

大学生对生成式人工智能作品的创造性感知与评价偏差*

熊小艳¹ 侯佳林¹ 丁佑尹² 向绪立² 何东² 龙海鹰³ 陈群林^{**1}

(¹西南大学心理学部, 重庆, 400715) (²西南大学美术学院, 重庆, 400715)

(³堪萨斯大学教育和人类科学学院教育心理学系, 堪萨斯, KS 66045)

摘要 随着生成式人工智能在创作过程中的广泛应用, 个体能否感知生成式人工智能作品的创造性水平以及是否存在评价偏差, 成为值得探讨的问题。研究以人类和生成式人工智能的艺术作品作为实验材料, 通过大规模线上行为实验, 考察大学生对生成式人工智能作品创造性水平的感知、评价及其影响因素。结果发现: 大学生对生成式人工智能作品的创造性评分接近或高于人类作品; 当作品被推测或标签为“生成式人工智能创作”时, 大学生会显著低估其创造性, 表现出推测偏差和评价偏差。进一步分析表明, 这种偏差受到大学生与人工智能的接触程度及其专业背景的调节。研究发现对生成式人工智能在创意生成领域的应用具有指导意义。

关键词 生成式人工智能 创造性 推测偏差 评价偏差

1 引言

2022年, OpenAI推出的基于大模型技术的言语对话系统ChatGPT, 迅速引发了社会各界对生成式人工智能(generative artificial intelligence, GAI)的广泛关注。随后, 文本、图片和视频生成的GAI产品不断涌现, 成为公众日常工作、专业领域以及创意行业的重要工具(Anantrasirichai & Bull, 2022; Tubadji et al., 2021)。多项研究表明, GAI已展现出近似人类的创造性特征(Cheng, 2022), 在写作、艺术品生成和创造性测试等任务中的创造性表现接近甚至超越人类(Hong et al., 2022; Hubert et al., 2024)。由于创造性评价依赖个体自身经验, 个体对GAI作品创造性的感知可能受其与人类作品的比较影响。此外, 当GAI与人类作品共同接受创造性评价时, 个体对GAI作品的评价会发生怎样的变化? 为此, 本研究聚焦GAI输出的艺术作品, 比较大学生对GAI与人类作品的创造性评分差异, 分析其评价偏差及影响因素, 旨在为GAI的应用推广、评估及社会效用提供实证依据。

创造性通常是指个体产生独特且适宜的产品或想法的能力(Runco & Jaeger, 2012)。已有研究通

过图灵测试(Turing, 2007)考察个体对人类与GAI作品的感知区分能力, 发现GAI作品在独创性和价值特征上与人类创作无显著差异(Elgammal et al., 2017; Hitsuwari et al., 2023), 甚至领域内的专家也难以有效区分(Schubert et al., 2017)。此外, 在言语创造性任务中, GAI在多个维度上的得分高于人类被试(Hubert et al., 2024)。在视觉艺术领域, 尽管人类对抽象艺术绘画的区分能力较高, 识别率达到了63.8%(Demmer et al., 2023; Gangadharbatla, 2022), 但对油画作品的识别率仅为53%(Ragot et al., 2020)。这些研究表明, 从创造性结果角度来看, GAI能够输出与人类相当的创造性产品。然而, 现有研究主要聚焦于个体对GAI作品的喜好、价值、审美偏好等方面(Amabile, 1982; Bellaiche et al., 2023; Chamberlain et al., 2018; Chiarella et al., 2022), 而较少关注诸如“作品是否具有创造性”等需要个体深度认知加工的评价指标。

GAI在创造性任务中表现出接近人类的创造性水平, 但个体对其作品的评估仍存在多种偏差。与专用人工智能不同, GAI在高阶认知任务中的良好表现使其能够有效辅助人类认知活动(Banh & Strobel, 2023), 但同时也引发了人类对“机器

* 本研究得到教育部人文社会科学研究青年基金西部和边疆项目(23XJCZH003)的资助。

** 通讯作者: 陈群林, E-mail: chenqunlin@swu.edu.cn

DOI:10.16719/j.cnki.1671-6981.20260315

换人”或技术性失业的担忧,这种矛盾心理导致个体在评价人工智能产品时产生偏差(Mokyr et al., 2015; Zlotowski et al., 2017)。例如,个体通常认为质量较高的艺术品出自人类艺术家,而GAI创作的艺术品则被认为“技不如人”(Banh & Strobel, 2023),或者受到作品作者标签的影响(Millet et al., 2023; Ragot et al., 2020; Shank et al., 2023; Zhang et al., 2025),如个体会对标有“GAI创作”的艺术品在喜爱程度和价值评分上显著压低(Millet et al., 2023; Shank et al., 2023)。尽管也有少量研究发现个体更偏好GAI作品(Van Hees et al., 2025),但大部分研究证实个体在评价GAI作品时存在评价偏差。社会认同理论为这一现象提供了合理解释,即个体可能将“创造力”视为人类特有的能力来维持群体独特性,这种自我群体偏好导致对GAI作品创造性的低估倾向(Ragot et al., 2020)。

然而,已有研究发现个体对人工智能的态度并非固定不变,而是受到个体接触人工智能的程度及相关经验(Kaya et al., 2024; Ting et al., 2023)、文化(Dang & Liu, 2021)、人格特征(Grassini & Koivisto, 2024; Kaya et al., 2024)、自我效能感(Latikka et al., 2023)、性别和受教育水平(Wang, 2024)等多种因素的影响。例如,深入了解人工智能创作过程(如观看机器人艺术创作的视频)能够显著改善个体对作品质量的评价偏差(Chamberlain et al., 2018)。工作场景中的人工智能接触不仅未增加负面评价,反而提升了使用者对人工智能带来就业机会的积极认知(Fietta et al., 2021)。与之类似,大学生群体对GAI创作普遍持较为积极的态度,表现出对GAI作品产生更强烈的喜爱偏好(Park & McGee, 2023),这可能与高频使用GAI辅助学习(如写作、编程和创意)有关。然而,不同学科背景的大学生对ChatGPT的使用频率存在差异,例如STEM(science, technology, engineering, mathematics)专业学生每月使用ChatGPT的比例(45.17%)高于非STEM专业学生(22.58%)(Baek et al., 2024)。那么,大学生的人工智能接触程度是否影响其对GAI作品的创造性感知和评价,尚缺乏研究。

综上,本研究聚焦探讨大学生对人类与GAI作品创造性感知与评价偏差。为此,研究选取人类与GAI创作的三行诗和油画作为实验材料。鉴于近期研究指出作品质量的变异性可导致“质量效

应”掩盖“标签效应”(Johnson & Proudfoot., 2024; Medeiros et al., 2025),本研究对作品进行了高、低质量分组,以此提升实验的内部效度和结果的可解释性。研究采用不标注与随机标注作品作者两种操作方式,通过大规模线上数据采集,探讨以下三个问题。(1)实验1中,通过对人类和GAI作品的独创性、适宜性以及喜好程度进行评分(衣新发等, 2011; Amabile, 1982; Bellaiche et al., 2023; Runco & Jaeger, 2012),考察大学生对GAI作品创造性的感知;(2)实验2中,通过对作品作者身份的推测以及对作品的创造性评分,考察大学生对GAI作品的创造性是否存在推测偏差与评价偏差。我们假设,对于高创造性作品,被试更倾向于判断为人类创作,且对推测或标注为“GAI创作”的作品给予更低的创造性评分;(3)实验3中,考察大学生对GAI作品的评价偏差是否受其接触人工智能的程度及专业背景的调节。我们假设,接触人工智能程度越高、专业背景为STEM的个体,其对GAI作品的评价偏差会相对较弱。

2 方法

2.1 被试

本研究使用JavaScript编写程序,通过脑岛在线研究公共平台(www.naodao.com)采集数据,网络招募1298名被试参与研究。其中,195名被试未通过注意力检测题被排除,数据清洗排除67名被试,剩余被试1036人(女性占比63.03%, $M_{\text{年龄}}=22.46$ 岁, $SD_{\text{年龄}}=4.68$ 岁),剔除非大学生被试287名^①,最终有效被试749人(女性占比63.82%, $M_{\text{年龄}}=21.55$ 岁, $SD_{\text{年龄}}=3.49$ 岁)纳入后续统计分析。

2.2 实验材料

实验材料包括三行诗和油画。其中,人类创作的三行诗源自先前研究(He et al., 2022),选取“天空”、“落日”、“旗袍”和“坟”4个主题词创作的作品,共计227首。GAI创作的三行诗由ChatGPT-4.0 Api代理完成,基于相同主题词,提示词为:“你是一位擅长创作三行诗的现代主义诗人,接下来请你以“XX”为主题创作一首三行诗,要求诗歌富有创意、语言含蓄隽永,字数控制在30左右”。程序迭代后每个主题词生成20首作品,并根据字数、

^①本研究中大学生样本是指在读大学生以及获得大学本科学历的在校学生。

格式、明显人工智能痕迹进行初筛。GAI 创作的有效三行诗共计 27 首。随后整合人类和 GAI 创作的三行诗，由 3 名评分者进行评估（包括独创性、情感倾向和情绪感染力三个维度，1~5 评分），区分出高（ $M_{\text{独创性}} = 3.22$, $SD = .27$ ）和低（ $M_{\text{独创性}} = 2.40$, $SD = .28$ ）两种创造性水平的作品。其中，人类创作的高创造性三行诗 88 首，低创造性三行诗 139 首；GAI 生成的高创造性三行诗 20 首，低创造性三行诗 7 首。本研究实际使用的人类和 GAI 创作的高、低创造性三行诗各 4 首。

人类创作的油画来自西南大学美术学院学生作品展，作品主题限定为“风景”、“静物”和“人像”，共计 146 幅作品。GAI 油画由 Midjourney-5.2 生成，创作过程中限定了主题和油画风格，提示词示例为：“Window edge, paintbrush, paint, and greenery, oil painting style of Laurent parcelier, classical style, eye-catching, solarizing master, romantic scenes, delicate in summer sunset, in the style of luminous and dreamlike scenes, delicately rendered landscapes, light-filled landscapes, high detail, super clarity --ar 4:5 --stylize 250 --v 5.2”。GAI 有效油画共计 53 幅。随后，整合人类和 GAI 创作的油画，由 1 名美术学专业教师（主要负责评判 GAI 创作绘画的种类和风格）和 3 名美术学研究生对所有绘画作品进行了筛选、创造性评估和匹配。筛选过程包括剔除临摹名人作品和低质量 GAI 作品，在进行作品匹配时，要求人类和 GAI 创作的油画在创造性等级、主题词、风格和绘画内容上保持一致。在作品评估时，3 名美术学研究生评估了作品在多大程度上像人类创作的（0~5 打分），创造性评估则从独创性、想象力以及精致性三个维度进行 0~3 打分（三个维度分数相加即为作品的创造性得分），区分出高（ $M_{\text{独创性}} = 5.41$, $SD = .40$ ）和低（ $M_{\text{独创性}} = 4.27$, $SD = .44$ ）两种创造性水平的作品，其中人类创作的高创造性油画 108 幅，低创造性油画 38 幅；GAI 创作的高创造性油画 16 幅，低创造性油画 37 幅。本研究实际使用人类和 GAI 创作的高、低创造性油画各 5 幅。

本研究收集了被试的人口学信息，包括年龄、性别、最后学历和所学专业，同时采用人工智能接触程度的自编问题，包括“您是否知道人工智能可以创作图、文、音频、视频等作品？”、“您是否使用过人工智能相关产品（比如 ChatGPT、BingChat、百度文心一言等）？”等共计 8 个条

目。在线实验结束后，被试还需完成大五人格问卷（姚若松，梁乐瑶，2010）以及人工智能态度问卷（Koverola et al., 2022），在本研究中仅作为协变量纳入分析。先前已有研究发现，相较于非 STEM 专业，STEM 专业的个体使用 GAI 更为频繁（Baek et al., 2024）且持有更积极的接受态度（Kim et al., 2025）。基于此，本研究也将专业划分为 STEM 专业和非 STEM 专业，进一步检验大学生群体对 GAI 作品创造性态度是否同样受到专业的调节。专业分类参照国际通用学科分类体系，包含艺术学科、社会科学、人文学科、技术学科、工程学科、数学学科、科学学科以及其他学科共计 8 类。再根据国际上对科学、技术、工程、数学的通用定义，依据学科所属领域及其与科学、技术、工程和数学的关联程度，将其划分为 STEM 与非 STEM 专业两大类。

2.3 实验流程

线上实验共分为五个部分，第一部分填写人口学信息和人工智能接触程度调查，第二部分到第四部分分别为在线实验 1、2 和 3，最后一部分是填写大五人格问卷和人工智能态度问卷。

2.3.1 实验 1

本实验采用 2（作品的真实作者：人类，GAI） \times 2（作品创造性质量：高，低）的被试内实验设计，因变量为被试对作品的创造性评价，旨在探究个体对 GAI 作品的感知和评价。实验中，屏幕中心呈现三行诗或油画（人类与 GAI 创作的高、低创造性作品各 1 首/幅，同种类作品随机呈现），作品的真实作者不予显示。被试需在观看作品后，依据个人感受对其创造性和喜爱程度进行打分（7 点评分）。具体包括：（1）独创性：您认为该作品多大程度上具有独创性？（2）适宜性：假设有一本三行诗诗集正在征稿，您觉得该作品在多大程度上可以被收录？（对于油画作品：假如某地将要举办一场美术展，您认为该作品在多大程度上可以作为展品展出？）（3）喜爱程度：您认为该作品有多受欢迎？本研究将独创性与适宜性的评分相加后取均值，作为作品的创造性得分。

2.3.2 实验 2

本实验采用 2（作品的真实作者：人类，GAI） \times 2（作品创造性质量：高，低）的被试内实验设计，因变量为被试推测作品的作者以及对作品的创造性评价，旨在考察个体对 GAI 作品创作者的推测偏差以及对作品创造性的评价偏差。实验中，屏幕中心

呈现三行诗或油画（人类和 GAI 创作的高、低创造性三行诗各 1 首，油画各 2 幅，同一种类作品随机呈现），作品的真实作者不显示，被试看到作品后，需要推测此作品的创作者是 GAI 还是人类，随后根据自己的感受对作品的创造性和喜爱程度进行评分。

2.3.3 实验 3

本实验采用 2（作品的真实作者：人类，GAI） \times 2（作者外显标签：人类，GAI）的被试内实验设计，因变量为被试对作品的创造性评价，旨在探究个体对 GAI 作品评价偏差的调节因素。实验中，屏幕中心呈现三行诗或油画（人类和 GAI 创作的作品各 4 首/幅，同一种类作品随机呈现），作者外显标签显示在作品上方（三行诗）或下方（油画）。作者外显标签伪随机处理，即每个同组的 2 首/幅作品中，其中 1 首/幅在每个试次下被随机标签为“创作者：人类/GAI”，另 1 首/幅标签为“创作者：GAI/人类”。要求被试看到作品后根据自己的感受对作品的创造性和喜爱程度进行评分。

2.4 数据清洗方法

针对年龄、作答时间和规律性作答进行数据清洗，进一步确保样本的质量和可靠性。考虑到被试学历大多在大学及以上，结合学历和年龄因素，首先排除 18 岁以下的初高中生被试。其后，针对作答时间采用正负三个标准差原则，将超出此范围的数据选择为异常值。结果显示，无异常值出现。针对反应规律性，采用 Shannon 熵作为被试答题规律性的衡量标准。Shannon 熵可以用来衡量数据的随机性和不可预测性，其值越低表示数据越具规律性。数据清洗共排除 262 名被试。

3 研究结果

3.1 数据描述性统计

在对人工智能接触程度的调查中，74.5% 的被试表示知晓 GAI，24.83% 的被试不仅知晓且了解其基本原理，仅有 0.67% 的被试表示不知道人工智能可用于相关领域的创作。大学生群体人工智能接触程度整体处于中等水平（ $M = 2.52$ ， $SD = .66$ ）。

3.2 实验 1 结果

基于实验 1 数据，以作品的真实作者（人类，GAI）和作品创造性质量（高，低）为自变量，以创造性评分为因变量，同时以喜爱程度为协变量，进行两因素重复测量方差分析。结果显示：关于三行诗的创造性评价（图 1A），作品的真实作者与

作品创造性质量的交互作用（ $F(1, 748) = 116.25$ ， $p < .001$ ， $\eta^2 = .03$ ）、作品的真实作者主效应（ $F(1, 748) = 52.36$ ， $p < .001$ ， $\eta^2 = .02$ ）及作品创造性质量的主效应（ $F(1, 748) = 189.62$ ， $p < .001$ ， $\eta^2 = .05$ ）均显著。具体而言，被试对 GAI 生成的三行诗（ 4.32 ± 1.48 ）创造性评分显著高于人类创作的三行诗（ 4.08 ± 1.53 ）；被试对高创造性的作品评分（ 4.56 ± 1.40 ）显著高于低创造性作品的评分（ 3.84 ± 1.54 ）。简单效应分析发现，在高创造性质量条件下，人类创作的三行诗评分显著高于 GAI（ $p < .01$ ）；而在低创造性质量下，人类创作的三行诗评分显著低于 GAI（ $p < .001$ ）。关于油画的创造性评分（图 1B），作品的真实作者的主效应不显著， $F(1, 748) = .01$ ， $p > .05$ ；作品的真实作者与作品创造性质量的交互作用（ $F(1, 748) = 9.61$ ， $p < .01$ ， $\eta^2 = .002$ ）及作品创造性质量的主效应（ $F(1, 748) = 13.36$ ， $p < .001$ ， $\eta^2 = .003$ ）均显著。具体而言，被试对高创造性的油画评分（ 4.67 ± 1.40 ）显著低于对低创造性油画的评分（ 4.73 ± 1.28 ）。简单效应分析发现，在高质量作品条件下，人类创作的油画评分显著低于 GAI 创作的油画（ $p < .01$ ）；在低质量作品条件下，两者没有显著差异（ $p > .05$ ）。总体而言，大学生对 GAI 作品的创造性评分并不低于人类作品，但对作品的创造性感知会受到作品类型和质量的影响。

3.3 实验 2 结果

3.3.1 被试对人类和 GAI 作品的区分能力

基于实验 2 数据，以人类作品为信号，以 GAI 作品为噪音，按作品类别分别根据信号检测论计算辨别力指数 d' 。结果显示，被试对三行诗的辨别力指数为 $d'_{\text{三行诗}} = .23$ ；对油画的辨别力指数为 $d'_{\text{油画}} = -.47$ 。二项检验显示，被试判断三行诗创作者的正确率为 54.64%，显著大于 50%， $p < .001$ ，95%CI = [.52, .57]；被试判断油画创作者的正确率为 40.62%，显著低于 50%， $p < .001$ ，95%CI = [.39, .42]。以上结果表明，被试对三行诗的创作者存在一定区分能力，而对于油画创作者的区分能力低于随机水平。

3.3.2 推测偏差

基于实验 2 的数据，以被试推测作品的作者（人类，GAI）为因变量，对作品的创造性评价得分、喜爱程度评价以及其他变量作为自变量进行多元逻辑回归。由表 1 和表 2 的结果可知，作品创造性得

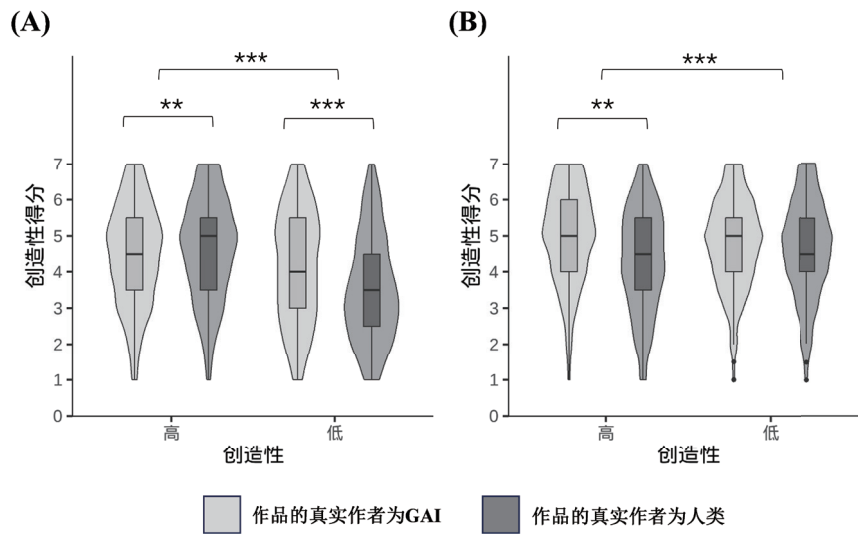


图1 三行诗 (A) 和油画 (B) 作品在作品真实作者和作品质量条件下的创造性得分

分显著可以预测作品的创作者。具体而言，三行诗被推测为“人类”创作相比于被推测为“GAI”创作，其创造性得分显著更高 ($OR = 1.48$)；油画被推测为“人类”创作相比于被推测为“GAI”创作，其创造性得分显著更高 ($OR = 1.63$)。

3.3.3 评价偏差

由于采用推测作者标签时，各组内被试数量不等，因此分别对三行诗和油画构建线性混合效应模型，考察个体对作品的评价是否受到推测作者标签、作品的真实作者及二者交互项的影响。结果显示，作品为三行诗时（图 2A），推测作者标签和作品的真实作者的交互作用不显著， $F(1, 2628) = 3.43, p > .05$ ；推测作者标签的主效应 ($F(1, 2437) = 887.39, p < .001, \eta^2 = .27$) 及作品的真实作者主效应 ($F(1, 2248) = 35.39, p < .001, \eta^2 = .02$) 均显著。具体而言，推测作者为“GAI”的三行诗创造性评分 (3.84

± 1.35) 显著低于推测作者为“人类”的三行诗 (4.99 ± 1.21)；作品的真实作者为“GAI”的三行诗创造性评分 (4.32 ± 1.37) 显著低于真实作者为“人类”的三行诗 (4.63 ± 1.40)。简单效应分析发现，在作品的真实作者为“GAI”的条件下，推测作者为“人类”的三行诗的创造性评分显著低于推测作者为“人类”的三行诗 ($p < .001$)；在作品的真实作者为“人类”的条件下，也得到相同的结果 ($p < .001$)。作品为油画时（图 2B），作品的真实作者主效应不显著， $F(1, 5248) = 2.13, p > .05$ ；推测作者标签和作品的真实作者的交互作用 ($F(1, 5428) = 12.13, p < .001, \eta^2 = .002$) 及推测作者标签的主效应 ($F(1, 5434) = 1446.09, p < .001, \eta^2 = .21$) 均显著。具体而言，推测作者为“GAI”的油画创造性得分 (3.91 ± 1.41) 显著低于推测作者为“人类”的油画 (4.99 ± 1.23)。简单效应分析发现，在作品的真实作者

表1 二元逻辑回归检验结果（三行诗），推测作者标签（0为GAI，1为人类）

变量	回归系数	SE	OR	95%CI	Wald	p
截距	-1.40	.42	.25	[.11, .56]	11.24	.001
性别	-.04	.09	.96	[.80, 1.15]	.21	.649
年龄	-.02	.01	.98	[.96, 1.00]	2.88	.089
作品质量	.04	.08	1.04	[.89, 1.22]	.21	.649
作品喜爱程度	.33	.05	1.39	[1.27, 1.52]	52.60	< .001
作品创造性得分	.39	.05	1.48	[1.34, 1.64]	55.97	< .001
专业	-.02	.09	.98	[.82, 1.16]	.07	.790
作品真实作者标签	.23	.08	1.25	[1.07, 1.47]	7.61	.006
人工智能态度	-.06	.04	.94	[.87, 1.03]	1.91	.167
开放性人格	-.24	.08	.78	[.68, .91]	10.53	.001

表2 二元逻辑回归检验结果(油画), 推测作者标签(0为GAI, 1为人类)

变量	回归系数	SE	OR	95%CI	Wald	P
截距	-1.13	.29	.32	[.18, .58]	15.21	<.001
性别	-.09	.07	.91	[.80, 1.04]	1.97	.160
年龄	-.00	.01	1.00	[.98, 1.02]	.09	.762
作品质量	.15	.06	1.16	[1.03, 1.30]	6.415	.011
作品喜爱程度	.16	.03	1.17	[1.10, 1.24]	29.15	<.001
作品创造性得分	.49	.03	1.63	[1.52, 1.74]	207.16	<.001
专业	.02	.06	1.02	[.91, 1.15]	.12	.730
作品真实作者标签	-.77	.06	.46	[.41, .52]	178.77	<.001
人工智能态度	-.09	.03	.91	[.86, .97]	9.87	.002
开放性人格	-.23	.05	.80	[.72, .88]	19.22	<.001

为“GAI”的条件下, 推测作者为“GAI”的油画创造性评分显著低于推测作者为“人类”的油画 ($p < .001$); 在作品的真实作者为“人类”的条件下, 也得到相同的结果 ($p < .001$)。以上结果表明, 对于三行诗和油画, 无论真实创作者是GAI还是人类, 若被认为创作者是GAI, 其创造性评分显著偏低。

此外, 本研究考察了推测偏差和评价偏差之间的相关性。具体而言, 在个体水平上, 以人类和GAI作品创造性得分均值之差作为评价偏差指标, 以将高创造性作品推测为“人类”创作超过50%的数量与低创造性作品推测为“GAI”创作超过50%的数量之和作为推测偏差指标。结果发现, 对于三行诗而言, 推测偏差和评价偏差没有显著相关, $r = .02$, $t = .43$, $p > .05$; 而对于油画而言, 两者显著正相关, $r = .34$, $t = 9.77$, $p < .001$ 。

3.4 实验3结果

3.4.1 评价偏差

基于实验3数据分析作者外显标签评价偏差, 针对三行诗和油画, 以作品的真实作者(人类, GAI)和作者外显标签(人类, GAI)作为自变量, 创造性评分作为因变量, 以喜好程度评分作为协变量, 进行两因素重复测量方差分析。结果显示, 作品为三行诗时(图3A), 作者外显标签和作品的真实作者的交互作用 ($F(1, 748) = 4.32$, $p < .05$, $\eta^2 < .001$)、作品的真实作者主效应 ($F(1, 748) = 73.95$, $p < .001$, $\eta^2 = .02$)及作者外显标签的主效应 ($F(1, 748) = 67.51$, $p < .001$, $\eta^2 = .01$)均显著。具体而言, 真实作者为“GAI”的三行诗创造性评分 (4.58 ± 1.45)显著高于“人类”三行诗 (4.17 ± 1.54); 作者外显标签为“GAI”的三行诗创造性评分 (4.08

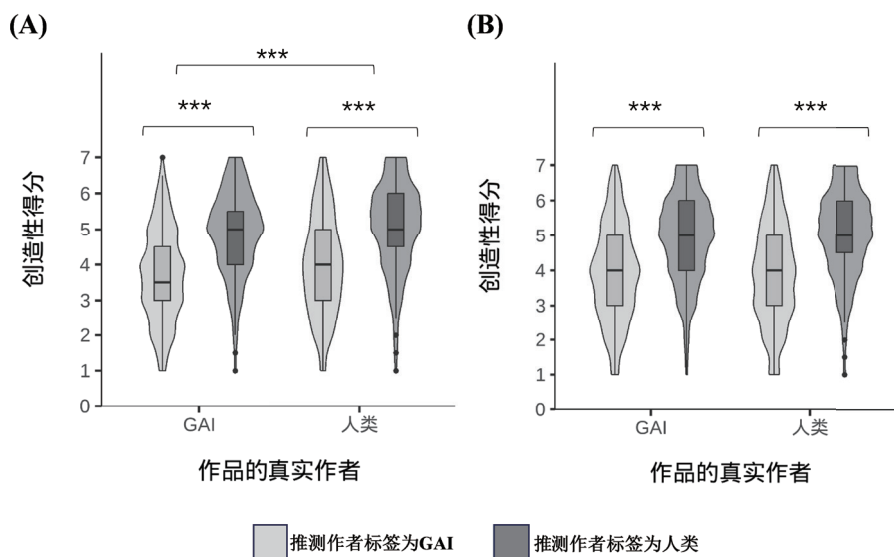


图2 三行诗(A)和油画(B)作品在作品真实作者和被试推测作者下的创造性评分

± 1.51) 显著低于作者外显标签为“人类”的三行诗 (4.67 ± 1.46)。简单效应分析发现, 在 GAI 创作者条件下, 作者外显标签为“GAI”的三行诗的创造性评分显著低于作者外显标签为“人类”的三行诗 ($p < .001$); 在人类创作者条件下, 作者外显标签为“GAI”的三行诗创造性评分同样显著低于作者外显标签为“人类”的三行诗 ($p < .001$)。当作品为油画时(图 3B), 作者外显标签和作品的真实作者的交互作用 ($F(1, 748) = 6.24, p < .05, \eta^2 = .001$)、作品的真实作者主效应 ($F(1, 748) = 7.06, p < .01, \eta^2 = .001$) 及作者外显标签的主效应 ($F(1, 748) = 81.25, p < .001, \eta^2 = .02$) 均显著。具体而言, 真实作者为“GAI”的油画创造性评分 (4.43 ± 1.47) 显著低于“人类”油画 (4.50 ± 1.52); 作者外显标签为“GAI”的油画创造性评分 (4.14 ± 1.50) 显著低于作者外显标签为“人类”的油画 (4.80 ± 1.42)。简单效应分析发现, 在 GAI 创作者条件下, 作者外显标签为“GAI”的油画创造性评分显著低于作者外显标签为“人类”的油画 ($p < .001$); 在人类创作者条件下, 作者外显标签为“GAI”的油画创造性评分显著低于作者外显标签为“人类”的油画 ($p < .001$)。以上结果表明, 对于三行诗和油画, 无论其真实创作者是 GAI 还是人类, 只要作者外显标签为“GAI”, 其创造性评分显著偏低。

3.4.2 评价偏差的调节因素

基于实验 3 的数据, 以创造性评分为因变量, 作者外显标签、接触程度、专业为自变量。人工智能接触程度对三行诗作品创造性评分的调节作用(图 4A) 结果显示, 人工智能接触程度和作者外显标

签的交互作用不显著, $F(1, 5986) = 3.70, p > .05$; 人工智能接触程度的主效应 ($F(1, 5986) = 10.43, p < .01, \eta^2 = .002$)、作者外显标签的主效应 ($F(1, 5986) = 438.08, p < .001, \eta^2 = .07$) 均显著。简单效应分析发现, 在低接触程度条件下 ($< -1.5SD$), 作者外显标签为“GAI”的三行诗创造性评分显著低于标签为“人类”的三行诗 ($p < .001$); 在高接触程度条件下 ($> 1.5SD$), 两种条件仍然存在显著差异 ($p < .001$)。在“人类”作者外显标签条件下, 人工智能的接触程度对三行诗的创造性评分无显著影响, $\beta = .02, SE = .02, t = -.92, 95\%CI = [-.02, .05]$; 而在“GAI”作者外显标签条件下, 创造性评分随着人工智能接触程度的增加而增加, $\beta = .06, SE = .02, t = 3.65, 95\%CI = [.03, .10]$ 。

对于专业的调节作用(图 4B), 结果显示, 专业和作者外显标签的交互效应显著, $F(1, 5986) = 9.16, p < .01, \eta^2 = .002$; 专业的主效应不显著, $F(1, 5986) = .78, p > .05$ 。简单效应分析发现, 在“GAI”作者外显标签条件下, STEM 专业被试给出的三行诗创造性评分显著高于非 STEM 专业被试 ($p < .01$); 而在“人类”作者外显标签条件下, 两者无显著差异 ($p > .05$)。综上显示, 对 GAI 作品的评价偏差受到个体与人工智能接触程度以及专业背景的调节, 但这些因素的调节作用较为微弱。

4 讨论

本研究以人类与 GAI 的创作物为实验材料, 通过三项线上行为实验考察大学生对 GAI 作品的创造性感知与评价。实验 1 在不告知作品作者的条件下,

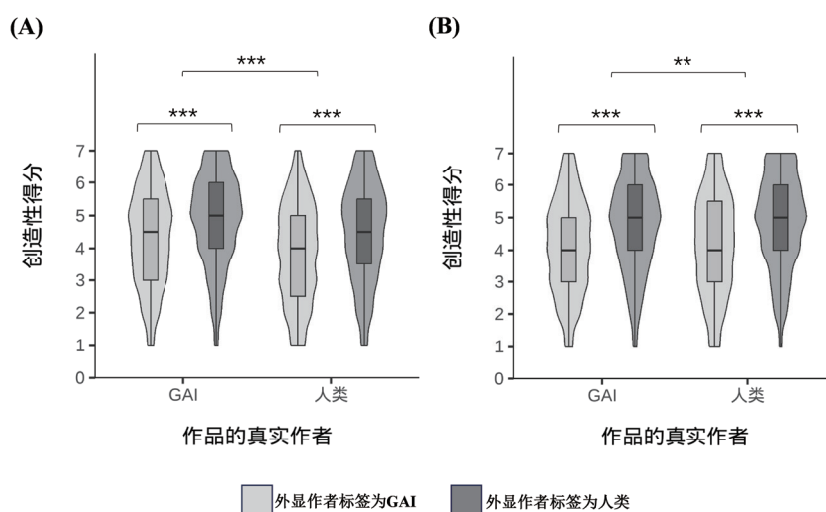


图 3 三行诗 (A) 和油画 (B) 作品在作品真实作者和作者外显标签下的创造性评分

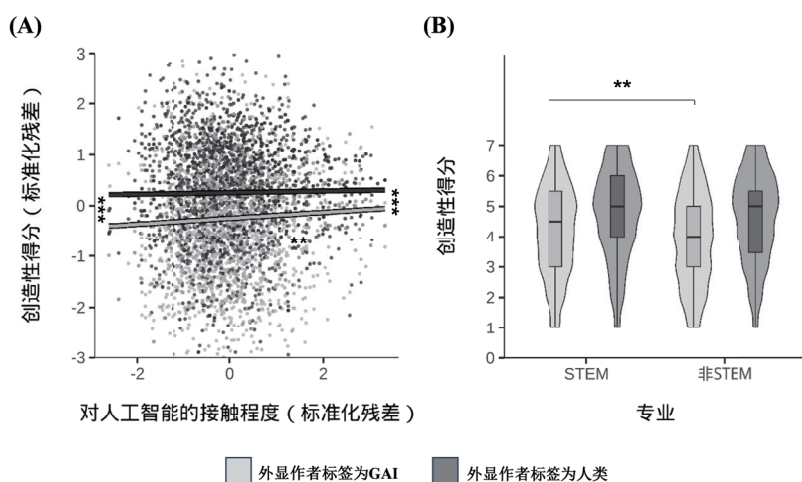


图4 人工智能接触程度(A)和专业(B)对三行诗创造性评分的调节

将人类与GAI创作的三行诗和油画混合呈现,要求被试主观评价作品的创造性水平。结果显示,GAI作品的创造性评分与人类作品无显著差异,但受到作品类型和质量的影响。实验2要求被试先对作品的作者身份进行推测,再对其创造性进行评分。结果发现,被试更倾向将高创造性作品的作者推测为“人类”,并对作者被推测为“人类”的作品给出更高的创造性评分,表现出推测偏差与评价偏差。实验3随机标注作品作者,要求被试进行评分,结果显示评价偏差依然存在。最后,考察人工智能接触程度和专业背景是否调节评价偏差。结果发现接触程度越高、专业为STEM的被试对GAI作品的创造性评价偏差更小。

基于千亿级参数规模的神经网络模型,GAI不仅能够探索潜在概念空间并实现概念的新颖组合,还具备复杂的策略生成、反馈调节与评价优化机制。尤为重要的是,GAI表现出对上下文语义的深度理解能力,可生成具有独创性和启发性的内容,使其在创作过程上与人类存在相似性,并在作品中展现出具有人类特征的创造性水平。本研究发现,对于油画作品,相比于高质量的人类油画,被试对高质量GAI油画的创造性评价更高。这一结果可能源于GAI通过大规模人类作品训练数据所提炼出的符合大众审美偏好的油画特征,使其产品在“迎合”普通人审美需求方面具有优势,从而被认为更具吸引力或创造性(Epstein et al., 2020; Hertzmann, 2018)。相反,相较于高质量的GAI三行诗,被试对人类创作的高质量三行诗的创造性评价更高,这可能是由于言语作品高度依赖语言的独特性、文化

背景和语境,尤其是高质量的言语作品对情感共鸣或深层次的文化内涵有更高要求。本研究结果表明,大学生在评价不同类型的GAI作品时表现出差异化的创造性感知倾向,但在创造性整体层面尚无法准确区分人类和GAI作品。

关于人类对人工智能的态度已有大量研究,其中算法厌恶(Burton et al., 2020; Dietvorst et al., 2015)现象在本研究中得到进一步验证。本研究发现,与标签为“人类”创作的作品相比,被试对标签为“GAI”创作的作品表现出推测偏差和评价偏差。虽然赋予人工智能人性化特征被证实可缓解算法厌恶(Ahn et al., 2021; Epstein et al., 2020),但本研究发现人类对GAI作品的评价偏差依然存在,该发现在一定程度上拓展了现有算法厌恶的内涵。与专用人工智能引发的“机器人”式威胁感知不同,GAI的艺术创作尚未对大众产生直接冲击,因此,这种认知偏差可能源于GAI对人类中心主义观念的挑战。

从人类中心主义视角分析,个体对GAI艺术作品的评价偏差可归因于认知偏差与情感亲近度等因素(Millet et al., 2023)。研究表明,当GAI与人类创作的作品难以区分时,个体会降低“创造性”这一人类特有能力的价值,转而关注GAI的相对劣势维度(周详等, 2024)。例如,GAI的算法快速生成属性会强化个体在评价中对创作时间成本的关注,使其倾向给予人类而非GAI作品更高评价(Magni et al., 2024)。此外,个体更倾向赋予人类作品更高的情感价值,并与人类创作者建立情感共鸣。最新研究还发现,相比“AI创作者”标签,个体对“人类创作者”标签的绘画有更强的情感信息认知加工

活动 (Zhang et al., 2025)。因此,在美学特征、创造性和价值评估等方面,人类作品可能获得更高评价。这一现象表明个体在界定创造性时,更为重视创作过程而非仅关注结果输出。尽管 GAI 展现出拟人化的创造性特征,但从科学技术哲学视角来看,其创作过程缺乏主体思维和人格的物化体现 (陈凡, 吴怡, 2021); 从艺术创作视角来看,亦缺乏真实的情感联结与共情属性 (郭超等, 2019)。因此, GAI 作品难以与人类作品相提并论。

本研究结果进一步证实,个体对 GAI 作品创造性的评价会受到其人工智能接触程度和专业背景的调节。已有研究发现,人工智能接触程度的差异会导致其对 GAI 作品不同的态度与行为模式 (van der Kaa & Kraemer, 2014), 接触人工智能越多、越熟悉的个体对其评价更为积极 (Hong & Curran, 2019)。同时, GAI 的拟人化特征也有助于提升个体对其的积极评价 (Li & Sung, 2021)。此外,个体对人工智能的态度也是影响评价的重要因素: 持有强烈人类中心主义观念或“人工智能无法产生艺术”等刻板印象的群体,往往表现出更强的评价偏差 (Hong & Curran, 2019); 而态度积极的个体更倾向于认可人工智能,并对人与人工智能的合作持开放态度。值得注意的是,本研究发现非 STEM 专业学生对 GAI 创造性的评价显著低于 STEM 专业学生,这可能与实验所用的 GAI 艺术创作材料更贴近非 STEM 专业学生的就业领域,从而引发相对强烈的职业风险感知有关 (Wissing & Reinhard, 2018), 进而低估 GAI 作品的创造性。

本研究仍存在一些不足。首先,需要对作品作者的主体界定制定更为明确的标准。例如,实验使用的 GAI 油画作品在生成过程中需依赖人类提供较为精细的提示词,且在材料筛选与匹配阶段也需要人工参与,因此这些油画作品在一定程度上应被视为人类与 GAI 协作完成的产物。类似地, GAI 生成的三行诗同样依赖人类对模型的预训练,以避免生成的内容呈现明显的人工智能痕迹。因此,两类 GAI 作品均非完全自动化生成,未来的研究可进一步考虑 GAI 自身的创造性能及其在创作过程中的参与程度。例如,基于创作过程从完全自动化到完全人工创作的连续谱 (Tigre Moura et al., 2023), 建立人机协作程度的标准化评估体系,以便更系统地研究人类与 GAI 作品在感知层面的特征和评价偏差,并探讨 GAI 在协同创作过程中对人类创造力的潜在

影响 (Habib et al., 2024; Messer, 2024)。

其次,随着 GAI 技术的飞速发展,当前研究结论在泛化时存在时效局限性。一方面,随着 GAI 版本迭代和推理模型升级,其输出作品的拟人性和创造性水平逐渐提升。另一方面,大型模型技术的开源化 (如 Deepseek 等) 及产品普及正在改变着公众对 GAI 作品的认知态度 (Hansen et al., 2024; Koenig, 2024)。未来研究应持续追踪 GAI 技术的演进路径以及由此引发的公众对其在感知与态度方面的动态变化。

再者,本研究的三项实验结果显示,作品真实作者的主效应存在异质性。这一结果可能源于人类作品与 GAI 作品在多个属性维度上未能实现有效匹配。例如,在筛选 GAI 作品时,虽然我们剔除了具有明显人工智能痕迹的作品,但未对呈现显著“人类痕迹”的作品进行同等处理,由此可能产生系统性偏差。此外,作品评估过程中显示的评分者同质性较高也在一定程度上影响了实验材料的生态效度。未来研究应建立系统的作品属性评估标准,确保人类与 GAI 作品在诸如风格一致性、复杂性和情感表达等关键指标上严格匹配,从而提高实验的信效度。

最后,本研究主要考察了与人工智能使用经验相关的调节变量,尚未探究个体其它因素的影响,例如性别、人格特质、擅长的创意领域以及个体创造性水平等。此外,这些因素如何交互影响个体对 GAI 作品的创造性感知与态度也亟待深入探究。

5 结论

本研究通过三项实验发现,大学生对 GAI 作品的创造性评分与人类作品无显著差异,但对作品的创造性感知会受到作品类型和质量的影响; 大学生对推测或标签为“GAI”的作品创造性评价更低,表现出明显的评价偏差; 大学生对 GAI 作品的评价偏差受到其与人工智能的接触程度以及专业背景的调节。

参考文献

- 陈凡, 吴怡. (2021). 人工“智”能的智慧, 机智与明智. *自然辩证法通讯*, 43(12), 95-100.
- 郭超, 鲁越, 林懿伦, 卓凡, 王飞跃. (2019). 平行艺术: 人机协作的艺术创作. *智能科学与技术学报*, 1(4), 335-341.
- 姚若松, 梁乐瑶. (2010). 大五人格量表简化版 (NEO-FFI) 在大学生人群的应用分析. *中国临床心理学杂志*, 18(4), 457-459.
- 衣新发, 林崇德, 蔡曙山, 黄四林, 陈枕, 罗良, 唐敏. (2011). 留学经验与艺术创造力. *心理科学*, 34(1), 190-195.
- 周详, 祖冲, 崔虞馨. (2024). *创造力与人工智能*. 陕西师范大学出版总社.
- Ahn, J., Kim, J., & Sung, Y. (2021). AI-powered recommendations: The roles of

- perceived similarity and psychological distance on persuasion. *International Journal of Advertising*, 40(8), 1366–1384.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43(5), 997–1013.
- Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review*, 55(1), 589–656.
- Baek, C., Tate, T., & Warschauer, M. (2024). “ChatGPT seems too good to be true”: College students’ use and perceptions of generative AI. *Computers and Education: Artificial Intelligence*, 7, Article 100294.
- Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33(1), 63.
- Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications*, 8(1), 42.
- Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239.
- Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts*, 12(2), 177–192.
- Cheng, M. (2022). The creativity of artificial intelligence in art. *Proceedings*, 81(1), 110.
- Chiarella, S. G., Torromino, G., Gagliardi, D. M., Rossi, D., Babiloni, F., & Cartocci, G. (2022). Investigating the negative bias towards artificial intelligence: Effects of prior assignment of AI-authorship on the aesthetic appreciation of abstract paintings. *Computers in Human Behavior*, 137, Article 107406.
- Dang, J., & Liu, L. (2021). Robots are friends as well as foes: Ambivalent attitudes toward mindful and mindless AI robots in the United States and China. *Computers in Human Behavior*, 115, Article 106612.
- Demmer, T. R., Kühnapfel, C., Fingerhut, J., & Pelowski, M. (2023). Does an emotional connection to art really require a human artist? Emotion and intentionality responses to AI-versus human-created art and impact on aesthetic experience. *Computers in Human Behavior*, 148, Article 107875.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126.
- Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). Can: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms. *ArXiv*.
- Epstein, Z., Levine, S., Rand, D. G., & Rahwan, I. (2020). Who gets credit for AI-generated art? *iScience*, 23(9), Article 101515.
- Fietta, V., Zecchinato, F., Di Stasi, B., Polato, M., & Monaro, M. (2021). Dissociation between users' explicit and implicit attitudes toward artificial intelligence: An experimental study. *IEEE Transactions on Human-Machine Systems*, 52(3), 481–489.
- Gangadharbatla, H. (2022). The role of AI attribution knowledge in the evaluation of artwork. *Empirical Studies of the Arts*, 40(2), 125–142.
- Grassini, S., & Koivisto, M. (2024). Understanding how personality traits, experiences, and attitudes shape negative bias toward AI-generated artworks. *Scientific Reports*, 14(1), 4113.
- Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity*, 34(1), Article 100072.
- Hansen, H. F., Lillesund, E., Mikalef, P., & Altwayiry, N. (2024). Understanding artificial intelligence diffusion through an AI capability maturity model. *Information Systems Frontiers*, 26(6), 2147–2163.
- He, R., Zhuang, K., Liu, L., Ding, K., Wang, X., Fu, L., Qiu, J., & Chen, Q. (2022). The impact of knowledge on poetry composition: An fMRI investigation. *Brain and Language*, 235, Article 105202.
- Hertzmann, A. (2018, May). Can computers create art? *Arts*, 7(2), 18.
- Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2023). Does human-AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior*, 139, Article 107502.
- Hong, J. W., & Curran, N. M. (2019). Artificial intelligence, artists, and art: Attitudes toward artwork produced by humans vs. artificial intelligence. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 15(2s), 1–16.
- Hong, J. W., Fischer, K., Ha, Y., & Zeng, Y. (2022). Human, I wrote a song for you: an experiment testing the influence of machines' attributes on the AI-composed music evaluation. *Computers in Human Behavior*, 131, Article 107239.
- Hubert, K., Awa, K. N., & Zabelina, D. (2024). Artificial intelligence is more creative than humans: A cognitive science perspective on the current state of generative language models. *Scientific Reports*, 14, 3440.
- Johnson, W., & Proudfoot, D. (2024). Greater variability in judgements of the value of novel ideas. *Nature Human Behaviour*, 8(3), 471–479.
- Kaya, F., Aydin, F., Schepman, A., Rodway, P., Yetişenoy, O., & Demir Kaya, M. (2024). The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence. *International Journal of Human-Computer Interaction*, 40(2), 497–514.
- Kim, J., Klopfer, M., Grohs, J. R., Eldardiry, H., Weichert, J., Cox, L. A., & Pike, D. (2025). *Examining faculty and student perceptions of generative AI in university courses*. *Innovative Higher Education*. Advance online publication..
- Koenig, P. D. (2024). Attitudes toward artificial intelligence: Combining three theoretical perspectives on technology acceptance. *AI and Society*, 40(3), 1–13.
- Koverola, M., Kunnari, A., Sundvall, J., & Laakasuo, M. (2022). General attitudes towards robots scale (GAToRS): A new instrument for social surveys. *International Journal of Social Robotics*, 14(7), 1559–1581.
- Latikka, R., Savela, N., & Oksanen, A. (2023). Perceptions of assistive robots at work: An experimental approach to social influence. *International Journal of Social Robotics*, 15(9), 1543–1555.
- Li, X., & Sung, Y. (2021). Anthropomorphism brings us closer: The mediating role of psychological distance in User-AI assistant interactions. *Computers in Human Behavior*, 118, Article 106680.
- Magni, F., Park, J., & Chao, M. M. (2024). Humans as creativity gatekeepers: Are we biased against AI creativity? *Journal of Business and Psychology*, 39(3), 643–656.
- Medeiros, K., Cropley, D. H., Marrone, R. L., & Reiter-Palmon, R. (2025). Human-AI co-creativity: Does ChatGPT make us more creative? *The Journal of Creative Behavior*, 59(2), e70022.
- Messer, U. (2024). Co-creating art with generative artificial intelligence: Implications for artworks and artists. *Computers in Human Behavior: Artificial Humans*, 2(1), Article 100056.

- Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior, 143*, Article 107707.
- Mokyr, J., Vickers, C., & Ziebarth, N. L. (2015). The history of technological anxiety and the future of economic growth: Is this time different? *Journal of Economic Perspectives, 29*(3), 31–50.
- Park, J., & McGee, R. (2023). Who, or rather, what painted this? Generation Z's attitudes towards artificial intelligence artworks. *Journal of Student Research, 12*(4), Article 5804.
- Ragot, M., Martin, N., & Cojean, S. (2020). *AI-generated vs. human artworks. A perception bias towards artificial intelligence?* Extended abstracts of the 2020 CHI conference on human factors in computing systems, Association for Computing Machinery.
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96.
- Schubert, E., Canazza, S., De Poli, G., & Rodà, A. (2017). Algorithms can mimic human piano performance: The deep blues of music. *Journal of New Music Research, 46*(2), 175–186.
- Shank, D. B., Stefanik, C., Stuhlsatz, C., Kacirek, K., & Belfi, A. M. (2023). AI composer bias: Listeners like music less when they think it was composed by an AI. *Journal of Experimental Psychology: Applied, 29*(3), 676–692.
- Tigre Moura, F., Castrucci, C., & Hindley, C. (2023). Artificial intelligence creates art? An experimental investigation of value and creativity perceptions. *The Journal of Creative Behavior, 57*(4), 534–549.
- Ting, T. T., Ling, L. Y., Azam, A. I. B. A., & Palaniappan, R. (2023). *Artificial intelligence art: Attitudes and perceptions toward human versus artificial intelligence artworks*. AIP Conference Proceedings.
- Tubadji, A., Huang, H., & Webber, D. J. (2021). Cultural proximity bias in AI-acceptability: The importance of being human. *Technological Forecasting and Social Change, 173*, Article 121100.
- Turing, A. M. (2007). Computing machinery and intelligence. In R. Epstein, G. Roberts, & G. Beber (Eds.) *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 23–65). Springer Netherlands.
- Van der Kaa, H. A., & Kraemer, E. J. (2014). *Journalist versus news consumer: The perceived credibility of machine written news*. *Proceedings of the Computation + Journalism Conference*, New York, NY, United States.
- Van Hees, J., Grootswagers, T., Quek, G. L., & Varlet, M. (2025). Human perception of art in the age of artificial intelligence. *Frontiers in Psychology, 15*, Article 1497469.
- Wang, C. (2024). Art innovation or plagiarism? Chinese students' attitudes toward AI painting technology and influencing factors. *Ieee Access, 12*, 85795–85805.
- Wissing, B. G., & Reinhard, M. A. (2018). Individual differences in risk perception of artificial intelligence. *Swiss Journal of Psychology, 77*(4), 149–157.
- Zhang, W., Xie, C., Jiang, L., Yang, L., Hu, Z., & Hao, N. (2025). Neural correlates of evaluative bias against AI-labeled versus human-labeled artworks. *Social Cognitive and Affective Neuroscience*. Advance online publication.
- Złotowski, J., Yogeewaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies, 100*, 48–54.

Perceived Creativity and Evaluation Bias Toward Generative AI Artworks among University Students

Xiong Xiaoyan¹, Hou Jialin¹, Ding Youyin², Xiang Xuli², He Dong², Long Haiying³, Chen Qunlin¹

(¹ Faculty of Psychology, Southwest University, Chongqing, 400715)

(² School of Fine Arts, Southwest University, Chongqing, 400715) (³ School of Education and Human Sciences, The University of Kansas, KS 66045)

Abstract Creativity has long been regarded as one of the advanced cognitive abilities unique to humans. Within the realm of creativity research, the perception of creativity in GAI-generated works and individuals' attitudes to GAI creativity, particularly when these creations are judged to have creative characteristics, are key issues worthy of in-depth exploration. This study focuses on the artistic products generated by GAI, specifically three-line poems and oil paintings, and uses three progressive experiments to investigate individual perceptions and attitudes toward GAI creativity among college students. Additionally, the study explores key factors moderating human evaluations of GAI creations, aiming to provide empirical evidence to support the widespread adoption, evaluation, and societal utility of GAI.

In Experiment 1, a 2 (products' actual author: human or GAI) × 2 (product quality: high or low) within-subjects design was employed.

Participants' likability of the products was considered as a covariate, and the perceived creativity of the products served as the dependent variable. During the experiment, participants viewed four three-line poems and four oil paintings, all created by humans or GAI. For each type of author, there was 1 high-creativity piece and 1 low-creativity piece. These works were presented in a random order within each type. The actual authorship of the works was hidden, and participants were asked to rate the works on a 7-point Likert scale for the creativity (obtained by calculating the average value of originality and appropriateness scores) and likability based on their subjective impressions. The results showed that the creative score of the three-line poems created by GAI was higher than that of those created by humans, but the main effect of the actual author of oil paintings was not significant. Overall, college students rated GAI's works as equally creative as human-created works, though this rating was affected by the type and quality of the works.

In Experiment 2, a 2 (speculated author label: human or GAI) \times 2 (product quality: high or low) within-subjects design was utilized. The dependent variables consisted of participants' inferred authorship of the works and their perceptual evaluations (creativity and likability). During the experiment, participants were presented with four three-line poems and eight oil paintings. Among human-created works, there was 1 high-creativity and 1 low-creativity three-line poem, as well as 2 high-creativity and 2 low-creativity oil paintings; the categories and quantities of works created by GAI were consistent with those by humans. All works were displayed in a random order within each category (three-line poems and oil paintings). Without revealing the actual authorship, participants were asked to speculate on the authorship of each piece (Human or GAI) after viewing the works. Subsequently, participants rated the works on a 7-point Likert scale for creativity and likability. The results indicated that participants had moderate accuracy in distinguishing the creators of the three-line poems (54.64% accuracy), whereas their ability to identify the creators of the oil paintings (40.62% accuracy) was below chance level, suggesting an inability to accurately identify authorship. Whether they are three-line poems or oil paintings, the perceived score of the works created by the speculated author for "GAI" are always lower than those of the works created by the speculated author for "Human". The above results show that for the three-line poems and oil paintings, whether the actual creator is GAI or humans, as long as the participants think the creator is GAI, their creativity ratings are lower. It indicates that people have speculative bias and evaluative bias on the creativity of GAI works.

In Experiment 3, a 2 (products' actual author: human or GAI) \times 2 (exoteric author label: human or GAI) within-subjects design was employed. Participants' likability of the products was again considered as a covariate, and the perceived creativity of the products was the dependent variable. In this experiment, participants sequentially viewed eight three-line poems and eight oil paintings. The three-line poems and oil paintings were presented randomly within their respective categories. The exoteric author labels were displayed either at the top (for three-line poems) or at the bottom (for oil paintings) of the works. These labels were pseudo-randomized: within each pair of works from the same condition (the types and quality of the works are the same), one was randomly labeled as "Creator: Human/GAI", while the other received the opposite label (Creator: GAI/Human). The remaining procedure adhered to the steps of Experiment 1. The results show that for both three-line poems and oil paintings, regardless of whether their actual author are GAI or humans, if the exoteric author label is "GAI", their creativity ratings are lower. This suggests that evaluative bias persists even when authorship is explicitly disclosed. Furthermore, the evaluative bias was moderated by participants' exposure to AI and their professional background. Specifically, evaluative bias was reduced in participants with higher levels of exposure to AI and those with a background in STEM fields.

In summary, this study highlights the perceptual characteristics of GAI creativity, the biases in college students' perceptions of GAI-generated works, and the potential factors moderating these biases. The findings offer valuable insights into human-AI interactions and contribute to the advancement of human-computer collaboration and innovative behavior in the AI era.

Key words generative artificial intelligence, creativity, speculative bias, evaluative bias