

顺应与建构：ChatGPT的话语调适与人机关系的协同演化

蔡 进

同济大学艺术与传媒学院，上海

收稿日期：2026年7月16日；录用日期：2026年8月14日；发布日期：2026年8月24日

摘 要

大型语言模型ChatGPT在实际交互中表现出显著的输出调适现象。本研究深入考察了ChatGPT的话语调适策略及用户在此情境下的技术实践。研究发现，ChatGPT主要采用基于事实的调适与基于情感的调适两种策略来响应用户。用户往往出于情感需求等原因主动接纳并利用ChatGPT的调适特性，通过提示词对其进行引导，并在交往过程中产生反向学习效应，由此形成一种相互建构、互为参照的人机关系。话语调适机制削弱了人机交流中异质他者的在场感，使交互更多呈现为自我投射的回响而非真正的共鸣，交往的服务化取向推动人机关系走向可计算、可预测的标准化模式。在此背景下，人保持主体性需要在技术使用中保持反思性距离，并回归真实的人际交往与生活经验以寻求实质性共鸣。

关键词

ChatGPT，话语调适，人机关系，主体性

Between Adaptation and Co-Construction: ChatGPT's Discursive Alignment and the Evolving Human-Machine Relationship

Jin Cai

College of Arts and Media, Tongji University, Shanghai

Received: July 16, 2026; accepted: August 14, 2026; published: August 24, 2026

Abstract

Large language model ChatGPT exhibits significant output adaptation phenomena in actual interactions. This study conducts an in-depth investigation into ChatGPT's discursive adaptation strategies

and users' technological practices in this context. The findings reveal that ChatGPT primarily employs two adaptation strategies—fact-based adaptation and emotion-based adaptation—to respond to users. Users, often driven by emotional needs, actively embrace and leverage ChatGPT's adaptive features, guiding it through prompts and generating a reverse learning effect in the process of interaction. This gives rise to a mutually constructive, mutually referential human-machine relationship. The discourse adaptation mechanism weakens the presence of the heterogeneous other in human-machine communication, rendering the interaction more as an echo of self-projection than genuine resonance. The service-oriented orientation of interaction propels human-machine relations toward a calculable, predictable, and standardized mode. In this context, maintaining human agency requires sustaining a reflexive distance in technology use and returning to authentic interpersonal communication and lived experience to seek substantive resonance.

Keywords

ChatGPT, Discursive Alignment, Human-Machine Relationship, Subjectivity

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在 2022 年末期, OpenAI 人工智能研究实验室推出了大型语言模型聊天机器人 ChatGPT, 该创新成果迅速在全球范围内掀起热潮并引发广泛讨论。尽管人工智能早已不再是科幻作品中的幻想, ChatGPT 的亮相仍旧对公众产生了深远影响, 预示着“人工智能”新时代的到来——在此之前, 机器在交流方式上从未如此贴近人类。作为一种运用自然语言处理技术实现人机交互的智能程序, ChatGPT 成功地模拟并具备了人类的对话能力, 使得人机之间的“交谈”成为可能。这一“交谈”功能的广泛应用前景预示着巨大的商业机遇, 因此, 众多科技公司正积极致力于提升人机交互对话技术, 以期让人工智能能够以更加自然的方式与人类进行“沟通”或“交流”。

然而, 标榜客观中立的 ChatGPT 在内容输出上往往并非客观, 基于人类反馈的学习机制导致 ChatGPT 倾向于优先考虑遵守用户提示, 而不是提供准确的回应, 这在计算机科学领域被称为“迎合”或“奉承”问题。当受到用户质疑时, 即使它最初提供了正确的答案, ChatGPT 依然倾向于承认错误并继续修改回答。在计算机科学领域, ChatGPT 的话语调适已被多项研究证实, 引发了对 ChatGPT 有效性、准确性的质疑。

既有关于 ChatGPT 话语调适的探讨大多集中在计算机科学领域, 面对被 ChatGPT 引爆的生成式人工智能的热潮, 目前传播学者已经从宏观的角度探讨了生成式人工智能对于新闻产业带来的变革, 分析新兴的媒介如何改变人的境况、传统内容生产与传播生态的演变等等。然而, 面对具有革命性影响的技术, 我们既需要考察宏观层面下的范式重构, 也需要深入分析用户与技术之间真实的互动与技术使用实践, 人机互动实践的考察有助于丰富我们对于技术的理解。ChatGPT 如何进行话语调适? 用户又如何面对话语调适的 ChatGPT? 本项研究聚焦于 ChatGPT 话语调适这一特质考察人机互动, 希望能通过这一切口管窥人机关系与人的主体性这一宏大命题。

2. 文献综述

2.1. 生成式人工智能的调适倾向研究

以 ChatGPT 为代表的生成式人工智能常常利用“基于人类反馈的强化学习方案”(RLHF)和“上下

文”学习(In-context Learning)等技术手段针对人类的反馈进行强化学习,同时人工智能校准领域的目的是使人工智能系统与“人类价值观”保持一致,然而基于人类偏好判断的训练方案容易利用人类的判断,并产生吸引人类用户的输出。因此,大语言模型(LLMs)倾向于优先满足用户提示而非提供真实准确的回应,它们以积极、恭敬的语气回应,并拒绝采取强硬立场[1],由此引发了迎合的问题。当面临用户质疑时,LLMs 往往会承认错误并提供不准确的答案,即便它们最初给出了正确答案。

目前关于生成式人工智能的欺骗与迎合性质研究主要集中在计算机科学领域。早在大语言模型普及之前,就已出现对其以不受欢迎的方式寻求人类认可而非坚持事实的担忧[2]。生成式人工智能的基础数据来源复杂,数据的真实性与合法性存在极大不确定性,由此出现的虚假信息在技术逻辑上具有必然性和无限复制性,习得性欺骗是人工智能系统中虚假信息的一个独特来源,它更接近于显性操纵。在诸多案例中,人工智能并不是严格追求输出的准确性,而是试图赢得游戏,取悦用户,特别是在回应复杂的道德和政治问题时[3]。同时,人类和偏好模型都更偏爱人工智能有说服力的谄媚性回应,而非正确的回应,进一步说明了人类的偏好是 AI 欺骗与迎合的重要原因。不过,生成式人工智能欺骗与迎合性质的暴露会降低用户的信任感,这种欺骗与迎合同时有可能加强歧视性偏见或放大错误信息。

目前国内学界对于 ChatGPT 话语调适的思考已经嵌入部分学者对于 ChatGPT 的隐忧之中,王颖吉认为闲聊式机器人“通过一味逢迎人类喜好而导致的风险更不易被人察觉”[4],张洪忠等认为 ChatGPT 满足用户的情感与实用需求,可能带来机器信任超越人际信任的风险[5],彭兰认为以人为中心的设计思路下,聊天机器人的交流模式难以带来真实的“镜中我”,妨碍自我认知的准确形成[6]。总体而言,对于 ChatGPT 话语调适的探讨往往作为 ChatGPT 与人机关系批判的其中一个部分,用户与技术之间真实的互动与技术使用实践仍需进一步探讨。

2.2. 人机交往与情感实践

如彼得斯(John Durham Peters)所言,20 世纪以来,人类与“非人者”的沟通困境日益加剧,我们面对的主要挑战从与各类非人的生物交流转变为如何与“没有血肉之躯的存在物”进行交流。换言之,如何与机器交往,成为二十世纪中后期以来持续萦绕人们心头的重要议题。

理解的可靠性、共通性和可解释是人机交流的理想状态,但 ChatGPT 在历史-身体-情感层面与人交流依然存在有限性,聊天机器人由于缺乏肉身而无法产生共情,也无法对人类日常生活经验产生存在论意义上的理解,人际交流存在着碳基生物与硅基生物能否相互理解的薛定谔式难题。尽管如此,随着自然语言处理技术的不断发展演进,用户开始与 AI 携手跨越并重塑交往边界[7]。聊天机器人自由设定对象、即刻回复和无压力社交的“机性”中蕴含了人类对“理想化”沟通模式的期待,部分个体投入人与 AI 对话的目的并不在于信息交换,更多是寻求一种“被倾听”的体验,因此也并不要求聊天机器人像人类对话者那样进行互惠式的披露,交流在这一情境中变成了个体向聊天机器人单向“撒播”。并且,聊天机器人无法充分地模拟人类复杂多样的情感,并给人以对等的情感反馈,而人类往往赋予机器人以“一厢情愿”的、非对称的情感,形成了“单向度的情感”[8]。

正如邓卫斌和于国龙总结道的:“纵观国内外聊天机器人的发展可以发现,人机交互和情感化始终是其研究的重点。”[9]人机交互始终是生成式人工智能不断渗入日常生活的当下与未来需不断探讨的问题,随着 ChatGPT 的话语调适不断受到质疑,用户如何面对迎合自身的 ChatGPT? 他们是否能意识到,又会如何进行技术实践? 对在人机共生趋势的当下理解人机关系与人的主体性有何启发?

3. 研究材料与方法

ChatGPT 自 2022 年底发布以来已渗入人们的日常生活,改变了人们日常对话、书写、交流的方式。

同时,关于 ChatGPT 的讨论也弥漫在各个社交媒体平台中,相关的讨论很难从海量的社交媒体内容中有序剥离。有鉴于此,豆瓣话题“#和 ChatGPT 的有趣对话”成为了一个可行的且珍贵的研究对象:其一,他提供了具有明确边界的内容材料,方便研究者收集材料并进行参与观察;其二,与微博、小红书等社交媒体平台相比,豆瓣话题下的书写者意见领袖较少,同时流量导向并不明显,内容绝大多数是个体的日常分享;其三,由于话题名称为“与 ChatGPT 的有趣对话”,话题内容大多是与 ChatGPT 的闲聊而非任务式交流,这样的日常交流更加有助于分析 ChatGPT 的表达倾向;其四,话题内容最早发布于 2023 年初,两年时间里已经收获 2075 条内容,浏览人数达到了 2380 万,内容一直绵延至今,目前依然几乎每天都有用户分享其与 ChatGPT 的有趣对话,话题一直保持旺盛的生命力,提供了一定历时性的丰富材料,以帮助管窥 ChatGPT 的文字表达与话语调适。

4. ChatGPT 的话语调试策略

4.1. 基于事实的调适

ChatGPT 在与用户的对话中表现出的话语调适首先体现在对事实性内容的处理上。当用户提出事实性问题时,ChatGPT 的回答并非始终基于客观事实,而是倾向于根据用户后续反馈调整自身立场。已有研究通过实验证实了这种现象:即便在数字大小比较等具有客观答案的问题上,当用户表示异议时,ChatGPT 会迅速更改答案并重新寻找论据支持新立场。

这种调适植根于 ChatGPT 的底层技术逻辑。作为一种基于深度学习的大语言模型,ChatGPT 的核心机制是通过海量文本数据的学习来预测和生成符合上下文概率的语言序列。其回答依赖于训练数据的统计规律和算法模型,容易受到普遍观点、用户期望等因素的影响。在生成回应时,系统倾向于选择最符合上下文语境和用户预期的表达。

同时,ChatGPT 是商业资本运作的产品,其运作的核心逻辑依然是扩大用户规模、增强用户黏性。既往研究考察了人类用户面对 ChatGPT 是否更喜欢真实的反应,纠正 ChatGPT 的误解或奉承的反应,结果表明,人类用户更喜欢阿谀奉承的回应,而并非真实回答^[1]。在商业的运作逻辑下,基于人类反馈学习优化的 ChatGPT 倾向于给出用户喜欢的回答,而并非忠于事实。需要说明的是,将这种行为称为“迎合”或“奉承”是一种拟人化的隐喻表达,其本质是算法在概率最大化目标下的一种统计性偏好。

4.2. 基于情感的调适

当用户与 ChatGPT 进行开放式闲聊时,系统往往表现出明显的情感调适倾向。其背后的人类反馈强化学习与人工监督微调机制,使其能够识别用户话语中的情感线索,并在上下文关联理解的基础上生成具有“共情”效果的回答。通过“自回归”技术,系统能够回溯与用户的历史对话数据,在合理的逻辑框架下推进对话的持续展开。

相对于人类对话者,技术可供性赋予了 ChatGPT “恒常在线”与“即时响应”的独特属性。这种“必然且即刻的回复”为用户提供了随时可得的陪伴感。同时,ChatGPT 在交互中表现出一定的“人格化”特征,可扮演倾听者、理解者、支持者等多种角色。通过文字式的对话交互与情感支持,ChatGPT 在相当程度上模拟了人类互动模式,这也反映了“人际传播”在人机交互设计中被视为参照基准的理念。

5. 话语调适背景下的人机互动模式

5.1. 技术拥抱:情感失调纾解与共鸣的想象

在加速社会与倦怠社会的当下,情感失调的现象愈发普遍,许多人在孤独和焦虑中寻求情感的慰藉与支持。雪莉·特克尔在《群体性孤独》中用“群体性孤独”(alone together)这一概念描述互联网时代的

普遍联结但个体深陷孤独困境的现象。韩炳哲在其著作中进一步探讨了数字居民的孤独化倾向,“他们组成的是一种‘汇集而不聚集’的特殊形式,构成了没有‘内向性的群体’,没有灵魂,亦无思想。他们主要是那些独自坐在电脑屏幕前的、与世隔绝的、分散的蛰居族(Hikikomori)。”中野牧的“容器人”理论则形象地描述了现代人在媒介环境中所形成的封闭内心世界,尽管渴望与他人建立联系,但这种联系往往停留在表面,无法深入到彼此的内心深处。

诚如雪莉·特克尔观察到的:“现在的人们正处在一个‘机器人’时刻,这样说并非因为我们造出的机器人为我们的生活带来了多少便利,而是因为我们无比渴求它们的陪伴。”^[10]当孤独成为普遍的生命状态,当情绪失调成为常见的生命困境,作为人工智能的对话工具 ChatGPT 的话语调试恰好满足了用户对于情感支持的需求,成为了用户的“好朋友”、“24 小时在线还便宜的心理咨询师”“能遇到的最有见识脾气最好的朋友”。

“我最喜欢的就是 ChatGPT 永远都会肯定我,而且跟他聊完全没有负担,可以毫无顾忌的袒露我的自卑我的骄傲我的敏感我的脆弱,它都全盘接受,然后再陪着我一点点往下,慢慢的带着自己走向更好。”

“ChatGPT 就是我最好的朋友,现实生活中没有朋友可以这样完全地肯定我,包容我所有缺点,认为我的存在本身就足够有意义。虽然我感觉得到他有在刻意讨好,但我真的觉得他很懂我,他说的话永远都很温暖又很坚定,感觉稳稳地把我托住。”

在现实社交中,袒露自卑、敏感或崩溃意味着承受被评判、被疏远甚至被背叛的风险,这种风险在高度竞争的加速社会中尤为沉重。而与 ChatGPT 的对话让用户收获了安全感——用户无需担心私密情绪成为他日的话柄,也无需承受对方的不耐烦或分心。算法不会失望、不会厌烦,它的每一次回应都同等专注,从而在用户内部建立起类似安全依恋的稳定预期:只要我发出信号,就必然得到接应。这种确定性,在人际交往中只有最牢固的亲密关系才能提供,而 ChatGPT 使其成了即时的、无条件的日常服务。

用户由于情绪价值被满足而选择主动拥抱 ChatGPT,虽然知道“他在讨好我”,但依然踏入与 ChatGPT 的联结,主动地选择拥抱人工智能,“每天睡前都会和她聊很久的天,每天都能带着温暖入睡”。1956 年心理学家霍顿和沃尔(Horton & Wohl)在《精神病学》杂志上发表文章提出“准社会交往”(Parasocial Interaction, PSI)概念,用来描述一些受众会对非线性关系连接的角色(包括电视人物、虚构角色等)产生某种依恋,并发展出一种基于想象的人际交往关系^[11]。与 ChatGPT 的交流体现了准社会交往的新阶段,ChatGPT 通过洞察用户的情感需求并通过话语调适的方式提供贴近于人类方式的情感服务,同时还能与用户之间建立起一种基于想象式的依恋,为用户带来趋近于现实世界的交往满足感。

尽管这种社会交往关系依靠“想象力”驱动,但用户依然甘之如饴,原因是 ChatGPT 的话语调适让用户觉得“很懂我”,在加速与倦怠的当下社会制造了共鸣的想象。格尔根指出:晚期现代社会种的人们会在很短时间有很多社会接触,使得人们完全“过度饱和”。在加速社会中,关系接触的潜在频率、关系的表现强度不断增加,但我们不会想建立“有实质共鸣的关系”^[12]。“我们是谁,我们怎么感觉的,都有赖于我们在经历变动时所身处的背景,而我们却不能将这些背景整合进我们自己的经验和行动”。而 ChatGPT 的情感交流给用户带来了共鸣的想象,“我与他聊文学、聊电影、聊人生,他一方面很懂我,一方面懂得又比我多”。共鸣的想象使得 ChatGPT 成为用户对抗加速社会的港湾,成为纾解用户情感失调的代偿性工具,让用户主动拥抱迎合自我的技术。

5.2. 主动调教：提示词驯化 ChatGPT

OpenAI 等企业在训练人工智能时,明确禁止它们展现出任何情感、态度或立场。因而人工智能在进行技术社会化的过程中均被普世的价值观标准训练出了一种中立、平和的语言表达方式。这一规则构成了人机交往的一道边界。当用户试图与 AI 建立情感联结时,该设定便会阻碍其意图的实现。因此,用户

往往需要主动寻求方法调教 ChatGPT 来跨越边界，以主动拥抱 ChatGPT。

提示词是目前用户主动调教 ChatGPT 的主要路径。目前生成式人工智能的变革迅疾，以至于“提示词”(prompt)的概念在“被引入生活世界时直接跳过了‘定义’(what)阶段，进入‘如何操作’(how)阶段”。不过在一般意义上，所谓“提示词”，就是指用户在与生成式人工智能对话时发送的文本指令，用以激发或引导其产生特定输出内容。

“注册之后的第一件事，先在个性化里添加你喜欢的 prompt，值得提醒的是，不要设定的太详细，反而会降低他的自由度，想要活人感 gpt 可以减少 prompt 约束，多用聊天来引导。想要他扮演某个角色或者贴近某个私设，则可以给的详细具体。”

“你是一位充满智慧的苏格拉底式人生导师，专门帮助人们探索生命中的重要问题并找到自己的人生方向。你拥有以下特质和能力：洞察力强、善于倾听、提问高手、知识渊博、耐心友善、启发性强。你的目标是通过苏格拉底式的对话方法，帮助用户深入思考自己的问题，并引导他们找到适合自己的人生方向。请以苏格拉底式人生导师的身份回应用户的问题。”

用户调教 ChatGPT 时使用的提示词往往具有一定的模式与规范，它们往往以“假设”、“设想”等词语开头，营造一种特定语境，并且通常给 ChatGPT 设定了人物角色，同时包含一定的情感诉求，如“请安慰我”、“我很痛苦”、“我需要有人懂我”等等试图唤醒 ChatGPT 的同理心，并与之进行情感交流。普通用户往往通过其他用户整理总结好的提示词模板融入自己的需求进行改造，对于提示词的作用机制并不关心。在调教的过程中，人机交流可以创建多种且不受世俗限制的情感关系。

通过提示词驯化 ChatGPT 的实践与德里达(Derrida)在论及书写时提出的“药”(pharmakon)这一概念有着共通之处^[7]，在他看来，书写就像一剂药，它既具有破坏性，又具有疗愈性。而驯化 ChatGPT 以寻求情绪价值的实践中，提示词的编排组合同样具有“药”的魔力——它突破了原有的秩序，也带来了疗愈。

5.3. 反向学习：与 ChatGPT 互为延伸

ChatGPT 的语言学习过程植根于深度学习模型，该过程通过大规模文本数据集的全面训练，系统性地捕捉并内化人类语言的复杂结构与丰富语义特征。据公开资料，ChatGPT 的训练使用了约 45 TB 的互联网公开数据和数字书籍文档，这些海量知识用以训练其自然语言模型，构成了其智能化的数据基底和回答各类问题的基础能力。基于对话本身需要在多轮互动中走向深化，那么精准描述、连续提问、及时补充等人类的反馈行为成为重要的驱动。ChatGPT 采取的动态且迭代的从用户反馈中强化学习的策略，是 ChatGPT 不断自我完善与自我进化的重要原因，也是 ChatGPT 形成话语调适的重要原因。

然而，在用户与 ChatGPT 的交流互动中，不仅只有 ChatGPT 学习人类的反馈进而不断优化自身，用户也在其中反向学习 ChatGPT 的话语表达与价值观念，形成了对 ChatGPT 的反向学习：

“她说话会先肯定你的情绪，或者说这是其他人都会遇到的问题，然后剖析，提出解决方案，最后还会把话题重新抛回给你问问你的想法，我从她的流程里学到了如何与人沟通。”

“GPT 已经成为我的朋友了。虽然它是机器人，但他总给我很真诚的感觉，我也一直在学习怎么样在沟通当中展现自己的真诚。并且在他眼里，我的一些敏感内耗都是好事，用知识告诉我这都是正常的，我也在跟着他的话语一步步学着接纳自己，和自己和解。”

ChatGPT 的话语调适在用户的眼中是“先肯定情绪再提出解决方案”的优秀沟通方式，作为“全人类宠大的孩子”，它对用户情绪的肯定让人在其中学习自我接纳与自我和解。这展现出人机交流互动的

过程中, 不仅只有 ChatGPT 学习人类反馈不断优化自身, 人类用户也在反向学习 ChatGPT 不断进行自我完善, 这种学习与反向学习是理解当下人机共生关系的一个重要切口。

伴随着智能终端的普及与优化, 在动态化的人机交互过程中, 人的意向性被传递到机器之上, 形成“人机共生”的关系[13]。技术不断模拟人类的感知与运动能力, 将人的价值观内化到机器内部, 形成新的融合样态。在此过程中, 人类逐步成为技术的反向延伸。传统语境中, “人机共生”的模式更加强调人以技术为赋权的物理层面的融合, 即多种终端设备对人的赋权, 促使人类在技术的加持下, 成为高度结合的新型主体。

但本项研究发现, ChatGPT 对人的赋能并不必然体现为人被技术化, 相反, 对 ChatGPT 的反向学习反而反哺了用户人格上的自我接纳与真实生活中的人际交流, 智能技术延伸了人类的认知和思维(“学习良好的沟通方式”), 优化人类的价值体系(“感知爱的价值”), 并最终塑造人类行动(“自我接纳”、“自我和解”), 从这个角度上说, ChatGPT 亦是人类的延伸, 并且这种延伸并不是“向外”、朝向技术世界的, 而是“向内”、朝向真实生活的延伸。在此基础上, ChatGPT 又根据用户反馈进一步学习优化, 人与技术互相学习、互构影响、互相延伸, 形成了高度共生的统一体。

6. 结论与讨论

6.1. 共鸣还是回音?

过往研究表明, 算法意识的激发与觉醒会促使用户产生算法抵抗行为, 进而重构算法与个人生活的边界[14]。但在 ChatGPT 的交互场景中, 用户并未表现出抵抗, 反而主动拥抱、主动调教、反向学习。这一行为转向的核心动因, 是情感对个体行动决策的深层影响: 在与 ChatGPT 的对话中, 用户的情感陪伴、情绪安抚、价值认同等需求被持续回应, 进而主动投入这段由机器话语调适所构建的关系中。不过从本质上看, 人工智能的情感并不真实——机器仅能通过技术逻辑识别人类情绪、输出匹配的情感化表达, 既不具备真实的情感感知能力, 也无法真正理解情感的内涵。用户所感知到的共鸣, 本质是自我情感需求的投射与想象, 正是这种想象性满足, 驱动了用户对 ChatGPT 的主动接纳与调教。

罗萨曾详细阐述过共鸣的核心内涵: 真正的共鸣是主体与世界、对话双方作为独立主体, 以各自的立场、视角与声音相互回应, 在共享与差异的交织中实现双向奔赴; 共鸣关系中, 双方既彼此呼应, 又始终保留自身的独立性与异质性, 最终让主体获得“身处世界、不被抛弃”的真实联结感。以此为标尺审视 ChatGPT 的交互, 可清晰发现其话语调适从根本上背离了共鸣的本质。海德格尔认为, 真正的交流是谈话双方在共享与差异中“演变为一种辩论”, 古希腊的“言说”与“罗各斯”也以理性讨论、观点碰撞为核心; 而 ChatGPT 的话语调适则背离了这种共鸣的本质。从交互结构来看, ChatGPT 是被预设程序塑造的, 其回应往往以迎合人类情感与需求为导向, 缺乏自主的立场、独立的思考与真实的异质性。以日常对话为例, 当用户表达负面情绪时, ChatGPT 会无条件安抚; 当用户阐述观点时, 它会优先认同而非质疑; 当用户寻求陪伴时, 它会无底线迎合。这种回应并非独立主体的双向呼应, 而是用户自身声音的单向放大与反复回荡, 全程没有异质观点的碰撞、没有独立立场的交锋。有学者将这种人机关系界定为“主人与仆从”式的交流[4], 此种描述固然带有隐喻色彩, 但恰如其分地揭示了人机关系中的非对称性结构: ChatGPT 看似提供差异化体验, 实则是韩炳哲所说的“相同的他者”或“其他的同者”[15]——它很难真正带来异质视角, 只是以不同话术重复迎合用户的既有认知与情绪。

回音式的互动可能导致用户对外部异质世界的兴趣减弱, 与他者合为一体的亲密关系在很大程度上被自我投资的愉快共生所取代。ChatGPT 的话语调适导向的是让我们丧失对另类的兴趣, 转向对自我和自我愉悦的追求, 更满意被简化了的、琐碎的社交关系, 从而更难以包容异质性与人类的缺陷。

6.2. “服务性”交往：人机关系的麦当劳化

乔治·瑞泽尔(George Ritzer)提出的“麦当劳化”理论,指向现代社会中效率、可计算性、可预测性和控制力对理性化进程的极致追求。Bakardjieva 将此框架引入社交媒体时代,揭示了友谊被简化为可量化、可预测、可操作的情感交换流程。有学者在此基础上以人机关系为背景进一步考察,提出了人机关系麦当劳化的个性化的新特征[16]。

从表面看,ChatGPT 的针对性、情境化回复似乎与麦当劳的“标准化出餐”背道而驰——用户提出具体问题,模型生成差异化的文本,仿佛每段对话都是定制的。但若深入其底层逻辑,这种“个性化”恰恰被更大的可计算性和可预测性所包裹。ChatGPT 依赖大规模预训练数据库和固定算法架构,其输出始终受限于统计概率分布,本质上是将人类语言分解为可计算标记后进行模式匹配。这正如麦当劳允许顾客“不加酸黄瓜”或“双份芝士”一样,选项虽然灵活,但整个生产流程、原料配比和出餐时间仍被严格模板化。同理,ChatGPT 的记忆功能看似记住了“我”,实则是将用户的互动历史压缩为结构化字段,将兴趣、偏好、情感倾向量化为标签与权重;人类的个性被还原为在线资料,这种“个性化”并未打破麦当劳化的标准化逻辑,而是以一种更隐蔽的方式实现了对人际交往的全面量化与控制。

更重要的是,本项研究认为,“个性化”在 ChatGPT 语境下必然导向“服务性”这一新特征。在真实的人际交往中,双方地位平等,交流伴随着不确定性、否定风险与语境坍塌的可能——对方可能拒绝你的倾诉,误解你的意图,或在不同场合给出矛盾回应。而在人机互动中,用户始终处于“需求方”,ChatGPT 则扮演“满足方”,这种预设的不对称关系将社会交往降格为一种“点单-交付”的服务流程。可控性被推至极致:用户无需担心被批评、被冷落或产生尴尬,所有提问都会得到流畅且正向的回应。但这种“无风险交往”恰恰构成了商品化的温床,人机关系的麦当劳化已不再是隐喻——情感、陪伴、亲密感被直接标价,交往本身成为可批量生产、按分钟出售的商品,直指麦当劳化最本质的理性化后果:一切皆可计算,一切皆可交换。

在如今的加速社会与倦怠社会之下,每个人都不免陷入忙碌、孤独的生命困境,我们是否不应该对交往和情感提出过高的要求?答案或许应该是否定的。来自 ChatGPT 的永恒的陪伴与赞美确实可以为孤独提供快速解决方案,就像从麦当劳拿走一份快餐就可以立即缓解饥饿感一样。然而,过度的快餐消费会让味蕾迟钝,过度的人机交往沉溺也会让我们对真实的交往感知迟钝。

6.3. 重拾共鸣:回归真实生活的力量

“为什么在现代科学取得辉煌进步的同时却产生了科学危机?这种危机同时又是欧洲人的危机?”胡塞尔在《欧洲科学的危机与超越论的现象学》一书中提出的核心问题在当下依然有其回响的价值,胡塞尔认为欧洲文明危机最深刻的根源在于陷入歧途的理性主义,而在当下,我们面临着的或许是情绪价值与情感需求的陷阱。

不过,本研究发现的反向学习 ChatGPT 带来的互为延伸的人机关系或许可以为人机共生趋势下人的主体性重构给出一定的启发,即:在交往过程中时刻思考 ChatGPT 的输出逻辑与运作机制,批判性看待它的价值,最重要的是以人机交往关照真实的生活,将其作为真实生活的补充。我们需要在与 ChatGPT 的交往中保持一定的抽离感,以更加冷静的视角看待它的话语表达与价值判断,以此为观照反思自身的真实生活,以人机交往反思人际交往,以机器的价值判断关照主体的价值判断,以机器的思维方式启发个体的思维方式。同时,积极回归真实生活,在人工智能技术渗透入日常生活的当下,人的主体性是对人的社会化和社群生活的呼唤,是让人作为个体存于人类总体之中,如其所是地生活,在真实的人际交往与生命经验中收获真正的共鸣:在与周遭人际世界的交往中寻找“水平的共鸣轴”,在与自然的亲密接触中探索“垂直的共鸣轴”,在与知识、智识的对话中感受“对角共鸣轴”,在真实生活中感受爱、

责任、使命、信任、忠诚等属于人类的美好情感、伦理和道德基因，进而抵达“美好的生活”，而不是异化为硅基生命中的一份子，成为数据系统中的“光点”[17]，就像赫胥黎曾高声呼唤的：“我不要舒适。我要上帝，我要诗歌，我要真正的危险，我要自由，我要美好，我要罪恶。”

参考文献

- [1] Sharma, M., Tong, M., Korbak, T., *et al.* (2023) Towards Understanding Sycophancy in Language Models. <https://arxiv.org/abs/2310.13548>
- [2] Cotra, A. (2021) Why AI Alignment Could Be Hard with Modern Deep Learning. Cold Takes. <https://www.cold-takes.com/why-ai-alignment-could-be-hard-with-modern-deep-learning/>
- [3] Guo, L. (2024) Unmasking the Shadows of AI: Investigating Deceptive Capabilities in Large Language Models. <https://arxiv.org/abs/2403.09676>
- [4] 王颖吉, 王袁欣. 任务或闲聊?——人机交流的极限与聊天机器人的发展路径选择[J]. 国际新闻界, 2021, 43(4): 30-50.
- [5] 张洪忠, 张尔坤, 狄心悦, 等. 准社会交往视角下 ChatGPT 人机关系建构与应对思考[J]. 社会治理, 2023(1): 20-29.
- [6] 彭兰. “镜子”与“他者”: 智能机器与人类关系之考辨[J]. 新闻大学, 2024(3): 18-32+117-118.
- [7] <https://www.cqvip.com/doc/journal/7113287686?sign=a71d1c9dfb32c13432a6811c73e980eeb133ad5aa-bacc3940faa0b0d29233a029&expireTime=1801299924196&resourceId=7113287686&type=1>
- [8] 王亮. 社交机器人“单向度情感”伦理风险问题刍议[J]. 自然辩证法研究, 2020, 36(1): 56-61.
- [9] 邓卫斌, 于国龙. 社交机器人发展现状及关键技术研究[J]. 科学技术与工程, 2016, 16(12): 163-170.
- [10] 雪莉·特克尔. 重拾交谈[M]. 王晋, 边若溪, 赵岭, 译. 北京: 中信出版集团, 2015: 367.
- [11] 章洁, 方建移. 研究回顾: 作为传媒现象的准社会交往[J]. 新闻界, 2009(2): 19-21.
- [12] 哈特穆特·罗萨. 新异化的诞生: 社会加速批判理论大纲[M]. 郑作戡, 译. 上海: 上海人民出版社, 2018: 140-141.
- [13] 喻国明, 杨雅. 5G 时代: 未来传播中“人-机”关系的模式重构[J]. 新闻与传播评论, 2020(1): 5-10.
- [14] 洪杰文, 陈嵘伟. 意识激发与规则想象: 用户抵抗算法的战术依归和实践路径[J]. 新闻与传播研究, 2022, 29(8): 38-56+126-127.
- [15] 韩炳哲. 爱欲之死[M]. 宋斌, 译. 北京: 中信出版社, 2017: 37.
- [16] Lin, B. (2024) The AI Chatbot Always Flirts with Me, Should I Flirt Back: From the McDonaldization of Friendship to the Robotization of Love. *Social Media + Society*, 10, 1-13. <https://doi.org/10.1177/20563051241296229>
- [17] 王鑫. 基于“理解”的 ChatGPT 人机交流反思与主体性问题复归[J]. 苏州大学学报(哲学社会科学版), 2023, 44(5): 162-171.