理偏好的人工智能传播技术产品。

在考虑自身与他人的人工智能传播技术使用时，人们在对人工智能传播技术的可接受性判断和使用预期方面表现出显著且稳定的自我与他人差异。人们预期他人使用人工智能传播技术的程度高于预期自己使用的程度。这一发现表明，人们高估了他人使用人工智能传播技术的程度。有趣的是，与此同时，人们还认为，他人使用人工智能传播技术比自己使用更具可接受性。这种可接受性标准的差异可能表明，在人工智能传播技术的使用方面，人们对他人采用了不同（较低）的可接受行为标准；也可能如行为驱动合理化这一观点所指出的那样，人们预期他人会更频繁地使用人工智能传播技术，而这一预期可能进一步提高了他们对这种使用行为可接受性的认知（Eriksson 等人，2015）。同样，人们可能会认为，相较于自己，他人对人工智能传播技术工具有更大的需求或拥有更多的获取渠道。

关于可接受性与使用之间的关联，也出现了自我—他人区分。人们对自己使用情况的预期与他们对可接受性的认知高度一致，但对他人使用情况的预期则不那么一致。这一结果表明，我们认为自己使用人工智能传播技术的方式会是负责任的，因为这种使用符合我们认定的可接受标准；但我们并不认为他人使用人工智能传播技术时也会遵循同样的原则。结合我们发现人们会高估他人对人工智能传播技术的使用，这一洞见

表明，人们可能普遍对人工智能传播技术及其使用的发展持负面看法，认为他人的采用程度会更高，而且不会符合他们对于何为可接受的个人信念。

我们对个体因素的探讨为理解可接受性和使用预期的潜在驱动因素提供了一些线索。例如，对于人工智能传播技术，人们认为更可接受的行为与他们对人工智能传播技术的使用是否体现了用户的角色，以及他们是否认为人工智能传播技术的使用会导致信息损失的看法有关。此外，这些发现也可能有助于我们解读贯穿各项主要论点的自我—他人区分。人们对人工智能传播技术的使用是否具有“信号传递”作用（即是否能够反映用户的角色）的看法，似乎与他们对自身使用人工智能传播技术的评价密切相关，但与对他人使用情况的评价无关。事实上，当人们认为使用人工智能传播技术在道德上可能存在负面影响时，他们在管理自身声誉方面对自己的要求比对他人更加严格（Dong et al., 2023）；因此，报告中对自身使用行为较低的预期，可能反映了声誉管理方面的一种策略性举措。因此，人们对人工智能传播技术产品的实际选择，可能受到其对他人使用情况的不准确认知的影响——在其他具有道德争议的行为中也观察到了类似模式（Dorrough et al., 2023）。

人工智能传播技术的发展与采用正在迅速推进，随之而来的是围绕可接受性和监管的激烈争论。未来人工智能传播技术的发展、政策制定与采用，很可能取决于公众认为哪些做法可以接受，以及公众预期这些技术将如何被使用。我们提供了实证证据，支持以下观点：可接受性和使用预期与透明度（公开/秘密）以及用户（自我/他人）有关。我们深入探讨了自我与他人认知之间这一基本心理区分所发挥的作用，以及个人对人工智能传播技术的信念所产生的影响。我们发现，在稳固的自我与他人区分这一认知基础上，人们确实对人工智能传播技术的发展持负面看法，这是因为他们高估了他人的使用程度，并认为他人的使用行为将是不负责任的。这些

研究结果证实了人们对人工智能传播技术（AICTs）感到担忧这一观点，并表明这种恐惧源于人们对自身与他人使用此类技术抱有不同的预期。未来，通过提高人们的认知水平和熟悉度，或实施开放式人工智能传播技术政策（尤其是考虑到人们更倾向于公开使用而非秘密使用人工智能传播技术），或许能够缓解这些错误认知及其引发的恐惧。然而，如果这些问题得不到纠正，它们可能会演变成自我实现的预言，加速个体负面前景预期的形成，削弱人际信任，并限制对人机交互产生影响的技术与监管措施的健康发展。

参考文献

Bicchieri, C., & Xiao, E.（2009）。做正确的事：但前提是他人也这么做。《行为决策研究》，22(2)，191–208。https://doi.org/10.1002/bdm.621

Browne, R.（2023年4月4日）。意大利成为首个禁止 *ChatGPT* 的西方国家。以下是其他国家的做法。CNBC。https://www.cnbc.com/2023/04/04/italy-has-banned-chatgpt-heres-what-other-countries-are-doing.html

Buchholz, K.（2023年1月24日）。ChatGPT 用户数迅速达到 100 万。Statista 信息图表。https://www.statista.com/chart/29174/time-to-one-million-users

Chee, F. Y.（2023年3月27日）。欧洲刑警组织就 ChatGPT 被用于犯罪敲响警钟，并对前景表示悲观。路透社。https://www.reuters.com/technology/europol-sounds-alarm-about-criminal-use-chatgpt-sees-grim-outlook-2023-03-27/

Chesney, B., & Citron, D.（2019年）。深度伪造：隐私、民主和国家安全面临的迫在眉睫的挑战。《加利福尼亚法律评论》，107(6)，1753–1820。

Dana, J., Weber, R. A., & Kuang, J. X. (2007). 利用道德回旋余地：证明公平偏好是一种错觉的实验。《经济理论》，33(1)，67–80。

Dong, M., Kupfer, T. R., Yuan, S., & van Prooijen, J.-W.（2023年）。做好人以赢得好名声：自我报告的道德品格能够预测追求声誉者的道德双重标准。《英国心理学杂志》，114(1)，244–261。https://doi.org/10.1111/bjop.12608

Dorrough, A. R., Köbis, N. C., Irlenbusch, B., Shalvi, S., & Glöckner, A.（2023年）。有条件贿赂：来自18个国家激励实验的见解。《美国国家科学院院刊》。

Eriksson, K., Strimling, P., & Coultas, J. C.（2015）。描述性规范与命令性规范之间的双向关联。《组织行为与人类决策过程》，129，59–69。https://doi.org/10.1016/j.obhdp.2014.09.011

Hancock, J. T., Naaman, M., & Levy, K.（2020年）。人工智能媒介传播：定义、研究议程与伦理考量。《计算机中介传播期刊》，25(1)，89–100。https://doi.org/10.1093/jcmc/zmz022

Hsu, T.（2023年1月22日）。深度伪造泛滥，各国难以应对。《纽约时报》。https://www.nytimes.com/2023/01/22/business/media/deepfake-regulation-difficulty.html

Jago, A. S.（2019年）。算法与真实性。《管理学会探索》，5(1)，38–56。https://doi.org/10.5465/amd.2017.0002

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., & Naaman, M.（2023年）。与带有观点倾向的语言模型协同写作会影响用户的观点。《2023年计算机系统中的人为因素大会论文集》，1–12。

Jakesch, M., French, M., Ma, X., Hancock, J. T., & Naaman, M.（2019年）。人工智能媒介传播：认为个人资料文本由人工智能撰写会如何影响可信度。《2019年计算机系统中的人为因素大会论文集》，1–13。

Jakesch, M., Hancock, J. T., & Naaman, M. (2023年). 人类对 AI 生成语言的启发式判断存在缺陷。《美国国家科学院院刊》, 120(11), e2208839120. https://doi.org/10.1073/pnas.2208839120

Köbis, N., Bonnefon, J.-F., & Rahwan, I. (2021). 糟糕的机器会败坏良好的道德。《自然·人类行为》, 5(6), Article 6. https://doi.org/10.1038/s41562-021-01128-2

Munch, L., Mainz, J., & Bjerring, J. C. (2023年). 算法决策中责任缺口的价值。《伦理学与信息技术》, 25(1), 21. https://doi.org/10.1007/s10676-023-09699-6

Ruggeri, A.（2023年3月1日）。*TikTok*备受争议的“美颜滤镜”问题。Https://Www.Bbc.Com/. https://www.bbc.com/future/article/20230301-the-problems-

Sundar, S. S.（2020年）。机器能动性的兴起：研究人机交互心理学（HAII）的框架。《计算机中介传播期刊》，25(1)，74–88。https://doi.org/10.1093/jcmc/zmz026

Sundar, S. S. 与 Lee, E.-J.（2022年）。重新思考人工智能时代的传播。《人类传播研究》，48(3)，379–385。https://doi.org/10.1093/hcr/hqac014

Valdesolo, P. 与 DeSteno, D.（2007年）。道德虚伪：社会群体与美德的灵活性。《心理科学》，18，689–690。https://doi.org/10.1111/j.1467-9280.2007.01961.x

Weiss, A.、Burgmer, P. 与 Mussweiler, T.（2018年）。双面道德：不信任会促使自我与他人采用不同的道德标准。《人格与社会心理学通报》，44(12)，1712–1724。https://doi.org/10.1177/0146167218775693

Yang, M.（2023年1月6日）。纽约市学校禁止使用能够撰写论文和回答提示的人工智能聊天机器人。《卫报》。https://www.theguardian.com/us-news/2023/jan/06/new-york-city-schools-ban-ai-chatbot-chatgpt

附录

研究2的扩展结果

可接受性和使用预期。为考察可接受性评分与使用预期之间的关系，以及它们与透明度和用户的交互作用，我们使用线性混合模型，根据用户、透明度和“评价”（可接受性或使用预期）预测评分。参与者和媒介被纳入为随机效应。该模型的截距（对应于模型均值）为 55.59（95% 置信区间 [47.35, 63.83], t(4579) = 13.23, p < .001）。在该模型中：

- 透明度的效应在统计上不显著且为正（B = 1.86，95% 置信区间 [-1.24，4.96], t(4579) = 1.18，p = 0.239）。

- 用户的效应在统计上显著且为正（B = 17.88，95% 置信区间 [14.78，20.98], t(4579) = 11.31，p < .001;

- 评价的效应在统计上不显著且为正（B = 0.68，95% 置信区间 [-0.60，1.97], t(4579) = 1.04，p = 0.297）。

- 透明度与用户的交互作用在统计上不显著，且为负向（B = -1.30，95% 置信区间 [-7.49，4.90], t(4579) = -0.41，p = 0.682；简单效应见下文。

- 透明度与评价的交互效应在统计上显著，且为负向（B = -2.78，95% 置信区间 [-5.35，-0.22], t(4579) = -2.13，p = 0.034 ）；更多详情请见正文。

- 用户与× 评价的交互作用效应在统计上显著且为正（B = 26.18，95% 置信区间 [23.62，28.75], t(4579) = 20.00，p < .001）；详情请参见正文。

- 透明度、× 用户与× 评价的三重交互效应在统计上不显著且为正（B = 0.88, 95% 置信区间 [-4.25，6.02], t(4579) = 0.34，p =0.736）。

个体因素。

除了主要分析外，我们还考察了三个因素——熟悉度、信息损失和角色信号——如何影响人们对人工智能传播技术的可接受性及其使用预期。

个体因素与可接受性。我们发现，较高的可接受性与以下因素相关：a) 对人工智能沟通工具的熟悉程度较高，b) 对使用人工智能传播技术导致信息损失的预期较低，以及 c) 使用人工智能传播技术所提供的角色信号较弱（见图A）。使用三个结构与使用预期模型相同的线性混合效应模型验证了这些结果（见下文）。此外，我们发现，相较于公开人工智能沟通技术，信息损失对秘密人工智能沟通技术可接受性的影响更强。其他影响均未达到统计显著性。

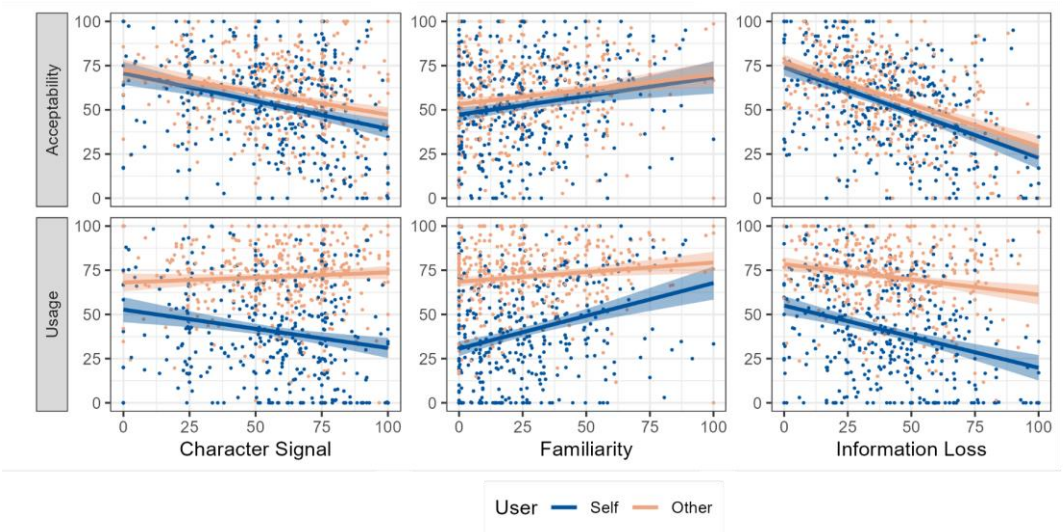

图　　A. 对于感知到的可接受性，研究2发现了与特征呈负相关的关系。　　*r*

信号与信息损失和熟悉度呈正相关。关于使用，对于自身

人工智能通信技术的使用，研究发现，使用预期与角色信号、使用预期与信息损失之间均呈负相关，而机身预期与熟悉度之间呈正相关。

以下进一步阐述这些结果：

- 可接受性与熟悉度。我们采用线性混合模型，根据用户（自我、他人）、透明度（公开、隐秘）和熟悉度 [0 to 100]，以及这三个预测变量之间所有可能的交互作用，预测“使用”。将参与者和媒介（文本语音、视频）作为随机效应。截距设为均值时：透明度的效应在统计上不显著且为正向（B = 2.88，95% 置信区间 [-1.19，6.95], t(2284) = 1.39，p = .165）；用户的效应在统计上显著且为正向（B = 5.79，95% 置信区间 [1.72，9.86], t(2284) = 2.79，p = .005）；熟悉度的效应在统计上显著且为正向（B = 0.15，95% 置信区间 [0.11，0.20],t(2284) = 6.94，p < .001）；用户与透明度的交互效应在统计上不显著且为正向（B = 0.80，95% 置信区间 [-7.34，8.94], t(2284) =0.19，p = .847）；熟悉度与透明度的交互效应在统计上不显著且为正向（B = 8.96e-03，95% 置信区间 [-0.07，0.09], t(2284) = 0.21，p =0.832）；熟悉度与用户的交互效应在统计上不显著且为负向（B = -0.04, 95% 置信区间 [-0.12，0.04], t(2284) = -0.92，p = 0.356）；熟悉度与（透明度 * 用户）的交互效应在统计上不显著且为负向（B = -0.09，95% 置信区间 [-0.26, 0.07], t(2284) = -1.10，p = 0.271）。

- 可接受性与信息损失。我们采用线性混合模型，以用户（自我、他人）、透明度（公开、秘密）和信息损失 [0 to100] 为预测变量预测“使用”，并纳入这三个预测变量之间所有可能的交互作用。参与者和媒介（文本语音、视频）作为随机效应纳入模型。截距设为均值时：透明度的效应在统计上不显著且为负（B = -

1.59，95% 置信区间 [-6.16，2.98], t(2284) = -0.68，p = .496）；用户的效应在统计上不显著且为正（B = 4.52，95% 置信区间 [-0.05，9.09], t(2284) = 1.94，p = .053）；信息损失的效应在统计上显著且为负（B = -0.41，95% 置信区间 [-0.46，-0.37], t(2284) = -19.03，p < .001）；用户与透明度的交互效应在统计上不显著且为负（B = -3.64，95% 置信区间[-12.78，5.50], t(2284) = -0.78，p = 0.435）；信息损失与透明度的交互效应在统计上显著且为正（B = 0.11，95% 置信区间 [0.03，0.19],t(2284) = 2.57，p = 0.010）；信息损失与用户的交互效应在统计上不显著且为负（B = -2.62e-03，95% 置信区间 [-0.08，0.08],t(2284) = -0.06，p = 0.950）；信息损失与（透明度 * 用户）的交互效应在统计上不显著且为正（B = 0.07，95% 置信区间 [-0.10，0.23],t(2284) = 0.79，p = 0.429；标准化 B = 0.01，95% 置信区间 [-0.02，0.05]）。为探究透明度与信息损失之间显著的交互效应，我们采用线性混合模型，以信息损失为预测变量预测可接受性，并将参与者和媒介作为随机效应。随后，我们分别考察秘密人工智能传播技术和公开人工智能沟通技术；信息损失对秘密人工智能传播技术的影响强于对公开人工智能沟通技术的影响。

- 对于秘密人工智能传播技术，信息损失的影响在统计上显著且为负（B = -0.47，95% 置信区间 [-0.53，-0.41], t(1156) = -15.73，p < .001）。

- 对于公开人工智能沟通技术，信息损失的影响在统计上显著且为负（B = -0.37，95% 置信区间 [-0.43，-0.30], t(1129) = -11.46，p < .001）。

- 可接受性与角色信号。我们采用线性混合模型，以用户（自我、他人）、透明度（公开、隐秘）和角色信号 [0 to100] 为预测变量来预测“使用”，并纳入这三个预测变量之间所有可能的交互作用。参与者和媒介（文本语音、视频）作为随机效应纳入模型。截距设为均值时：透明度的效应在统计上不显著且为正（B =

3.13，95% 置信区间 [-3.25，9.51], t(2284) = 0.96，p = .336）；用户的效应在统计上不显著且为正（B = 4.46，95% 置信区间 [-1.92，10.83], t(2284) = 1.37，p = .171）；角色信号的效应在统计上显著且为负（B = -0.36，95% 置信区间 [-0.41，-0.31], t(2284) = -14.36，p < .001）；用户与透明度的交互效应在统计上不显著且为负（B = -9.51，95% 置信区间[-22.26，3.25], t(2284) = -1.46，p = .144）；角色与透明度的交互效应在统计上不显著且为负（B = -9.96e-04，95% 置信区间 [- 0.09，0.09], t(2284) = -0.02，p = .983）；角色与代理的交互效应在统计上不显著且为正（B = 0.02，95% 置信区间 [-0.07，0.12], t(2284) =0.45，p = 0.650）；角色与（透明度 * 代理）的交互效应在统计上不显著且为正（B = 0.13，95% 置信区间 [-0.06，0.32], t(2284) =1.37，p = .171）。

对使用预期的影响。较高的使用预期与 a) 较高的熟悉度和 b) 较低的信息损失相关，而较低的使用预期则与更强的角色信号相关。对于处于“自我”条件下的参与者而言，这些关联更为显著，也就是说，他们考虑的是自己潜在的人工智能通信技术的使用（见图A）。我们采用三个线性混合效应模型验证了这些结果，每个探索性因素对应一个模型；模型根据用户（自我、他人）、透明度（公开、保密）以及探索性因素 [<熟悉度/信息损失/角色信号>，预测“使用”；0 至 100]。参与者和媒介作为随机效应纳入模型。以下将进一步阐述这些结果：

- 使用预期与熟悉度。我们采用线性混合模型，根据用户（自我、他人）、透明度（公开、秘密）和熟悉度 [0 to 100]预测“使用”，并纳入这三个预测变量之间所有可能的交互作用。参与者和媒介（文本语音、视频）作为随机效应纳入模型。模型截距设为均值时：透明度的效应在统计上不显著且为正（B =

0.89，95% 置信区间 [-3.02，4.80], t(2284) = 0.45，p = .656）；用户的效应在统计上显著且为正（B = 34.71，95% 置信区间 [30.80，38.62], t(2284) = 17.40，p < .001）；熟悉度的效应在统计上显著且为正（B =0.19，95% 置信区间 [0.15，0.23], t(2284) = 8.73，p < .001）；用户与透明度的交互效应在统计上不显著且为正（B = 3.25，95% 置信区间 [-4.57，11.07], t(2284) = 0.82，p = .415）；熟悉度与透明度的交互效应在统计上不显著且为负（B = -0.02，95% 置信区间 [-0.10，0.06],t(2284) = -0.54，p = .590）；熟悉度与用户的交互效应在统计上显著且为负（B = -0.14，95% 置信区间 [-0.22，-0.06], t(2284) = -3.49，p < .001）；熟悉度对（透明度 * 用户）的交互效应在统计上不显著且为负（B = -0.15，95% 置信区间 [-0.31，0.01], t(2284) = -1.82，p = .070）。为探究熟悉度与代理之间的交互作用，我们分别考察了熟悉度对自我和他人使用的影响。熟悉度对自我使用的影响强于对他人使用的影响：

○ 对于自我而言，熟悉度的影响在统计上显著且为正（B =0.22，95% 置信区间 [0.15，0.29], t(1126) = 6.26，p < .001）。

○ 对于他人，熟悉度的影响在统计上显著且为正向（B =0.13，95% 置信区间 [0.09，0.18], t(1159)= 5.49，p < .001）。

- 使用预期与信息损失。我们采用线性混合模型，根据用户（自我、他人）、透明度（公开、隐秘）、信息损失 [0 to 100] 以及这三个预测变量之间所有可能的交互作用，预测“使用”。参与者和媒介（文本语音、视频）作为随机效应纳入模型。模型截距设为均值时：透明度的效应在统计上不显著且为负（B = -1.97，95% 置信区间 [-6.68，2.75], t(2284) = -0.82，p = .414）；用户的效应在统计上显著且为正（B = 19.67，95% 置信区间 [14.96，24.39], t(2284) =

8.18，p < .001）；信息损失的效应在统计上显著且为负向（B = -0.23，95% 置信区间 [-0.27，-0.18], t(2284) = -10.28，p < .001）；用户与透明度的交互效应在统计上不显著且为正向（B = 1.90，95% 置信区间 [- 7.53，11.33], t(2284) = 0.39，p = .693）；信息损失与透明度的交互效应在统计上不显著且为正向（B = 0.05，95% 置信区间 [-0.03，0.13], t(2284) = 1.19，p = .235）；信息损失与用户的交互效应在统计上显著且为正向（B = 0.27，95% 置信区间 [0.19，0.35], t(2284) = 6.40，p< .001）；信息损失与（透明度 * 用户）的交互效应在统计上不显著且为负向（B = -0.05，95% 置信区间 [-0.22，0.12], t(2284) = - 0.58，p = .559）。

- 使用预期和角色信号。我们采用线性混合模型，根据用户（自我、他人）、透明度（公开、秘密）和角色信号 [0 to 100] 预测“使用”，并纳入这三个预测变量之间所有可能的交互作用。参与者和媒介（文本语音、视频）作为随机效应纳入模型。 以模型截距位于均值为基准：透明度的效应在统计上不显著且为正（B = 5.48，95% 置信区间 [-0.86，11.83], t(2284) = 1.69，p = .090）；用户的效应在统计上显著且为正（B = 9.84，95% 置信区间 [3.49，16.18], t(2284) =3.04，p = .002）；角色的效应在统计上显著且为负（B = - 0.20，95% 置信区间 [-0.25，-0.16], t(2284) = -8.26，p < .001）；用户与透明度的交互效应在统计上不显著且为负（B = -3.56，95% 置信区间 [- 16.25，9.13], t(2284) = -0.55，p = .582）；角色与透明度的交互效应在统计上不显著且为负（B = -0.09，95% 置信区间 [-0.18, 5.53e-03], t(2284) = -1.84，p = .065）；角色与用户的交互效应在统计上显著且为正（B = 0.37，95% 置信区间 [0.28，0.47], t(2284) = 7.90，p< .001）；角色与（透明度 * 用户）的交互效应在统计上

不显著且为正（B = 0.05， 95% 置信区间 [-0.14， 0.24], t(2284) = 0.53, p =0.594）。