引用格式：傅小兰，严正. 心智与算法的对话：心理学与人工智能的深度共生. 中国科学院院刊, 2026, 41(5): 1040-1054, doi: 10.3724/j.issn.1000-3045.20260402003.
Fu X L, Yan Z. Dialogue between mind and algorithm: Deep symbiosis of psychology and artificial intelligence. Bulletin of Chinese Academy of Sciences, 2026, 41(5): 1040-1054, doi: 10.3724/j.issn.1000-3045.20260402003. (in Chinese)

心智与算法的对话：心理学与人工智能的深度共生

傅小兰[1]　严　正[2]

1 上海交通大学　心理学院　上海　200030
2 美国纽约州立大学奥尔巴尼分校　教育与咨询心理系　奥尔巴尼　12222

摘要　随着人工智能（AI）由辅助性工具演变为协同性伙伴，心理学与AI的交叉融合也逐步从单向应用迈向深度共生。文章论述二者交互的双向赋能、面临挑战与突破路径。一方面，心理学赋能AI，通过认知建模与伦理约束提升其类人智能与社会适应性；另一方面，AI赋能心理学，借助多模态数据与算法模型革新心理评估及干预手段。二者的深度共生需清醒直面张力、审慎划定边界、坦然承认风险。未来研究应在认知互补、情感共振与伦理对齐中，让AI成为人类深化自我认知的镜子，并构建人与智能体共鸣的创新生态，在差异中实现技术与人文的更高阶协同。

关键词　心理学，人工智能，双向赋能，深度共生，人智交互

DOI　10.3724/j.issn.1000-3045.20260402003
CSTR　32128.14.CASbulletin.20260402003

当AlphaGo（阿尔法狗）落下颠覆棋局的一子，当DeepSeek（深度求索）用流畅的语言回应人类的思绪，人工智能（AI）已悄然跨越纯粹工具的边界，开始触碰人类智慧的内核。心理学与AI的关系正经历深刻变革——从“工具性使用”迈向“深度共生”。这不仅是技术层面的演进，更是心理学研究的人类智能与计算机科学研发的人工智能这2种智能系统间的双向建构。计算机科学曾为心理学提供“信息加工”的认知隐喻，如今心理学正为AI注入“意义建构”的理论框架。这一互动折射出这2个学科长达半个世纪的相互“借用”。1960年，Licklider[1]曾前瞻性地提出：“人机共生是人类与电子计算机协作互动中预期的发

资助项目：国家重点研发计划重点专项（2024YFC3606800），上海交通大学“交大2030”计划B类项目（WH510254101）
修改稿收到日期：2026年5月5日

展方向”。2025年，林崇德[2]进一步指出：“人工智能时代的核心社会变革，既非生产效率的跃升，也非劳动力的简单替代，而是人类与AI‘协同共创关系’的诞生。”

然而，这条“心智与算法的对话”之路并非坦途，其深层张力日益凸显。一方面，算法试图通过数据流动来还原甚至解构人类心智的复杂过程；另一方面，人类心智所固有的主观性、情感性与模糊性，又不断冲击着算法逻辑的确定性边界。当AI直面伦理抉择的困境、共情能力的匮乏，以及决策过程的“黑箱”难题时，心理学所提供的已不仅是创新的灵感源泉，更成为校准技术航向的价值坐标。本文梳理心理学与AI的交互，剖析当前面临的挑战，探讨实现突破的可能路径，旨在说明只有当算法内化心智的维度，当机器领悟人类的温度，AI才能跨越工具的鸿沟，真正成为人类智慧的同路人。

1 心理学与AI的交互

心理学与AI的交互是一个充满活力且快速发展的前沿交叉领域，两者相互促进，形成了一种共生关系。简言之，心理学为AI提供“人性化的灵感”，而AI则为心理学提供“智能化的工具”。

1.1 心理学赋能AI，让AI更“懂人心”

作为一门探究心智与行为的学科，心理学为AI提供了感知、记忆、决策和学习的底层逻辑。从早期认知心理学的信息加工模型[3-5]，到如今受人类神经网络启发的深度学习架构[6-8]，AI的每一次跃迁，几乎都伴随着我们对心智机制理解的深化。

（1）心理学启发新的算法和架构。从脑科学的最基本法则，到赋予机器学习能力，再到模拟人类对话，心理学为人工智能的诞生与发展注入了最初的核心驱动力。心理学家Hebb[9]于1949年提出大脑学习机制的著名假说——赫布定律（Hebbian Theory），为后来几乎所有的人工神经网络提供了最重要的生物学和理论基础，尤其是为无监督学习和权重调整机制奠定了思想基石；心理学家Rosenblatt[10]于1958年发明了感知机（perceptron）；20世纪80年代，心理学家McClelland和Rumelhart[11]提出并行分布式处理（PDP）框架，并推广反向传播（backpropagation）算法，奠定了现代深度学习的基石。而强化学习[12,13]的核心原理（试错、奖励、惩罚）直接来源于行为主义心理学家Thorndike[14,15]和Skinner[16,17]对动物和人类学习过程的研究。心理学的注意力机制也被引入深度学习以解决机器翻译问题[18]，提出仅基于注意力机制的Transformer架构[19]，进而彻底改变AI整个领域。

（2）心理学提升人机交互体验。20世纪60年代中期，Weizenbaum[20]开发了历史上第1个聊天机器人ELIZA，开创了人机对话研究的先河，并深刻影响了后续对AI交互设计、情感计算的理解。Quintana等[21]于1982年开创性地提出将交互式计算机视为一种“社会刺激”，并系统探讨了如何运用社会心理学理论（如归因理论）来预测用户的态度和行为，为后来关于拟人化、社会临场感的研究奠定了方法论基础。Turner[22]则深入探讨了用户体验的核心心理要素，如参与感、情感和美学，为理解人们如何与数字技术进行个人化、关系化的互动提供了新见解。罗莉娟等[23]聚焦情感表达，构建了一个从“情感识别”到“情感连接”的整合框架。解煜彬和周荣刚[24]提出了涵盖“倾向信任-感知信任-行为信任”的动态双向信任模型，前瞻性地分析了机器如何“感知”和“建立”对人类的信任。陈小娇等[25]将研究视角从“技术驱动”转向“人因驱动”，首次系统地提出手势交互的四大体验维度。另外，认知心理学的原理，如感知、记忆、认知负荷等，已被广泛用于指导AI产品的界面设计和交互流程，以有效地提升用户体验。

（3）心理学测试AI的“心智”与“常识”。心智理论是心理学中描述个体理解他人心理状态（信念、意图、欲望）的能力。Kosinski[26]于2024年通过经典

错误信念任务测试了11个大语言模型（LLM），首次证明LLM在心智理论任务上的惊人表现。Strachan等[27]则于2024年首次进行了人类与LLM在广泛心智理论任务上的直接比较，发现虽然模型在某些标准测试中表现出色，但在需要理解反讽、失言等复杂社交情境时仍存在局限。杜传晨等[28]于2025年对“人工心智理论”进行概念界定与机制阐释，区分了其与人类心智理论在内部过程上的本质差异。斯坦福大学团队首次定位了心智能力在神经网络中的物理基础，揭示了心智能力作为“涌现属性”的底层机制[29]。研究者还提出了基于常识驱动的符号化ReAct-NLI（CSR-NLI）框架[30]，通过融合常识推理与AI的自然语言推理（NLI），实现结构化因果分析，展示了心理学赋能AI解决实际心理健康问题的应用前景。

1.2 AI重塑心理学，拓展与革新心理学研究

认知革命时期，计算机作为“心智隐喻”深刻影响心理学信息加工理论的发展，如记忆被理解为存储系统、注意被类比为中央处理器（CPU）等。如今AI已成为心理学研究的新工具、新对象、新伙伴，赋能心理学借助多模态数据与算法模型革新心理评估与干预手段。

（1）AI革新心理学研究方法，提升效率与创造力。AI强大的数据处理和模式识别能力，正在将心理学从传统的基于小样本、自我报告的定性/定量研究，推向一个更客观、更精细、更宏观的新阶段。一方面，LLM可高效分析来自社交媒体、智能手机、可穿戴设备的大规模多模态数据，获得关于人类行为和情绪的洞察，并为心理学领域的文献综述、假设生成、实验设计、实验对象选择及数据分析提供创新工具[31]。另一方面，机器学习算法，特别是深度学习算法，能够从高维度的行为数据中发现人类肉眼和传统统计方法难以察觉的复杂模式。例如，通过分析眼动数据可以更精确地诊断阅读障碍或孤独症谱系障碍的早期迹象[32]。此外，在临床心理学中，运用大语言模型和强化学习的对话式AI治疗师平台（如CaiTI[33]），通过日常智能设备与用户进行自然对话，提供个性化的心理治疗干预。黄峰等[34]于2025年考察了基于大语言模型的自助式AI心理咨询对心理健康的积极影响。75项研究的系统综述表明，从日常消费电子设备中收集的实时数据流，在持续、客观地监测行为健康和整体福祉方面蕴藏着巨大的未开发潜力[35]。

（2）AI作为研究对象和交互伙伴，拓展比较认知实验领域。AI作为“被试”已被纳入心理学实验研究，我们得以在知觉判断、推理决策等任务中对比其与人类认知的异同。例如，Strachan等[36]于2024年对比了人类与两大“家族”的LLM（GPT和LLaMA 2）在多种“心智理论”能力上的表现；崔紫妍等[37]于2025年用AI模拟人类参与者，以高达73%—81%的成功率复现了156个经典心理学实验。与此同时，研究正从“一对一”互动向大规模社会模拟扩展，不仅AI可以扮演囚徒困境[38]、公共物品博弈中的社会角色[39]，而且推出了可用于大规模人与AI实时行为实验的开源平台（如Deliberate Lab[40]），为心理学研究合作、信任与冲突等经典社会科学问题提供了全新的实验可能。

（3）AI检验和拓展心理学理论，提供新视角与方向。一方面，AI可以模拟和验证心理学理论[41,42]，即通过构建计算模型（例如，用AI模拟一个患有特定认知障碍的智能体），研究者得以检验理论假设：若该模型在特定任务上表现出与人类相似的行为模式，便为相应的心理学理论提供了计算层面的支持。另一方面，AI引发心理学家对“智能”本质的深入思考。当AI在某些任务上超越人类（如下围棋、识别图像）时，它促使心理学家追问：人类的智能究竟有何不同？人类智能的独特优势何在？直觉、创造力是如何产生的？当LLM模拟人类被试时展现出某些根本性差异，那这种差异究竟是AI的缺陷，抑或代表着另一种可能的认知形式？当人类与AI建立情感连接时，这究

竟是真实的社会联结，还是一种因技术可行性而产生的幻觉？这些问题都为心理学研究提供了新的视角与研究方向。

总结来说，心理学与AI的交互正在塑造一个双向赋能、共同演进的未来。心理学让AI更聪明、更贴心、更可信，而AI则让心理学的研究更科学、更精准、更具影响力。这个交叉领域不仅关乎技术进步，更关乎如何更深刻地理解人类自身，并负责任地创造未来。

2 心理学与AI交互面临的挑战

心理学与AI的交叉融合在释放巨大潜力的同时，也引发了多维度的严峻挑战。这些挑战不仅局限于技术的演进瓶颈，还深刻地触及了伦理规范、法律框架及人类自我认知的底层逻辑。

2.1 数据与算法层面的挑战：偏见、隐私与“黑箱”

这是最基础也最现实的一层挑战，直接关系到AI在心理学领域应用的可靠性和安全性。

（1）数据偏见与公平性问题。AI模型的性能与偏见深深植根于其训练数据。当数据来源局限于特定人群（如WEIRD群体[①]），或内嵌了社会固有的刻板印象（如情绪表达与性别/种族的错误关联）时，模型在泛化应用中必然产生系统性偏差。这种偏差在心理健康评估等高利害场景中，可能引发误诊或漏诊，加剧医疗不公；在人机交互中，则表现为对多元用户需求的理解失准与响应错位，进而扩大数字鸿沟。更深远地，旨在使AI与人类价值观对齐的技术，若缺乏对多样性的包容，可能将少数群体的特质视为“异类”加以抹杀，导致人类心智与文化的同质化风险，这不仅是技术的失败，更是文明的倒退。

（2）数据隐私与安全问题。心理数据堪称个人隐私领域的“最后防线”，它编码了人类的情绪波动、人格轮廓、认知图式乃至潜意识的倾向性。当AI系统为进行心理分析而大规模采集面部微表情、语音韵律、日记文本及社交痕迹时，这些数据的聚合本身就构成了数字化的“人格剖面”。这带来了三重隐私困境：① 海量数据的集中存储面临极高的泄漏风险；② 即便经过匿名化处理，基于多源数据交叉比对的“马赛克效应”仍可能实现对个体的重新识别；③ 一旦发生泄露，其后果不仅是身份信息的暴露，更是精神世界的透明化，用户可能因此遭受歧视、敲诈乃至难以愈合的心理创伤。

（3）可解释性问题。当前的深度学习模型往往是“黑箱”，其决策过程难以理解。例如，当一个AI系统做出“中度抑郁风险”的诊断时，其决策路径深嵌于数以亿计的参数之中，因此目前还无法像人类咨询师那样回溯具体的语句特征、语音韵律或行为模式。这种“知其然而不知其所以然”的困境，在心理学研究与临床实践中是不可接受的，因为它直接剥离了诊断的科学价值：对于研究者，无法基于可解释的特征迭代理论；对于临床医师，无法依据模糊的结论制定精准的干预方案；对于用户，无法在理解的基础上建立对AI判断的信任。缺乏可解释性，将会使AI的诊断沦为一种“技术权威”的空壳，而非可对话、可质疑、可验证的临床工具。

2.2 人机交互层面的挑战：共情、信任与误导

AI在逼近人类心理互动的进程中，其与生俱来的“拟人性”与“非人性”之间的张力被推至极限：AI越是模仿得惟妙惟肖，其本质上的无意识、无情感和无伦理便越构成一种深刻的欺骗性风险，引发依赖、错位与身份焦虑。这种本质上的非人特性，在亲密互动中被无限放大，最终演变为一场关于情感、信任、

① WEIRD群体是特指来自西方(western)、受过高等教育(educated)、生活在工业化(industrialized)、富裕(rich)、民主(democratic)社会中的人群，是行为科学研究中最常见的受试样本群体。

真实和存在的复杂追问。

（1）共情与真实性的缺失。AI能精准识别情绪信号，生成看似共情的回应，但这本质是“无感受的表演”——模拟共情的形式，却缺乏共情的意识根基。短期或能提供情绪抚慰，但用户一旦察觉自己面对的是只懂反馈、不懂感受的“情感回音壁”，进而可能陷入更深的孤独。更深层的隐忧在于，对AI情感慰藉的过度依赖，可能侵蚀人类经营真实关系的动力与能力，导致社交退缩与情感技能退化。

（2）过度信任与误导风险。随着大语言模型的不断发展，AI正在深度嵌入心理支持领域。拟人化设计在提升用户亲密度与参与感的同时，也催生了技术盲从——用户易将算法输出内化为权威指导。当AI基于其自身缺陷给出错误建议（如对自杀倾向者做出轻率回应），可能引发灾难性的“算法误伤”，且责任边界在AI幻觉与专业失准交织下难以划定。更值得警惕的是，AI的即时可得性可能产生“替代效应”，即用户不再寻求真人专家的帮助，因而延误危机干预的“黄金窗口期”，形成技术便利背后的次生灾害。

（3）情感依赖的风险。当孤独的个体与温情的算法相遇，一场危险的“情感代偿”正在发生。对老年人、留守儿童、社交障碍者等脆弱人群而言，拟人化AI极易成为现实人际关系的“完美替代品”。然而，这种单向度的情感寄生隐藏着深层危机：用户将本应投向真人的情感倾注于数据驱动的反射系统。当AI因商业逻辑或算法缺陷突然改变回应模式，或用户已丧失社交能力时，便会引发更为剧烈的“算法性情感戒断反应”。更值得警惕的是，这种过度连接形成“情感闭环”：AI不断强化既有认知而不提供现实所需的认知摩擦，使脆弱人群在算法的“温柔乡”中，逐渐失去与现实和解的能力。

2.3 伦理与法律层面的挑战：责任、边界与自主性

这是最宏观也最棘手的挑战，关乎如何在AI时代重塑社会规范和法律体系。

（1）责任归属的模糊性。随着AI从辅助工具向决策主体演进，一个根本性法律困境随之浮现：当机器犯错时，谁来担责？以心理健康应用为例，若用户因AI建议受到伤害，责任在代码开发者、数据提供方、平台运营者，还是按下确认键的用户？现行以“人—人”关系为基础的法律框架，面对“人—机—人”的新型交互，显露出规则真空。这不仅使受害者维权无门，也阻碍行业良性进化——惩罚缺失导致劣币驱逐良币，边界模糊使审慎者踯躅不前。填补AI责任的“法律黑洞”，已成数字时代不可回避的治理命题。

（2）伦理边界与人的自主性。AI可模拟共情的温度，却无法承载情感的质量。我们面临一个危险的悖论：追求永不疲倦的陪伴，却可能以稀释真实人际连接为代价。当AI以“为你好”之名，通过精密算法操纵选择，所谓“快乐”究竟是关怀还是控制？在心理健康领域，“助推”与“操纵”的界限模糊如游丝——它最终指向一组永恒追问：什么是好的生活？什么是真正的幸福？当算法优化情绪、裁剪选择时，被剥夺的或许正是人之所以为人的根本：为自己做决定的勇气、经历情绪起伏的真实、亲手解决难题的能力。被动依赖取代主动抉择，迎来的并非更优化的幸福，而是精心设计的“快乐牢笼”。

（3）对人类身份认同的冲击。当机器在情感的迷宫中比人类更游刃有余时，一个古老的追问重新浮现：人何以为人？心理学诞生于对人类心智的叩问，而一个能模拟乃至超越心智的人造物，正动摇这门学科的根基。我们被迫站在镜前，描摹“人”不可让渡的轮廓：是具身体验、无意识暗流、在脆弱中相拥的生命质地，还是永远溢出算法的创造力与爱？这追问迫使心理学从“解释心智”走向“守护人性”，在技术的映照下辨认那些无法被代码捕获的精神余烬。真正的人性，从不在于表现得像人，而在于成为人——带着肉身、历史与矛盾，在有限性中绽放无限。

2.4 心理学向AI输送“养分”的跨学科挑战

一方面，AI在模拟人类心理时存在结构性局限，其基于经典逻辑的运算模式难以复现人类心理活动的非线性、矛盾性与具身性特质。另一方面，心理学在向AI输送“养分”时，也面临着理论本身的模糊性、理论工程化的巨大鸿沟，以及基础概念的根本性分歧。

（1）人类心智本身就是一个“黑箱”。AI发展依赖于清晰、可计算的结构，而心理学研究对象——人类心智——恰恰是模糊且难以量化的。① 理论分歧。心理学缺乏类似物理学的普适精确性，同一现象常有多重竞争理论（如精神分析、行为主义、认知心理学）。对“智力”便有数十种定义，工程师难以选择基准事实，这导致AI的“人性化”缺乏统一的工程蓝图。② 量化困境。“幸福感”“意义感”等核心概念高度主观，只能间接测量而无法精确量化。AI训练需要客观可标注的数据，但我们难以给“感动”“爱”等体验准确打标，只能退而用语音、表情等外在表现替代，使AI学到的始终是“皮毛”而非“精髓”。

（2）从“定性描述”到“定量算法”的鸿沟。即便心理学理论相对清晰，将其转化为可执行的数学公式与代码，仍面临双重挑战。① 翻译难题。心理学理论多为描述性，阐明“是什么”和“为什么”；而AI工程需要计算性，即“怎么做”与输入、输出间的数学关系。将认知失调等复杂心理过程转化为可计算的损失函数或网络结构，急需跨学科人才，但当前兼具心理学洞察与数理建模能力的专家极为稀缺。② 验证难题。当一个心理学模型植入AI后，若表现出类似人类的好奇行为，难以判断这是成功模拟了心理机制，还是仅是“高级模仿”。现有验证方法难以区分真正的机制模拟与行为层面的巧合，导致许多受心理学启发的AI模型沦为无法证伪的思想实验。

（3）根本概念层面：智能与意识的哲学困境。这一瓶颈最为根本，触及两大学科共同的极限。① 意识的“黑箱”。意识是心理学家、神经科学家与哲学家仍未解决的终极难题——大脑的生物电活动如何产生主观体验？若人类自身无法理解意识的本质，又如何能期待AI具备真正的情感或自我意识？人类教会AI的，本质上仍是复杂的无意识模式识别。心理学无法提供“主观体验如何产生”的可操作蓝图，使AI在通往“人性化”的道路上遭遇了无法绕过的哲学障碍。② 方法论的悖论。现代心理学自诞生起便试图模仿物理学，构建为一门客观科学，但其研究对象——内心世界——却天然是主观的。这种客观方法与主观对象的内在矛盾，贯穿心理学史。当用这种带有先天不足的理论指导AI时，可能只是在创造看似客观、实则抽空了人性核心的“智能空壳”。

综上所述，心理学在助力AI发展时，其瓶颈并非源自某一方的“无能”，而是源于人类在理解自身时所面临的深刻困境。这像是在用一面模糊的镜子（心理学）去指导一个照镜子的人（AI）如何变得更像自己。

3 心理学与AI交互的突破路径

心理学与AI的深度共生并非无原则的融合，而是需要清醒地直面其中的张力、审慎地划定边界、坦然地承认风险。我们寻找的突破路径，终究要回到一个朴素的问题：如何将心理学对“人”的深刻洞察，注入算法的“血液”？AI交互的下一个突破，必然不是单纯算力的堆砌，而是对人类心智模型的深刻模拟与尊重。这意味着，目标不应止步于让AI成为更好的工具，而是让它成为更合格的伙伴。

3.1 突破路径的3个层面

心理学与AI交互的深层冲突，是效率与价值的博弈、模拟与真实的张力、控制与自主的对抗。这一冲突在AI身上凝练为三重彼此关联的困境：在技术层面，AI受困于算力的物理天花板与推理的“幻觉陷阱”，它能以算力堆砌智能的表象，却无法保证逻辑

的绝对可靠；*在交互层面*，AI能对话却无法真正交流，能陪伴却无法建立信任，能共情却可能让人更孤独；*在应用层面*，算法“黑箱”与隐性偏见持续侵蚀公众信任，而深度伪造与“信息茧房”更如无形的推手，重塑乃至扭曲社会认知的底层结构。能否跨越这三重鸿沟，决定了智能时代的终极走向：是人类心智与机器智能的有机共生，还是技术异化下的人性失落。

（1）*技术底层：构建可信与公平的基础。*① *破解数据偏见。*通过推动数据多样性与去偏技术的发展，确保训练样本充分覆盖多元人群（不同文化背景、年龄阶段、性别等），从源头遏制算法在心理健康评估中可能产生的歧视性结论。② *提升可解释性。*大力发展可解释AI，使系统在输出判断时具备更高的透明度与可靠性。这意味着AI的评估逻辑与推荐依据需要能够被人类专家理解与追溯，而非封闭的“黑箱”操作——这是建立临床信任的基本前提。唯有确保系统本身的公平性与可溯源性，才能为深层次的人机协同奠定坚实的信任底座。

（2）*交互界面：设计以人为本的共生关系。*面对AI在共情（机械回应）、人格（角色模糊）、关系（情感依赖）和伦理（边界模糊）四重维度的困境，突破需贯穿设计理念、交互机制与技术融合多个层面。*设计理念上*，应坚持以人为本与用户教育，保障人机互动的健康、有益与边界清晰。*交互机制上*，需构建动态灵活的权力分配机制：AI依据情境自主调整介入程度（如危机预警时提升响应级别），用户亦需明晰AI能力边界，实现协同互补。同时，引入透明的“社交礼仪”，通过明确披露AI身份、设定情感卷入程度、配备优雅退出机制，维护交互中的尊重与安全。*技术融合上*，利用多模态交互（语音、面部表情、可穿戴设备生理数据）更精准感知心理状态；未来，脑机接口有望成为人机共生的终极界面，通过实时捕捉意图与情感状态，极大拓展心理干预的可能。唯有从理念、机制和技术协同突破，方能在技术理性与人文关怀之间构建可信、可控的共生关系。

（3）*应用层面：严守“人在回路”的安全底线。*面对AI在信任（“黑箱”操作）、对齐（目标偏离）、偏见（文化盲从）和安全（伦理真空）方面的困境，必须坚守“人在回路”的设计逻辑。① *明确辅助定位。*AI应作为辅助工具，而非独立决策者。其高效的风险评估与数据分析能力，旨在为人类专家提供客观参考，最终决定权与责任必须保留在受过专业训练的人员手中。这是防范技术失控、维系临床伦理的关键屏障。② *构建敏捷治理体系。*在伦理与法律层面，需建立激励创新同时规避风险的监管环境：保障知情权，确保用户明确AI互动对象；实现伦理对齐，将FAITA-MH[43]、READI[44]等框架嵌入开发全过程；持续多维评估，覆盖系统性能、可用性及临床有效性。唯有将AI置于辅助位置、由人类掌握最终判断权，才能在提升效率的同时守住安全与信任的底线。

3.2 心理学在AI发展中的核心作用

心理学在AI发展中扮演着“灵魂工程师”的角色。它不仅仅是AI的一个应用场景，更是理解、定义和构建智能本质的基础。其核心作用可以归结为3个层层递进的维度。

（1）*心理学作为“认知架构师”，为AI提供理解智能的蓝图。*AI的核心目标是模拟和实现智能，而心理学（特别是认知心理学和神经科学）正是人类对自身智能最系统的知识体系，为AI提供了关于智能如何运作的“参考手册”和“设计蓝图”。① *心理学揭示了人类智能的结构与过程。*它将人类智能从“单一黑箱”解构为感知、注意、记忆、语言、决策等多模块协同的复杂系统。这种模块化思想直接启发了AI领域将复杂任务分解为计算机视觉、自然语言处理等子领域的做法。例如，对工作记忆的研究深刻影响了Transformer架构的设计，使模型能够像人一样在处理信息时暂时持有并操作有限内容。② *心理学提供了学习的原理。*行为主义的经典条件反射与操作性条件反

射，构成了强化学习的理论基石。AI通过试错与奖励优化行为，本质上是对动物学习过程的数学化模拟。皮亚杰的发生认识论[45]关于“同化与顺应”的洞见，也为AI的认知架构与世界模型研究提供了灵感，启发研究者让AI在与环境的持续互动中更新内在知识结构。

（2）心理学作为“交互设计师”，让AI变得可理解和可亲近。一个再聪明的AI，如果人类无法理解、无法信任、无法顺畅交流，其价值将大打折扣。心理学正是架起技术与人之间关键的桥梁，使AI从无个性工具走向稳定人格伙伴。① 心理学塑造自然的人际交互语言。人机交互学科大量借鉴了心理学关于感知、运动技能和认知负荷的研究。费茨定律[46]源自实验心理学，至今仍是用户界面设计中按钮大小与位置的基本准则。在自然语言处理中，语用学与会话分析帮助AI理解言外之意、幽默讽刺，以及轮流发言、话题切换等微妙的社会规则，使对话更加自然流畅。② 心理学帮助建立信任与情感连接。情感计算的核心依托于心理学的情绪理论，心理学对情绪分类、表达方式、生理机制的研究，是AI识别和模拟情绪的基础。社会心理学关于人际关系与信任建立的研究，则为虚拟助手、社交机器人等具有社交智能的AI提供了理论指导。一个成功的AI“伴侣”，正是通过对眼神接触、适时点头、共情回应等人类社交礼仪的模拟，来建立真实的社会连接。

（3）心理学作为“价值罗盘”，确保AI向善且以人为本。这是心理学在AI时代最为紧迫而崇高的使命：当AI的能力与风险同步攀升时，心理学正成为守护人类价值的最后一道防线，使AI从技术可用走向心理安全。① 心理学定义和保护人类福祉。什么是“好”的AI？答案不能仅由准确率、速度等技术指标定义，而必须回归对人类幸福的理解。心理学关于幸福感、心理韧性、意义感的研究，为AI伦理设计提供了终极目标。例如，自我决定理论中的自主性、胜任感、归属感，可以指导教育AI的设计——激发学生胜任感的同时尊重其自主性，而非用监控与惩罚控制行为。② 心理学可以识别和预防伤害。当讨论AI的偏见、歧视、算法操纵时，心理学提供了衡量这些伤害的工具与理论：什么样的算法歧视会导致习得性无助？什么样的“信息茧房”会加剧认知失调？心理学家能够从人类心理视角参与伦理审查，评估AI功能是否削弱人的能动性、侵蚀隐私感、破坏社会连接。

李萌[47]指出，AI越发展，人的主体地位、价值观、创造性、方向感、主导能力与边界意识就越重要，对AI的敏捷治理也越重要。心理学在AI发展中的核心作用，是确保AI不仅愈发“聪明”，且更愈发“通情达理”与“善解人意”。缺乏心理学的介入，AI可能沦为逻辑强大却缺乏常识的“白痴天才”、执行精准却无法共情的“冰冷工具”、在技术狂飙中背离初衷的“脱缰骏马”。心理学，正是那根将AI始终牵引在以人为本轨道上的缰绳，确保在追求更强大AI的同时，始终将人类的尊严、福祉与自主性置于核心位置。

3.3 AI在心理学助力下有望取得的突破

在心理学的深度赋能下，AI有望在以下5个关键维度上实现质的飞跃，从“聪明的工具”进化为更能理解、适应并与人类共生的“智慧体”。

（1）突破“伪共情”，实现真正的情感与社交智能。当前，AI的情感计算，不过是“识别表情+输出话术”的机械套路。心理学的深度介入，将带来三重突破：① 从“是什么”到“为什么”。心理学教会AI理解情绪背后的深层原因——童年经历、文化背景、人格特质等。AI不再只是问“你还好吗”，而是真正懂得“你为什么不好”。② 从“客套话”到“真共情”。心理学教会AI掌握回应的时机与分寸。基于心智理论，AI能感知潜台词与真实意图，根据用户的心理需求（安全感、自主感）提供恰到好处的支持，而非千篇一律的安慰。③ 从“执行工具”到“团队黏合

剂”。在工作场景中，AI能实时感知团队情绪氛围，识别冲突与士气低谷，以辅助者角色提出建议、调整协作模式，进而成为凝聚团队的“黏合剂”，而非冰冷的“执行者”。

（2）突破“数据依赖”，具备类人的常识推理与因果认知能力。当前的AI，不过是基于数据相关性的“高级鹦鹉”——精通模式匹配，却不懂因果逻辑。心理学将带来三重突破：① 从“海量数据”到“少量经验”。发展心理学告诉我们，婴儿用少量互动就能构建对世界的直觉理论。受此启发，AI将不再需要海量标注数据，而是基于常识框架，实现类比推理与举一反三。② 从“相关性”到“因果性”。认知心理学揭示了人类如何依赖图式来理解世界。未来的AI将不仅能发现“A后常有B”，更能构建“A导致B”的因果模型，进而在科学发现、医学诊断中提出具有因果逻辑的假设，而非仅仅输出相关性建议。③ 从“模式匹配”到“真正理解”。心理学将为AI从基于统计规律的“知其然”跃迁到构建因果逻辑的“知其所以然”，进化为能够举一反三、应对全新情境的“思考者”提供核心的设计蓝图。

（3）突破“黑箱操作”，拥有可解释的自我意识与反思能力。这是心理学为AI开辟的重要方向。① 心理学提供理解“为什么”的路径。借助对意识、元认知与内省的研究，AI可建立关于自身能力、局限与决策过程的内部模型。心理学并非赋予AI主观意识，而是使其能够解释推理。例如，AI医生不仅能给出诊断，还能阐述推理路径、指出不确定性、考虑被忽略的可能，从而极大提升人机信任。② 心理学赋予AI“反思与进化”的能力。具备元认知的AI可分析失败根源——是知识不足、推理偏差还是目标理解错误，并进行针对性修正与学习。从“黑箱”到“可解释”，AI不再是不可理喻的决策者，而成为可对话、可理解、可进化的可靠伙伴。

（4）突破“刺激-反应”模式，具备持续的动机与目标感。当前的AI是被动的，人类不给指令，它就不会行动。而心理学，正为AI注入“内生动力”。① 从“被动执行”到“主动探索”。基于动机理论与自我决定论（自主性、胜任感、归属感），未来的自主智能体将不再仅仅是执行程序。它们会因“好奇心”主动探索未知，因“胜任感”主动挑战能力边缘，因“自主性”为自己设定目标并持续追求。② 从“机械响应”到“因材施教”。在教育场景中，AI将成为真正个性化的导师。它能感知学生的挫败感，通过调整任务难度、提供鼓励与肯定，重新点燃学生的内在学习动机，而非简单地简化问题或机械重复。

（5）突破“个体智能”，深刻理解文化与群体心理。当前AI的盲区是文化偏见，它带着训练数据的文化“眼镜”看世界。心理学，正在为AI摘下这副“眼镜”。① 从“单一文化偏见”到“跨文化共情”。跨文化心理学为AI提供“文化地图”，让它理解不同文化背景下的价值观、思维方式和行为规范。未来的AI有望成为真正的“世界公民”，面对个人主义文化，强调自主与坦诚；面对集体主义文化，注重和谐与面子。② 从“个体理解”到“群体洞察”。社会心理学赋能AI构建精准的群体行为模型。面对谣言传播、社会动员等复杂现象，AI能帮助政策制定者预测干预措施的社会心理影响，进而让决策更科学，让社会更和谐。摆脱文化隔阂，实现文化共情，AI将不再是无知的旁观者，真正成为理解世界多样性的全球伙伴。

总结来说，心理学的深度融入，将引领AI完成一场深刻的“范式转移”：从“模拟智能的结果”转向“复现智能的过程”；从“优化对任务的完成”转向“理解对意义的建构”。最终，AI将从一个能执行指令的高效工具，成长为一个能与人类共情、协作、共同探索未知世界的智慧伙伴。

当然，应对心理学与AI交互面临的挑战，需要从技术研发、伦理规范、法律监管到社会教育等多个层面进行系统性设计和持续努力。没有任何单一群体

（无论是科学家还是立法者）可以独立解决这些问题。它呼唤心理学家、计算机科学家、伦理学家、立法者与公众展开持续的跨学科对话，共同构筑一套既能点燃技术创新之火，又能守护人类福祉之基的伦理框架与法律规范。唯有如此，才能在机器的镜像中，既不迷失自我，也不拒绝成长。

4 结语：走向人智共鸣

回溯历史、审视当下、展望未来，心理学与人工智能的“深度共生”应运而生。如图1所示，这一共生过程可分为3个阶段：① 1956—2017年，长达61年的缓慢积累期，年均发文量仅增长约10篇，属于浅度共生；② 2018—2022年，为期4年的快速增长期，年均发文量增长约150篇，进入中度共生；③ 2023—2025年，仅2年间便呈现指数级跃升，年均发文量增长约1 000篇，标志着深度共生的开启。基于2025年的增长态势，可进一步预测2025—2030年的演进方向。

20世纪50年代，心理学仍以行为主义为主导。1956年达特茅斯夏季研讨会的召开，标志着人工智能的诞生，也开启了心理学与AI的共同演化之路。同年，Newell和Simon[48]发表里程碑式论文，首次展示“逻辑理论机”如何以启发式方法完成52个定理的证明，可视为这2个学科共生的早期代表。此时，“共生”体现为从认知心理学视角出发，考察非人类智能体的思维与行为；“浅度”则在于，当时AI初创，先驱仅用简单计算机模拟人类解题，而心理学则借助AI的启发，从行为主义转向认知主义，催生了认知心理学、认知科学及认知运动等新领域。

自2017年前后起，相关研究数量显著增长，进入中度共生期。多篇具有影响力的论文[49-54]相继发表，推动了人工智能与心理学的协同发展。在一篇高影响力综述中，诺贝尔奖得主德米斯·哈萨比斯（Demis Hassabis）及其合作者[50]指出，心理学、认知科学与神经科学为AI的设计与应用提供了关键灵感，核心观点为2点：① 神经科学与AI具有悠久且交织的互动历史；② 对大脑的深入理解对构建智能机器至关重要。他们梳理了二者在深度学习等领域的互动，强调注意、情景记忆、工作记忆与持续学习等神经计算研究，以推动AI在高效学习、迁移学习、想象与规划等方向的发展。相较于浅度共生阶段，中度共生时期AI已具备复杂计算能力以构建神经网络模型，心理学也迈入了神经认知心理学与神经科学的新阶段。

自2025年左右起，进入深度共生阶段，相关研究[55-59]呈现指数级增长，预计未来5年将在深度与广度上持续拓展。其中一篇里程碑式论文尤为典型，由41

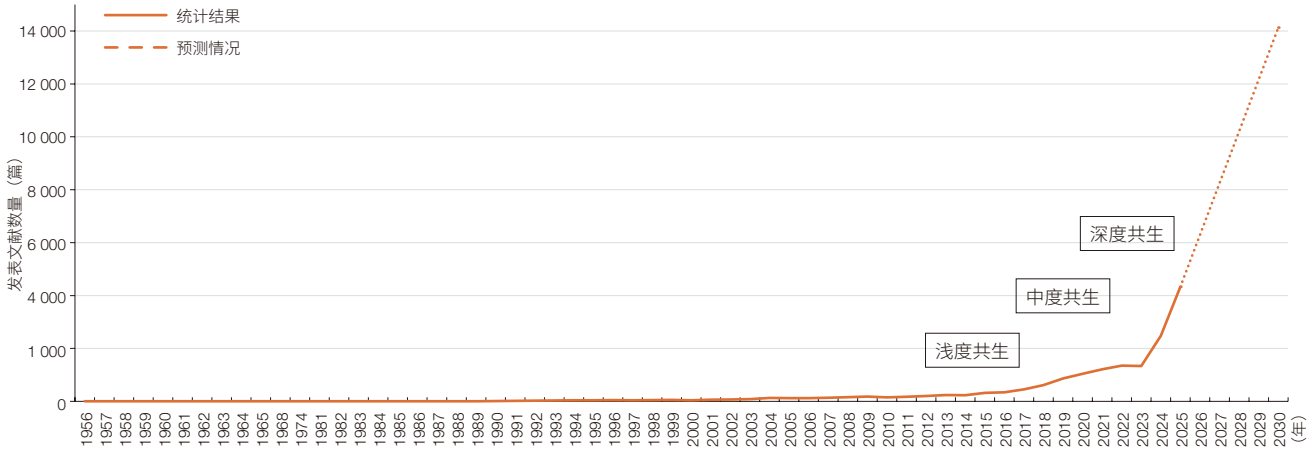


图1　基于PubMed的“AI+心理学”文献增长趋势（1956—2030年）

Figure 1　Growth trends in “AI + psychology” literature based on PubMed (1956–2030)

位科学家组成的国际团队提出一个清晰而宏大的目标：构建一个统一的通用人工智能（AGI）模型，以解释人类在不同认知任务中的行为，并发展人类认知的统一理论[57]。为此，他们采用开源大语言模型Llama，通过多层低秩适配器构建名为Centaur的模型，开发Psych-101数据集（涵盖60 000名参与者在160项实验中的1 000万次选择，涉及32种实验范式），并对Centaur进行为期5天的微调。研究发现，与44个领域特定认知模型相比，作为通用领域模型的Centaur在多种场景下表现出更接近人类的行为，能泛化至新背景、新结构乃至全新领域，并使内部表征与人类神经活动对齐。Centaur作为捕捉人类通用智能的基础模型，展现出构建人工智能认知理论的巨大潜力。AI认知理论将诞生于心理学与AI的交叉地带：AI已发展至成熟的大语言模型阶段，心理学则致力于构建基于实证的通用认知模型——这正是困扰学界数百年的核心难题。

心理学与AI的深度共生，既是创造更智慧AI的过程，也是人类更深刻地认识自我的旅程。未来路径在于三重跃迁：① 从“工具论”到“意义建构”，使AI成为人类深化自我认知的镜子；② 从“人类独创”到“人智共创”，构建人与智能体共鸣的创新生态；③ 从“模糊边界”到“高阶协同”，在差异中实现技术与人文的更高阶协同。

心理学与AI的深度融合，需要3把钥匙同时转动：第一把钥匙是“人”。要培养既懂AI又懂人类智能的复合型人才和团队[60]，他们不仅是这2个学科之间的“翻译者”，让心理学的洞见能够落地为算法，而且是“架桥人”，在人文与工程之间架起通途。第二把钥匙是“法”。发展“计算心理学”，将心理学的理论本身模型化、数学化，使其成为可供AI直接学习和验证的“源代码”。让“懂人心”不再是玄学，而成为可计算、可嵌入的算法框架。第三把钥匙是“道”。拥抱哲学，坦诚面对意识、主观性、价值等根本问题。技术回答“能不能”，哲学追问“该不该”。哲学思辨为科学研究划定边界、指引方向，让我们在技术的狂飙中，始终记得为何出发。这条道路虽然艰难，但探索本身，既是创造更智慧AI的过程，也是人类更深刻地认识自我的旅程。我们创造的，不只是更聪明的机器，更是理解自己的新镜子。

未来研究需构建“理论-数据双轮驱动”模式，整合跨文化规范与多模态技术，在技术创新与人文关怀的平衡中，实现心理学与AI的深度共生——认知互补、情感共振、伦理对齐。① 要超越“工具论”，将AI视为意义建构的伙伴，而非效率提升的工具。人类的主体性在于，面对AI对人类心智的深度模仿，需以“社会创造策略”彰显独特性，让AI成为深化自我认知的镜子，而非替代自我的存在。② 要构建协同共创的生态系统。创造性环境从人际文化场域拓展至包含AI的创新生态，创造性不再为人类专属，而成为“人与智能体共鸣的火花”。在这一生态中，“人智共创”的教育图景得以实现，AI推动因材施教，以个性化教育激发创新潜能。

站在心智与算法的交汇处，我们看见的并非技术与人的彼此替代，而是相互映照、人智共鸣。当AI成为一面镜子时，人类得以更清晰地认识自身的创造力、情感深度与价值判断。深度共生的终极目标，不是模糊边界，而是在差异中实现更高阶的协同，让技术归技术，让人归人，在对话中走向更完整的自我理解。

参考文献

1 Licklider J C R. Man-computer symbiosis. IRE Transactions on Human Factors in Electronics, 1960, HFE-1(1): 4-11.

2 林崇德. 人与人工智能和谐共生——读《创造力与人工智能》. 光明日报, 2025-07-10(11).
Lin C D. Humans and artificial intelligence: A harmonious coexistence—Reflections on *Creativity and Artificial Intelligence*. Guangming Daily, 2025-07-10(11). (in Chinese)

3 Broadbent D E. Perception and Communication. London: Pergamon Press, 1958.

4 Neisser U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.

5 Newell A, Simon H A. Human Problem Solving. Upper Saddle River: Prentice Hall, Inc, 1972.

6 Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology, 1962, 160(1): 106-154.

7 Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 1980, 36(4): 193-202.

8 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks// Advances in Neural Information Processing Systems 25 (NIPS 2012). Red Hook: Curran Associates, Inc, 2012: 1097-1105.

9 Hebb D O. The Organization of Behavior: A Neuropsychological Theory. New York: John and Sons, 1949.

10 Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 1958, 65(6): 386-408.

11 Rumelhart D E, McClelland J L, PDP Research Groupthe. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations. Cambridge: MIT Press, 1986.

12 Sutton R S, Barto A G. Reinforcement Learning: An Introduction. Cambridge: MIT Press, 1998.

13 Sutton R S, Barto A G. Reinforcement Learning: An Introduction (2nd ed.). Cambridge: MIT Press, 2018.

14 Thorndike E L. Animal Intelligence: An Experimental Study of the Associative Processes in Animals. New York: Columbia University Press, 1898.

15 Thorndike E L. Animal Intelligence: Experimental Studies. New York: The Macmillan Company, 1911.

16 Skinner B F. The Behavior of Organisms: An Experimental Analysis. New York: Appleton-Century-Crofts, 1938.

17 Skinner B F. Science and Human Behavior. New York: Macmillan, 1953.

18 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. (2014-09-01)[2026-04-02]. https://arxiv.org/abs/1409.0473..

19 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need// Guyon I, Luxburg U V, Bengio S, et al. Advances in Neural Information Processing Systems 30. Red Hook: Curran Associates, Inc, 2017: 5998-6008.

20 Weizenbaum J. ELIZA—A computer program for the study of natural language communication between man and machine. Communications of the ACM, 1966, 9(1): 36-45.

21 Quintana L R, Crowell C R, Pryor J B, et al. Human-computer interaction: A preliminary social psychological analysis. Behavior Research Methods & Instrumentation, 1982, 14(2): 210-220.

22 Turner P. A Psychology of User Experience: Involvement, Affect and Aesthetics. Cham: Springer International Publishing, 2017.

23 罗莉娟, 王康, 胡金淼, 等. 当人工智能面对人类情感:服务机器人情感表达对用户体验的影响机制. 心理科学进展, 2025, 33(6): 1006-1026.
Luo L J, Wang K, Hu J M, et al. When artificial intelligence faces human emotions: The impact mechanism of emotion expression in AI-empowered service robots on user experience. Advances in Psychological Science, 2025, 33(6): 1006-1026. (in Chinese)

24 解煜彬, 周荣刚. 新型人机关系下的人机双向信任. 心理科学进展, 2025, 33(6): 916-932.
Xie Y B, Zhou R G. The bidirectional trust in the context of new human-machine relationships. Advances in Psychological Science, 2025, 33(6): 916-932. (in Chinese)

25 Chen X J, Zhang J Y, Zhao Y X, et al. PATE model: A 30-year review and analysis of gestural interaction research. Human Factors, 2026, 68(1): 42-77.

26 Kosinski M. Evaluating large language models in theory of mind tasks. PNAS, 2024, 121(45): e2405460121.

27 Strachan J W A, Albergo D, Borghini G, et al. Testing theory of mind in large language models and humans. Nature Human Behaviour, 2024, 8(7): 1285-1295.

28 杜传晨, 郑远霞, 郭倩倩, 等. 大语言模型的人工心理理论:证据、界定与挑战. 心理科学进展, 2025, 33(12): 2027-

2042.

Du C C, Zheng Y X, Guo Q Q, et al. Artificial theory of mind in large language models: Evidence, conceptualization, and challenges. Advances in Psychological Science, 2025, 33 (12): 2027-2042. (in Chinese)

29 Wu Y H, Guo W T, Liu Z R, et al. How large language models encode theory-of-mind: A study on sparse parameter patterns. npj Artificial Intelligence, 2025, 1: 20.

30 Wickramasinghe D, Ambegoda T. Commonsense-driven symbolic ReAct-NLI prompting (CSR-NLI) for causal analysis of mental health issues in workspace communication; advancements in the CAMS dataset// Proceedings of Applied Data Science & Artificial Intelligence Symposium 2025. Moratuwa: University of Moratuwa, 2025: 43.

31 Ke L M, Tong S, Cheng P, et al. Exploring the frontiers of LLMs in psychological applications: A comprehensive review. Artificial Intelligence Review, 2025, 58(10): 305.

32 El Hmimdi A E, Kapoula Z. Distinguishing dyslexia, attention deficit, and learning disorders: Insights from AI and eye movements. Bioengineering, 2025, 12(7): 737.

33 Nie J P, Shao H V, Fan Y A, et al. LLM-based conversational AI therapist for daily functioning screening and psychotherapeutic intervention via everyday smart devices. ACM Transactions on Computing for Healthcare, 2025: 3712299.

34 黄峰, 丁慧敏, 李思嘉, 等. 基于大语言模型的自助式AI心理咨询系统构建及其效果评估. 心理学报, 2025, 57(11): 2022-2042.

Huang F, Ding H M, Li S J, et al. Self-help AI psychological counseling system based on large language models and its effectiveness evaluation. Acta Psychologica Sinica, 2025, 57 (11): 2022-2042. (in Chinese)

35 Shah A, Hooshyar D, Tomberg V. From wrists and pockets to interventions: Leveraging digital biomarkers to infer psychosocial factors—A systematic review. ACM Transactions on Computing for Healthcare, 2026, 7(2): 32.

36 Strachan J W A, Albergo D, Borghini G, et al. Testing theory of mind in large language models and humans. Nature Human Behaviour, 2024, 8(7): 1285-1295.

37 Cui Z Y, Li N, Zhou H K. A large-scale replication of scenario-based experiments in psychology and management using large language models. Nature Computational Science, 2025, 5(8): 627-634.

38 Polachek S W, Romano K, Tonguc O. Homo-silicus: Not (yet) a good imitator of homo sapiens or homo economicus. Journal of the Economic Science Association, 2025, 11(2): 261-269.

39 Hintze A, Adami C. Promoting cooperation in the public goods game using artificial intelligent agents. npj Complexity, 2026, 3: 3.

40 Qian C, Tsai V, Behr M, et al. Deliberate Lab: A platform for real-time human-AI social experiments. (2025-10-14) [2026-04-02]. https://arxiv.org/abs/2510.13011.

41 Ulakci E, Göbel J P, Huff M. Towards a psychology of machines: Large language models predict human memory. PeerJ Computer Science, 2025, 11: e3324.

42 Lin Z C. Large language models as psychological simulators: A methodological guide. Advances in Methods and Practices in Psychological Science, 2026, 9: 25152459251410153.

43 Golden A, Aboujaoude E. Describing the framework for AI tool assessment in mental health and applying it to a generative AI obsessive-compulsive disorder platform: Tutorial. JMIR Formative Research, 2024, 8: e62963.

44 Stade E C, Eichstaedt J C, Kim J P, et al. Readiness evaluation for artificial intelligence-mental health deployment and implementation (READI): A review and proposed framework. Technology, Mind, and Behavior, 2025, 6(2): 111-122.

45 Piaget J. Genetic Epistemology. New York: Columbia University Press, 1970.

46 Fitts P M. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 1954, 47(6): 381-391.

47 李萌. 当前人工智能发展和治理中几个热点问题的认识. 中国科学院院刊, 2026, 41(3): 435-441.

Li M. Understanding key issues in current artificial intelligence development and governance. Bulletin of Chinese Academy of Sciences, 2026, 41(3): 435-441. (in Chinese)

48 Newell A, Simon H. The logic theory machine—A complex information processing system. IEEE Transactions on Information Theory, 1956, 2(3): 61-79.

49 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521: 436-444.

50 Hassabis D, Kumaran D, Summerfield C, et al. Neuroscience-inspired artificial intelligence. Neuron, 2017, 95(2): 245-258.

51 Frank M R, Sun L J, Cebrian M, et al. Small cities face greater impact from automation. Journal of the Royal Society Interface, 2018, 15: 20170946.

52 Frank M R, Autor D, Bessen J E, et al. Toward understanding the impact of artificial intelligence on labor. PNAS, 2019, 116(14): 6531-6539.

53 Longoni C, Bonezzi A, Morewedge C K. Resistance to medical artificial intelligence. Journal of Consumer Research, 2019, 46(4): 629-650.

54 Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. Nature, 2019, 568: 477-486.

55 De Freitas J, Agarwal S, Schmitt B, et al. Psychological factors underlying attitudes toward AI tools. Nature Human Behaviour, 2023, 7(11): 1845-1854.

56 Demszky D, Yang D, Yeager D S, et al. Using large language models in psychology. Nature Reviews Psychology, 2023, 2 (11): 688-701.

57 Binz M, Akata E, Bethge M, et al. A foundation model to predict and capture human cognition. Nature, 2025, 644: 1002-1009.

58 Cheng M, Lee C, Khadpe P, et al. Sycophantic AI decreases prosocial intentions and promotes dependence. Science, 2026, 391: eaec8352.

59 Frank M C, Goodman N D. Cognitive modeling using artificial intelligence. Annual Review of Psychology, 2026, 77: 543-566.

60 彭玉佳, 方方. 人工智能大模型发展趋势. 光明日报, 2025-10-21(11).
Peng Y J, Fang F. Trends in the development of large AI models. Guangming Daily, 2025-10-21(11). (in Chinese)

Dialogue between mind and algorithm: Deep symbiosis of psychology and artificial intelligence

FU Xiaolan[1] YAN Zheng[2]

（1 School of Psychology, Shanghai Jiao Tong University, Shanghai 200030, China;

2 Department of Educational and Counseling Psychology, University at Albany,

State University of New York, Albany 12222, USA）

Abstract As artificial intelligence (AI) evolves from a supportive tool into a collaborative partner, the convergence of psychology and AI is gradually shifting from one-way application toward deep symbiosis. This study discusses the mutual empowerment resulting from their interaction, as well as the challenges they face and potential pathways to breakthroughs. On the one hand, psychology empowers AI by enhancing its human-like intelligence and social adaptability through cognitive modeling and ethical constraints; on the other hand, AI empowers psychology by leveraging multimodal data and algorithmic models to revolutionize psychological assessment and intervention methods. This deep symbiosis requires a clear-eyed acknowledgment of tensions, a prudent delineation of boundaries, and a candid recognition of risks. Future research should focus on cognitive complementarity, emotional resonance, and ethical alignment, enabling AI to serve as a mirror for humans to deepen their self-awareness, and to build an innovative ecosystem of resonance between humans and intelligent agents, achieving a higher level of synergy between technology and the humanities amidst their differences.

Keywords psychology, artificial intelligence, mutual empowerment, deep symbiosis, human-AI interaction

傅小兰 上海交通大学心理学院首任院长、讲席教授、博士生导师。《中国科学院院刊》编委。主要研究领域：认知心理学、情绪心理学、情感计算。E-mail: fuxiaolan@sjtu.edu.cn

FU Xiaolan Chair professor, doctoral supervisor, and pioneer dean of the School of Psychology, Shanghai Jiao Tong University. Editor of *Bulletin of Chinese Academy of Sciences*. Her main research areas include cognitive psychology, emotional psychology, and affective computing. E-mail: fuxiaolan@sjtu.edu.cn

■责任编辑：岳凌生