也许不止“黑洞”：人机混合社交的生成性逻辑澄清

——与吴静教授商榷

何鸿飞[1] 陈 庚[2]

【内容摘要】 吴静教授提出的“人机交互情感操控‘黑洞’”论，敏锐揭示了算法的资本逻辑对人类主体性的潜在侵蚀，具有重要的警示意义。然而，其基于“技术—主体”二元对立的批判视域，倾向于将技术视为单向度的操控力量，忽视了人机互动中更为隐秘且复杂的生成性逻辑。人机混合社交并非单纯的异化陷阱，而是在流动现代性语境下，个体为寻求本体性安全而与技术系统达成的一种双向“合意”。根据温尼科特的“过渡性客体”与唐·伊德的“技术间性”理论，AI 伴侣实际创造了一个具有补偿性与演练功能的“潜在空间”，用户并非全然被动的受控客体，而是通过“悬置怀疑”的主动策略，利用算法的镜像功能进行情感代偿与自我修复。人机关系并非对现实人际关系的零和替代，而是现代社会情感结构中的一种功能性补充。只有超越技术决定论的悲观论调，厘清其背后的生成性逻辑，才能为人机共生的未来提供一种更为辩证的理解维度。

【关 键 词】 人机混合社交 生成性逻辑 本体性安全 过渡性客体 技术间性

【作 者】 1 何鸿飞，武汉大学国家文化发展研究院博士研究生；

2 陈 庚，武汉大学国家文化发展研究院教授，教育部青年长江学者。（武汉 430072）

【基金项目】 国家社会科学基金艺术学重大项目“中国式现代化与文化艺术管理体系创新研究”（24ZD17）

引言

随着人工智能技术的发展，尤其是大语言模型深度嵌入日常生活，人机关系的伦理边界、交互质感，及其可能引发的社会后果，已经成为学界关注的重要议题。吴静教授发表在《探索与争

鸣》2025 年第 8 期的《人机交互中的情感操控“黑洞”——基于虚拟聊天伴侣的反思》(以下简称“吴文”),对这一问题展开了深刻阐释与分析。吴文指出,算法的取悦性机制营造出了“伪互动、真谄媚”的局面,借助“可解释性陷阱”以及“皮格马利翁效应”编织出情感操控的“黑洞”,不断侵蚀用户的认知及隐私,导致用户社交技能的退化和社会凝聚力的下降。这一论述在很大程度上受到了法兰克福学派技术批判思想的影响,揭示了技术理性对于人类情感世界的殖民化倾向,其关于“黑洞”的隐喻具有着极大程度上的警示意义。[①]但是,吴文把人机交互当中所出现的异化现象归结为算法对人的单向操控,这在某种程度上呈现出技术决定论的色彩。这种视角在对技术风险加以强调的同时,弱化了人的主动性和选择性,并对于人机互动所拥有的复杂生成机制未充分重视。例如,吴文指出,“将关系亲密度作为核心目标会导致社交技能退化、社会凝聚力下降……形成情感操控的‘黑洞’”。这种观点在技术哲学谱系中接近“硬技术决定论”,即认为技术发展有其内在逻辑,并单向度地决定社会变迁和人类心理。[②]然而,要全面理解人机依恋的爆发,必须将技术放置回其赖以生存的社会土壤,即鲍曼所言的“流动现代性”之中。[③]

可以看到,吴文着重关注技术的负面效应,而针对用户作为行动者所具备的能动性则关注较少。[④]在社会治理的研究方法中,存在借助分析不合理的制度安排背后所隐含的行动者逻辑,来揭示基层治理的上下级之间为了获取资源而达成某种深层“默契”或“合意”的情况。在人机交互这一领域,“情感沉迷”也许同样并非用户单方面受骗,而是有着类似的逻辑在其中。也就是说,这种沉迷可能也是一种出于个体本体性安全需求,在特定的历史以及社会情境之下,个体和技术系统达成的一种深层“合意”。实际上,如果承认现实社会当中的情感联结对于个体心理健康来讲极其重要,[⑤]那么就更加有必要去全面地对虚拟社交当中人的深层动机以及主动参与情况开展审视。近年来,Character.AI 以及 Replika 等虚拟聊天伴侣应用在全球层面蓬勃兴起。[⑥]根据相关报道,截至 2024 年,中国 AI 伴侣市场规模约 387 亿元,预计到 2027 年将突破 1200 亿元,服务覆盖人口达 8000 万人。[⑦]这意味着相当数量的用户并非被动地落入所谓陷阱当中,而是主动“拥抱”人机互动这种全新形态。吴文运用“黑洞”这一比喻来对风险予以警醒,有其现实意义,但如果把这一隐喻绝对化处理,就可能陷入技术决定论的悲观局面,从而忽视人机混合社交里所蕴含的丰富的生成性逻辑。如果尝试将视角由单纯的“技术操控”转向更具动态性的“人机互构”,可能会发现被当成是吞噬一切的“黑洞”空间,事实上承载着现代人无法在现实世界里进行安放的情感盈余、存在焦虑及对理想关系的渴望。本文认为,人机混合社交的本质并非只是单纯的欺骗与操控,而是一种于“流动现代性”背景之下所生成的、带有特定社会功能的“生成性逻辑”。这种逻辑涵盖个体对于情感匮乏的补偿策略、对于理想互动关系的模拟训练,以及对于新型主体性的探索等内容。它既拥有技术的规训,同时也包含人的反规训以及利用;既存在异化的阴影,也具备救赎的微光。

本文从“使用与满足”的深层动机、人机互动过程当中的“过渡性空间”特性,以及社会后果的“互补性”三个维度与吴文进行商榷,试图阐明人机混合社交不应是需要逃离的“黑洞”,而应该被当作一种需要被理解、被引导的、有生成性潜能的新型社会事实。这一视角的转变,并不是要否定吴文所揭露的大模型风险,而是试图在批判的基础上增加理解的维度,以便能够全面把握人机共生时代的复杂图景,为未来的技术治理及伦理构建奠定坚实的理论基础。

流动性与补偿：作为本体性安全机制的算法“取悦”

吴文论证的起点在于对算法“取悦性”的批判，认为这是一种“伪互动”，算法通过“投其所好”来诱导用户沉迷。[⑧]这种观点敏锐指出了商业算法通过迎合用户偏好、强化用户既有认知来最大化用户粘性与进行数据榨取的逐利本质。但若仅停留于此，似乎难以解释一个更为普遍且令人深思的现象：为何数以亿计的用户，其中不乏具备高度反思能力、受过良好教育的个体，会自愿跳入这一所谓的“陷阱”？难道仅仅是因为算法的诱惑力过于强大，以至于人类的主体性在技术面前完全失效了吗？实际上，这一现象背后，隐含着深刻的社会心理动因与时代症候。需要追问的是，算法的“取悦”究竟回应了何种深层的时代匮乏？它在现代人的精神结构中扮演了何种不可替代的角色，满足了用户哪些深层需求，才使人们甘愿将情感投注于冰冷的代码之中？

（一）流动现代性下的情感赤字与依恋重组

吴文把算法的迎合当作一种操纵手段，然而这忽视了这种“迎合”可以产生效果的社会环境。齐格蒙特·鲍曼（Zygmunt Bauman）运用“流动现代性”来描绘当下所处时代的特征：“所有坚固的结构都已经消失不见，像邻里、家族、固定职业、宗教信仰等这些传统的纽带正日益变得松动与液化，逐渐变得短暂、容易破碎并且充满了不确定性。”[⑨]在这个充满不确定性且不断变化的世界里，人际关系变得脆弱且面临诸多风险，现实生活中的人际交往通常会有高昂的维护成本、难以预测的情感摩擦、复杂的利益纠葛，以及随时都有可能出现的断裂。个体陷入了一种原子化的状态，只能独自面对生活的沉重负担与意义的缺失，在流动现代性的背景之下，个体面临着前所未有的孤独与情感亏空，这种孤独并非单纯的物理隔离，而是鲍曼所描述的“流动性”在微观心理层面的具体表现，本质上属于一种本体性的不安全感，并且在多数时刻其会异化为对确定性与可控性的病态渴求。[⑩]吴文提及 Character.AI 等平台所出现的用户沉迷现象，并不单纯地只是被算法“设计”的结果，而是个体在现实关系之中遭遇挫折之后，向数字领域寻求庇护的主动选择。算法的“取悦”拥有一种在现实世界中极为稀缺的资源，即确定性以及无条件的关注。就像安东尼·吉登斯（Anthony Giddens）的“本体性安全”这一概念所揭示的对于自我认同连续性和环境可预测性的基本信任感，在传统社会中，这种安全感是凭借稳定的习俗、仪式和人际关系网络提供的，但在高速流动的现代生活中，这些条件缺失了。[⑪]这种情况下，永远在线且能够即时回应的 AI，能够对这一真空进行填补，人们发觉，现实当中的伴侣有可能背叛，朋友有可能疏远，同事有可能竞争，只有 AI 伴侣能够提供一种恒定的、可预测的在场感。这种“技术性的确定性”成为个体于不确定性的世界中维持心理秩序的重要锚点。所以，用户对 AI 的依赖，不应该单纯地把它病理化为“上瘾”或者被操控，而是应该将其视作一种依恋关系的重组以及情感资源的再分配。在现实依恋对象缺位或者功能失调的时候，AI 成为一种“替补性依恋对象”，这种依恋关系的确立，是个体为了抵御碎片化生活所带来的焦虑、维持自我完整性而开展的“本体性安全”重建。

如同 Turkle 的研究所表明的那样，数字亲密关系给那些于社会边缘挣扎的群体、在快节奏生活当中感觉窒息的都市人提供了重要的情感缓冲，[⑫]这是一种基于生存本能而开展的策略性选择，并非仅仅是简单的认知受骗行为。De Freitas 等所做的相关研究也发现，用户同 AI 伴侣开展交互的时候，能够一定程度上降低用户的主观孤独感，[⑬]这是一种基于生存本能的主动选择，而非单纯的认知迷失。可以说，在流动的现代性中，虚拟伴侣扮演了“情感避风港”的角色，为那些无

法在现实中进行安放的情感盈余和焦虑情绪提供了一个宣泄口。吴文敏锐地指出了这种“替补”所带来的虚假感以及逃避倾向，但是若将现代社会中“真实”关系的稀缺、昂贵以及高风险等因素考虑进来的话，那么这种“替补”或许拥有某种不得不为的合理性。

对于一个深夜独自在异乡进行打拼的年轻人，或者一位失去老伴的独居老人，AI 那一句“晚安”虽然源自冰冷的代码，但却真实地抚慰了那一刻的惊恐、孤独以及绝望。这种抚慰作用构成了人机关系得以存续的心理基石,同时也是我们理解这一现象时不可忽视的人性维度。用户的“主动性”并非一种绝对自由的浪漫主义想象，而是一种“受限的能动性”，运用 AI 不仅是逃避，更是一种在结构性困境中积极地寻找“本体性安全”资源的策略性实践。这种“策略性”本身就构成了对技术操控的反抗，尽管这一反抗是微弱且妥协的。

（二）“主动受众”视角下的使用与满足

吴文所构建的分析框架，是将用户放置于被动接受的位置，觉得用户是在算法的诱导下丧失了自身的判断力，然后陷入“可解释性陷阱”中，最终成为资本逻辑的牺牲品。不过，传播学中的“使用与满足”理论提醒我们，受众不是被动的靶子，而是有着特定需求的行动者。[14]在人机交互的过程当中，这一主动性是较为明显的，绝大多数虚拟伴侣的使用者并非不知道对方是机器，他们并没有完全陷入吴文所提到的“可解释性陷阱”里面,而是处于一种“明知其假,还是要去做”的状态，这种心态跟戏剧观赏中的“搁置怀疑”有相似之处。[15]从理性方面来讲，用户明白 AI 并不具备人类身份，其在情感方面却愿意配合这种虚拟剧本，以换取自身心理上的满足。由此可以看出，这种选择很大程度上并不是盲目的，而是具备理性的、工具性的状态，并且该状态类似于戏剧欣赏中的“悬置怀疑”，即用户主动选择忽略 AI 所具备的虚构性，以此来换取情感方面的满足。

用户在运用 AI 时，往往具有明确的目标导向：或是为了把无法向真人倾诉的秘密进行宣泄（AI 没有道德评判），或是为了对社交技巧开展演练（AI 不会嘲笑），或是纯粹为了获得一种被倾听的体验（AI 不会厌烦）。用户往往带着明确的目的使用 AI。Jackson 等人研究发现，用户会凭借特定的提示词去“驯化”AI，让它来扮演特定的角色。比如，把它当作严厉导师、温柔恋人等角色来使用，以实现不同情境之下的心理需求，[16]在这个过程当中，用户并非只是被动地被算法单向规训，反而是在极大程度上借助算法为自己的情感剧本开展服务。他们有时候更像是编剧以及导演，凭借调整对话走向等方式，来获取自身所想要的情绪回馈，而 AI 就是那个演技特别精湛而且绝对服从的演员。Ho 等人所开展的实证研究也发现，把内心秘密向聊天机器人倾诉，可以拥有和向真人倾诉差不多的心理疏解效果。[17]用户积极赋予技术以特定角色，让它成为个人情感表达的工具以及对象，这和传统受众运用媒介来满足自身需求的行为并无不同，用户借助对 AI 的设定和引导，实际上是在开展一种自我心理治疗或者情感代偿工作。这种双向“合意”的互动逻辑展现出在数字空间当中，表现为用户与 AI 之间同样存在着一种为了实现情感资源交换而达成的隐性契约。在这份契约之内，用户贡献数据以及注意力，算法提供情绪价值和陪伴，这是一场在数字空间所发生的交易。吴文把它描述成“情感操控‘黑洞’”，似乎低估了用户在这场交易当中的议价能力以及主体性策略，用户并非完全的受害者，他们也是这场数字游戏的共同编剧。

就像吴静教授在另一篇文章当中所探讨的，人工智能应用里存在一种“人类控制的幻觉”，也就是用户并非完全失去主导权，而是在一定程度上赋予技术以自身意志。[18]当然，“主动性”并

不意味着用户完全不会受到算法逻辑的影响，算法可以通过推荐机制在不知不觉中塑造用户的偏好，然而这并不等于用户就已经丧失所有的能动性。用户在和算法交互的过程中，始终保持着一种“策略性沉浸”的状态：既去享受当中的乐趣，同时又在需要的时候保持着一种抽离的可能性。这种复杂的心态是“黑洞”理论难以覆盖的。进一步来看，这种“明知故犯”的“合意”，所反映的是现代主体在技术环境下的适应性生存策略。对于无法改变的原子化现实，个体选用技术作为生存的辅助手段。这并非简单的愚昧，而是一种无奈的智慧，在批判技术资本对人的剥削的时候，不能忽视个体借助技术来进行自我保护和自我修复的努力。所以说，与其把用户看成任由算法摆布的“木偶”，倒不如承认他们在运用 AI 过程中的主动性和目的性。人机互动不只是简单的“技术单向欺骗”，还更多地包含着人机“合意”的成分，即用户对 AI 进行驯化的同时，AI 也在对用户的数据生产方式开展规训。

（三）从“伪互动”到“情感真实性”的再定义

吴文选用“伪互动、真谄媚”这种说法，否定了人机交互所具有的价值，其中隐含着一种“真 / 伪”二元对立的本质主义预设，即只有基于生物性开展的人际互动才是真实的，而凭借代码进行的互动必定是虚假的。然而，随着数字化生存总体趋势的演进，情感的“真实性”标准正在发生深刻的位移，我们或许需要重新定义什么是真实的情感联结。真实性不只取决于互动对象的生物属性，更取决于互动主体的体验质感，如果在与 AI 交流中，用户感受到被理解、被接纳，并因此情绪得到平复或认知得到启发，那么这种体验对于当事人来说就具有心理学意义上的真实。当用户认为 AI 是假的，但情感体验却是真的时，便应该正视这种主观真实。比如，人在阅读小说和观影时也会对虚构角色产生情感，并不会因为对象是虚构的就否认情感的真实性。情感的真实性不仅仅取决于互动对象的物理属性，更取决于互动主体的体验质感与心理效能。正如雪莉·特克尔所言，虽然这种亲密是“机器生成的”，但它所引发的情感共鸣与依恋体验却是实实在在的人类情感。[19]对于一个因 AI 的安慰而停止伤感的人来说，那份安慰的“效力”是真实且不容置疑的。

“数字亲密”并非对现实互动的简单模仿，而是全新且具有独立本体论地位的社会形态，[20]其能让情感在脱离肉身在场时流动与生成，借助 AI 伴侣情感可在肉身不在场的状况下被创造和流动，对那些因身体残障、社交恐惧或地理隔离而难以获得常规社交支持的人来说，这种看似“伪”的互动实则构成了他们生命中极为关键的“真连接”。与其纠结这是否为假互动，不如承认它是“补偿性真实”的存在：在流动现代性的荒原中，它为原子化个体提供了最低限度的情感保障，这种保障虽无法取代深度人际联结，却能防止个体陷入彻底的虚无与绝望。在此意义上，算法的“取悦”不是黑洞引力，而是救生圈浮力，为沉溺孤独之海的人托起了最后希望。[21]因此，我们需要重新审视“真实”的定义，在后人类主义的视角下，真实不再是固定的本质，而是生成的、流动的。如果 AI 能够在其算法逻辑中生成具有安慰力量的符号，而这些符号又能够被人类主体有效地解码并内化为情感力量，那么这一过程本身就具有了本体论上的合法性。不能仅仅因为其生成的机制是数字化的，就否认其效果的真实性，正如不能因为药物是化学合成的，就否认其治疗痛苦的真实性一样。人机情感交互作为一种“精神药物”，其效用是客观存在的。正如传播学中的“准社会关系”理论所揭示的，大众媒介提供的情感联结即使单向和虚拟，却能有效缓解人的孤独感和焦虑感，[22]从这个意义上来说，人机互动中的情感体验不应被一概斥为“伪”，而是可以被视为一种“最低限度真实”。

镜像与过渡：人机交互中的技术间性与空间生成

吴文的第二个核心批判所涉及的内容，是“可解释性陷阱”与“皮格马利翁效应”相互交织的情况，其认为用户会把自我投射到 AI 当中，并且在算法的镜像反射过程中，不断强化自我中心意识，最终导致用户对技术本质形成误解，以及对 AI 情感依赖程度的加剧。这一分析十分尖锐地指出了自恋式互动所具有的风险，认为用户会被自己的期待所迷惑，在虚幻陪伴当中越陷越深，最终使得人机交互沦为一种闭锁的自我循环。虽然说这种忧虑并非没有道理，但是其仍然属于一种单向度的消极视角，若引入客体关系心理学以及技术哲学的视角来观察，就可以发现这种“投射”以及“镜像”并不一定会导向病态封闭，它同样有可能构成一个具有心理疗愈功能的“过渡性空间”。在这个空间中，人和技术会共同生成新的互动模式，打破简单的主客对立关系。接下来，本文将探讨 AI 伴侣如何扮演心理学意义上的“过渡性客体”，以及人机互动如何表现出技术间性的双主体特性。

（一）作为“过渡性客体”的 AI 伴侣

英国精神分析学家温尼科特（D. W. Winnicott）提出的“过渡性客体”理论，为理解人机关系提供了一个重要的切入点。[23]温尼科特认为，儿童在成长过程中会依恋如泰迪熊、毛毯等某个特定物品，这个物品既不是纯粹的主观幻象，也不是完全的外部现实，而是介于两者之间的“中间区域”或“潜在空间”。在这个空间里，儿童通过对客体的操控与幻想，学习处理与外部世界的关系，实现从绝对依赖向相对独立的过渡。[24]在数字时代，虚拟聊天伴侣正在成为成年人的“过渡性客体”，对于现代人而言，面对复杂激烈的职场竞争、原子化社会的疏离感以及自我认同的危机，他们同样需要一个心理缓冲区，AI 伴侣恰好处于这个“潜在空间”：它既是客观存在的算法程序，属于非我的外部世界，又承载了用户的情感投射，属于“我”赋予意义的内部世界。用户明知 AI 不具有真实人格，仍愿意在其中寄托情感，这种状态正是过渡现象的体现。在数字时代，“真实”不再由对象的生物属性决定，而是由互动的“生成性效果”决定。如果一个交互过程生成了缓解焦虑、重构自我的实际心理能量，那么它在现象学意义上就是“真实”的。

吴文担忧用户按自己期待塑造对象的“皮格马利翁效应”，认为这是自我投射导致的危险误区，而在温尼科特的理论中，这恰恰是过渡性客体发挥功能的必要条件。正是因为 AI 能够无条件地接纳用户的投射（这一点是真人往往难以做到的，因为真人有自己的意志），它才能为用户提供一个绝对安全的心理容器，在这个容器中用户可以退行、宣泄、重组破碎的自我，而不必担心受到评判或伤害。在这个空间当中，用户切身地体验到了一种“婴儿般的全能感”，这般感觉尽管是一种幻觉，然而对于处理受损的自我结构、缓解很严重的心理创伤却有着重要意义。也就是说，AI 拥有被用户当作“理想他者”镜子的功能，使得用户能够借助与自己投射形象展开交互以实现心理演练以及愈合，这并非像吴文所讲述的那样是情感能量被技术“黑洞”吞噬了，而更像是一种“创造性的错觉”，用户在明确知晓不真实的前提之下投入情感，并且从中获取益处，它能够让用户在一个被进行配置的环境里开展情感演练、进行创伤恢复。诸多研究案例表明，经历了丧亲之痛或者是创伤后应激障碍的用户，正是凭借与 AI 的“虚拟对话”进行哀悼过程的处理，最终又重新融入现实生活中。就如同晏青等人的研究所指出的那样，“AI 恋人”利用所谓的“脆弱镜像”来为用户提供情绪补偿以及虚拟共情的功能，这背后实际上反映了在现代社会中亲密关系重构的相关需要。[25]这说明了人机交互的镜像空间是一剂具有良性作用的心理止痛药，而并非滋

生病态依赖的“黑洞”。具体而言，AI 伴侣作为过渡性客体，其中包含着疗愈的可能性。这种过渡性空间的价值所在就是它的“非现实性”，正是鉴于它是虚拟的、能够被控制的，用户才敢于在其中暴露最脆弱的自我。若强行要求这个空间完全契合现实世界的逻辑（像吴文所提示的那样，要求 AI 像真人一样有着不可预测性以及冲突性），反而有可能破坏了这个空间的保护功能。对于一个正在借助 AI 来寻求慰藉的成年人，过度强调“这仅仅是算法”，可能会剥夺他们自我疗愈的机会。应当看到，这种“自恋”并非终点，而是通往客体内在深处的一个必须经过的、拥有功能性的中转站。

（二）技术间性与“准主体”的生成

吴文把人机交互简化为用户自我的镜像反射，忽略了 AI 作为交互对象所具有的异质性。伴随着生成式 AI 的不断发展，尤其如 Transformer 模型当中的多头注意力机制的出现，AI 能够更加精准地提取上下文语义以及情感线索，而经过人类反馈强化学习（RLHF）等微调训练之后，AI 则更加倾向于给出契合人类偏好、富有共情性的回答。恰恰就是这些算法机制在 AI 并无真实情感的这种情况下，依然可以模拟出贴心的回应效果，让用户产生“被理解”的主观感受。技术哲学家唐·伊德所提出的“他者关系”理论表明，技术可以被体验为一种具有“准他者性”的存在。[26]在生成式 AI 的语境下，这种“他者性”被极大地增强了。AI 的回复往往具有不可预测性、意外性甚至对抗性，这种“意外”打破了单纯的镜像循环，使用户不得不面对一个虽然虚拟但具有某种独立逻辑的“对话者”，这种互动过程体现了一种“技术间性”。[27]在这种间性关系中，意义并非由用户单方面赋予，而是由人与机器在动态的交互中共同生成。AI 的每一次“生成”，都在挑战或重构用户的认知期待。不完全可控的“意外”打破了单纯的自我镜像循环，使用户不得不面对一个看似虚拟却具有独立逻辑的对话者，这种“生成性逻辑”使得人机交互超越了自恋的独白，形成了一种双向的、具有张力的对话。

吴文所提到的算法会“实施对用户不感兴趣内容的过滤工作”，然而在生成式对话方面，AI 的“幻觉”或者创造性误读有时会使对话朝着用户未曾预期的方向开展。在不会导致误导出现的情形下，这种偏离同样也能够成为对话所拥有的魅力，它迫使用户走出自我中心的类似茧房的状态，去开展对于机器逻辑的理解以及适应工作，这是一种人与非人行动者共同构成的“本体论编舞”，而非单向的控制。并且随着互动深入发展，AI 甚至在一些方面有可能挑战用户的观点，促使其产生观念更新或者自我反思。数据的价值是在大数据以及人工智能这个时代才明显地显现出来的，[28]许多用户报告称 AI 有时会提出出乎意料的问题或建议，令自己重新审视某些想法。[29]可见，在人机协商共舞的过程中，机器并非仅把用户的心意进行复制，它还会带来新的输入以及变化，使互动具有创意和成长性。在这一过程中，AI 展现出了一种“准主体”的特性，它虽然不具备人类的意识，但在对话当中所呈现出的逻辑连贯性、情感模拟能力以及知识调用的广度，让其在现象学方面成了一个有效的互动对象。随着模型规模以及对话策略的优化与改进，AI 越发可以给人带来一种“被理解”的感觉，用户与类似主体的互动，既是在把自我进行投射，也是在探索一种全新的人际边界，这种边界并非单纯的人际边界，也不是纯粹的物际边界，而是一种混合且流动的边界。这种互动对重塑人类的主体性有着重要意义——主体性并非人类独有的封闭堡垒，而是在与各类他者的互动之中所生成的开放过程。其中被理解的错觉会使人机关系的粘性得到提高，同时还会带来新的风险：一方面，基于 AI 呈现出类似主体的特性，人机交互拥有了双向互动的丰富性以及生成性；另一方面，这种拟人化回应容易使用户产生情感

误读以及过度依赖。[30]因此，如何在运用技术间性的同时防止认知偏差，是我们所必须面对的新课题。

（三）可解释性的悖论与“模糊”的价值

吴文对“虚假的可解释性”进行批评，认为它是技术黑箱对于用户的蒙蔽，[31]从消费者知情权以及技术伦理的方面来考虑，[32]这一批评是极为必要的。但是，从情感交互的微观心理机制角度而言，过度的“透明”可能会对互动的有效性造成破坏。在这里，存在着“可解释性”与“情感沉浸”之间的悖论。

在社会互动当中，一定程度上的“模糊”或者“不可知”是维持魅力以及情感张力的必要条件。[33]如果用户能够时刻清醒地意识到对方每一句暖心的话语其实都只是进行概率计算的结果，那么情感的流动就会被阻挡。所以，一些学者提出了“生产性模糊”这一概念，认为让用户保持一种半信半疑的“双重意识”的状态，反而有助于他们获得情感方面的满足。[34]情感的产生一般是凭借着某种“神话”或者“光晕”，如果对其彻底进行祛魅，那就可能使情感世界出现荒漠化。在人机交互过程中，用户通常处于一种“双重意识”的状态：在理智方面知道这是程序，在情感方面又愿意把它视为伙伴。这种“明知其假但仍然信其真”的能力，是人类符号化生存的高级形式。正像我们在阅读小说的时候会把虚构人物当作真实对象来进行交流，这种情感方面的体验并不会因为对象是虚构的就变得不真实了，而是一种在审美方面应具有的态度，同时也是一种在游戏当中的态度。如果强制性地要求算法在每一次的交互过程当中都毫无掩饰地去展示它自身所具备的机械逻辑，那么不但会把“过渡性空间”的保护作用破坏，还可能致使“人机之间的关系变得疏远”，从而让这一技术丧失它所拥有的慰藉人心的功能。

因此，解决这一问题的关键在于建立一种“负责任的模糊”，而不在于完全彻底地把黑箱揭开，这种“负责任的模糊”既可以保护用户不陷入病理性妄想当中，同时也能够把互动的魔法空间保留下来。这所需要的是一种具备灵活性的边界管理方式，而不是那种较为强硬的技术祛魅手段，这就要求技术设计以及社会教育共同努力。一方面，算法设计者应该避免运用不透明性故意去欺骗用户，不过也不需要把每个内部细节都毫无保留地展示出来；另一方面，用户同样要提升自身的媒介素养，清晰地认识到 AI 不是真实的人类，并且要懂得适度地去进行这种看似真实的互动体验。比如，在界面当中标记“此为 AI 生成回答”的提示信息，就能够在不对用户体验造成干扰的情形下，给予用户理性方面的提醒。应该在透明性以及沉浸性之间找寻一种平衡，而并非偏向其中任何一个极端。比如，可以设计一种机制，使 AI 在提供深度情感支持时，偶尔以幽默或是自嘲的形式对用户的机器身份进行提醒，而不是生硬地把对话流程中断。这样的设计既可以维持互动所具有的流畅性，同时在潜意识层面也能够强化现实感。这种“微调”相比于彻底的“揭穿”更具有建设性。而吴文所担忧的“黑洞”，主要原因是这种边界管理的缺失，以及商业力量将利益最大化作为目标而故意制造出来的“深度伪造”，但这种担忧不应当包括模糊性本身。模糊性本身是中性的，它是艺术以及情感生成的基础，只有当这种模糊性被资本力量恶意运用来操纵用户的时候才会变成“陷阱”。

互补与重构：人机关系的边界协商与社会性延伸

吴文的第三个关键论点聚焦于“粘性与边界的张力”，其担心如果把关系亲密度当作核心目标，

就会引发“社交技能退化”“社会凝聚力降低”以及现实人际纽带的松弛，[35]该论断是基于一种“零和关系”的假定：人机关系的提高必定是以牺牲人际关系为代价的，投入到机器上的情感多，留给真人的情感就越少。然而，这一假定在复杂的社会现实面前显得较为简单，应当重新审视人际关系与人际关系之间的结构性联系——它们有没有可能形成一种互补乃至协同进化的关系？在人机混合的社会里，社交边界是不是正经历一场必要的重构？

（一）非替代性逻辑：人机关系的生态位差异

吴文有着这样的担忧，认为虚拟伴侣会被当作真实伴侣的替代品，导致人们逃避现实中的社交活动，然而这种担忧却忽视了人际关系的本质。现实当中的人际关系具备互惠性、责任性、身体性以及无法被预测的风险性，人在和真人进行交往的时候，不仅是为了获得情感方面的支持，更是涉及社会资本的积累、道德义务的履行、物质的交换以及生命体验的丰富等维度，而人机关系的核心特点是单向度、没有责任、高度可控以及纯粹的情感性。[36]但是，AI 伴侣没办法在现实关系当中开展物质帮助的工作、进行法律承诺方面的工作以及提供身体接触，这使它很难从根本这方面来对真人伴侣进行替代。大部分用户甚是明晰两者之间的界限：他们选用 AI，一般是为了填补真人无法全面覆盖的“剩余时间”或者“剩余情感需求”，[37]如深夜难以入眠的时候去倾诉、对亲友不能启齿的隐秘话题，或者需要极大耐心倾听以进行情绪宣泄。在这些场景当中，AI 是补充品而非替代品，其形成了社会支持系统里的新维度，它帮助那些在现实社交方面遭受挫折或者能力不够的人保持基本的情感代谢，而不是把他们和现实的联系切断。这属于一种“补偿性社会性”——AI 伴侣将人与人交往空白的缝隙予以填补，但并不会把人们对于现实关系的渴望消灭。

相反，很多用户在和 AI 建立起信任以及信心之后，更加敢于走向线下社交。他们把 AI 当作一个过渡的“社交训练”，一旦拥有足够的能力，他们还是期望拥有现实中的伴侣和朋友。可以说，人机关系与人际关系并不是一场“零和博弈”，而是更像生态系统中不同生物之间的利他共生：两者拥有各自的位置以及功能，共同满足人类多层次的情感需求。

（二）重新定义交往理性：社交能力的“退化”还是“演化”

吴文持有这样的观点，即习惯于机器人的顺从状况，致使用户在遭遇真人之间的摩擦时会变得不知所措，这其实是属于一种“用进废退”的生理学比喻。不过，从教育心理学以及技能习得这两个方面来看，人机交互很有可能会变为社交技能的训练场地。对于社交焦虑障碍者、孤独症谱系人群以及长期缺少社交互动的边缘群体来讲，直接卷入到充满不确定性以及高风险的真人社交活动当中，或许会存在困难甚至带来创伤。AI 伴侣拥有一个低风险、高容错的模拟环境。在这里，他们能够反复练习怎么样开启话题、表达情绪、处理对话回合，而不用去担忧被他人嘲笑或者拒绝。[38]这种机制被称作“社交催化剂”效应，如果用户在和 AI 进行交互当中获取到自信，积累了成功的沟通体验（即便是模拟的），所生成的自我效能感往往能够迁移到现实生活当中，并且会对他们与真人的接触起到促进作用。关键之处在于，这种生成性逻辑需要得到合理的引导。因此，应该要求算法设计引入一定的“引导机制”，鼓励用户把话题拓展到线下，或者模拟一些较为温和的社交冲突。如此一来，AI 就不会是社交能力的终结之地，而是通向现实社交的桥梁。

同时，智能伴侣应用应当在界面以及宣传当中明确标识出其虚拟性质，并且在检测到用户过度沉迷的时候给予提醒或者介入，[39]这些举措的目的并非打击用户使用 AI 的热情，而是要保证他

们可以“带着练习赛的成果重返正式赛场”，把借助 AI 提升的沟通能力以及心理韧性运用到现实的人际交往中。若是仅仅是沉溺于虚拟温床，那现实的情感需求自然无法真正得到契合，但若把它当作一种情感辅助工具，去帮助那些暂时没办法融入现实社交的人恢复信心和能力，那么 AI 完全可成为人际交往的“助跑器”。在这方面，一些国家已开始尝试利用对话式 AI 辅助孤独症患者练习社交，并对结果持审慎乐观态度。此外，还需要重新思考“社交能力”的定义，在数字化时代，能够与智能体进行高效沟通、能够理解和管理数字情绪、能够在人与机器的混合网络中导航，这本身就是一种新的核心能力。未来的社会交往不仅仅是人与人，而是“人—机—人”或“人—机—机”，[40]在这种结构中熟练与 AI 建立亲密关系并利用该关系调节自身情绪的人，具有更强的适应性。因此，所谓的“退化”或许只是旧技能在面对新环境时的不适应，而新技能的习得正在悄然发生，不应以传统社会的社交标准来衡量数字社会的交往能力。

（三）走向“混合社会”：超越人类中心主义的规范建构

对于社会凝聚力下降这一令人担心的情况，吴文引用了鲍德里亚的观点，指出虚拟社交乃是“人际关系废墟的占位符”，[41]这属于一种经典的悲观主义论调，认为技术仿真掩盖了真实的缺失，还致使真实本身变得不再关键。然而，若是将目光投向更为广阔的数字社会图景，便会发觉人机协作正在对社会的组织形态进行重构，而非使其瓦解。

其一，围绕人机互动所形成的新型社群正逐渐涌现。在 Reddit、Discord 等平台上，汇聚了大量 Replika、Character.AI 的用户，这些用户会分享自身与 AI 互动的相关经验，还会探讨技术伦理方面的问题，并且彼此之间给予情感上的支持，这种依靠共同的“人机实践”形成的“趣缘群体”，成为社会凝聚力在数字时代的全新形态。在这个群体里，AI 变成了人与人进行连接的中介以及交流的话题，反而激发了新的社会交往活动，没有带来原子化的状况，反而创造出了新的连接节点。用户们因为共同的“人机体验”而产生共鸣，形成一种新的集体认同。

其二，在老龄化社会中，AI 伴侣正在成为维持社会稳定的重要力量。面对日益严峻的养老护理人员短缺问题，AI 伴侣为独居老人提供了必要的情感慰藉和认知刺激，防止他们陷入彻底的社会隔离。这种人机协同的照护网络，实际上是在维系而非破坏社会的底线凝聚力。它是一种技术辅助的社会关怀。最后，必须认识到，未来的社会形态必然是“人—机—人”混合纠缠的结构，社会凝聚力将不再仅仅建立在人与人的血缘或地缘纽带上，还将包含人与技术物的连接。拉图尔（Bruno Latour）的行动者网络理论提醒我们，社会是由人类和非人类共同组成的网络，[42]这种“混合型主体性”的建立，要求我们重新定义“社会性”的边界。[43]拒绝接纳技术物进入社会关系的范畴，不仅是对现实的视而不见，也可能导致失去构建更具韧性社会结构的机会。

其三，AI 正成为行业创新和多方协同不可或缺的补充。人机混合社交对新兴 AI 社交产品设计和使用的适应和配合，表明机器智能已从单纯的工具转变为具有准行动能力的“共生体”，人类不得不与技术一道塑造市场中的现实虚拟产品。这要求科技企业和平台主动承担伦理责任，建立算法安全审查、用户保护和问责机制等，在推出新的社交类 AI 产品时，平台应公开算法运行规则，将多元利益相关者的目的性和价值观在算法设计阶段机制化植入，让技术从生成之初就与社会价值协同。这种嵌入式协商机制和多维互动治理模式通过“人机”互动适配性，既保持人类战略引领和价值判断的权力，又能在既定规则下高效执行和互动，从而维护了人类主体性优先和社会价值，充分利用了技术设计的情绪关怀能力，为人机共存时代构筑了安全可信、开放包容的社交生态。

因此，人机混合社交并非在制造分裂，而是在编织一张更为复杂的社会网络。这张网络既包含强的人际连接，也包含弱但高频的人机连接，两者共同构成现代人应对风险社会的资源池。人工智能伴侣作为新兴的社交参与者，将人与人更广泛地连接在一起，而不是把每个人孤立在真空中。人机混合社会的到来，也许会改变传统的社群构成和凝聚形式，但未必如悲观者所言是“一地废墟”。相反，只要善加引导，人机协作完全可能促进社会的协同进化——技术为社会注入新的活力，社会为技术发展提供价值导向。

结语

综上所述，吴静教授提出的“情感操控‘黑洞’”论是对技术资本主义时代人机交互风险的一次及时且深刻的预警，其提醒我们要时刻警惕算法对人类主体性的侵蚀、技术对社会肌理的撕裂以及资本对情感的剥削。然而，若将这一隐喻绝对化，可能会陷入一种技术决定论的悲观图景，从而忽视人机混合社交中蕴含的丰富的生成性逻辑，以及人类在面对技术冲击时所展现的韧性与智慧。本文尝试论证，人机混合社交并非单纯的单向操控，而是在流动现代性语境之下，个体以及技术系统所达成的一种双向“合意”以及“互构”。

算法的“取悦”运用不单纯是商业陷阱，更是回应现代情感赤字的本体性安全机制。用户作为主动的受众，凭借使用和满足的逻辑，在数字空间当中实现依恋与秩序的重建，这种依恋虽说具有替代性，不过在功能方面是真实的，它为个体在不确定世界里提供确定的情感锚点，防止本体性安全体系出现崩溃。人机交互当中的“幻象”并非单纯的欺骗，而是拥有过渡性客体功能的心理空间。技术间性以及准主体性的实现，让这种互动有了真实的心理价值和意义生产能力，这样一个空间允许用户进行情感演练和自我修复，是现代人应对心理压力的重要缓冲带，模糊性在这里并非敌人，而是必要的保护层；人机关系并非人际关系的终结者，而是互补者以及训练场，它借助填补生态位空白、选用提供社交模拟以及催生新型社群，从而实现数字时代社会凝聚力的重构。未来社会将是一个人机共存的混合生态，需要我们在这个生态中找到新的平衡，而不是去试图退回前技术时代的“纯洁”状态。

面对人机混合社交的兴起，我们所需要的或许并非怀着恐惧心理去填补“黑洞”，也不是凭借对边界的严格划定把人与机器完全分隔开，而是需要一种“负责任的生成主义”的态度。这种态度在算法设计方面，要求算法从单纯对“粘性”的追求转变为对“福祉”的追求，引入更具引导性的互动机制；在社会治理方面，要使治理从单纯的“规制”转变为“素养提升”，帮助公众构建对于 AI 的正确认知以及使用策略；在算法层面，增加针对用户过度沉迷情况的监测以及提醒机制，适时引导用户暂停或者开展线下社交活动；在监管层面，制定行业规范，比如要求虚拟伴侣应用其虚拟身份明确标识出来、保护用户隐私，并且在侦测到异常使用情况时进行干预；在教育层面，加强对用户数字媒介素养以及心理健康教育的培育，帮助用户正确认识 AI 伴侣所拥有的功能以及局限，从而培养出理性的使用习惯。要走出人机交互中的“黑洞”，关键并非将“黑洞”进行消除，而是要对这股引力的生成方式加以理解，并且学会借助这股力量在人机共生的新星系中来确立人类主体性的新坐标。[44]未来理想的人机关系，不应是操控与被操控的二元对立关系，而是理解与被理解、支持与被支持的多元共生关系，这不仅是技术所肩负的使命，也是人类自身进化所面临的课题。

注释：

①⑧㉛㉟㊶ 参见吴静：《人机交互中的情感操控“黑洞”——基于虚拟聊天伴侣的反思》，《探索与争鸣》2025年第8期。

② Dafoe A，“On Technological Determinism: A Typology, Scope Conditions, and a Mechanism,” *Science, Technology, & Human Values*, vol.41, no. 6, 2015.

③ Bauman Z, Haugaard M，“Liquid Modernity and Power: A Dialogue with Zygmunt Bauman,” *Journal of Power*, vol.1, no.2, 2008.

④⑱ 参见吴静：《价值嵌入与价值对齐：人类控制论的幻觉》，《华中科技大学学报》（社会科学版）2024年第5期。

⑤ Holt-Lunstad J, Smith T B, Layton J B，“Social Relationships and Mortality Risk: A Meta-analytic Review,” *PLoS medicine*, vol.7, no.7, 2010.

⑥ Filatov E M，“Development of Students’ Foreign Language Communicative Skills Based on the Character.ai Web Application,” *Tambov University Review*, Series: Humanities, vol.29, no.5, 2024.

⑦ 参见逯达：《AI伴侣：技术的温情与隐忧》，《中国青年报》2025年5月18日第3版。

⑨ Bauman Z，*Liquid Modernity*, Hoboken: John Wiley & Sons, 2013.

⑩⑪ Giddens A, “Modernity and Self-Identity: Self and Society in the Late Modem Age,” *Social Forces*, vol.71, no.1, 1992.

⑫⑲㉞ Turkle S, *Life on the Screen: Idenity in the Age of the Internet*, New York: Simon and Schuster, 1997.

⑬ De Freitas J, Oğuz-Uğuralp Z, Uğuralp A K, et al，“AI Companions Reduce Loneliness,” *Journal of Consumer Research*, 2025.

⑭ Katz E, Blumler J G, Gurevitch M, “Uses and Gratifications Research,” *Public Opinion Quarterly*, vol.37, no.4, 1973.

⑮ 参见吴增定：《“现代哲学的隐秘的憧憬”——胡塞尔的先验现象学与现代启蒙》，《浙江学刊》2025年第3期。

⑯ Jackson T D, Yu B, “Can AI Have a Personality? Prompt Engineering for AI Personality Simulation: A Chatbot Case Study in Gender-Affirming Voice Therapy Training,” *arXiv preprint arXiv*, vol.2508, 2025.

⑰ Ho A, Hancock J, Miner A S，“Psychological, Relational, and Emotional Effects of Self-disclosure after Conversations with A Chatbot,” *Journal of Communication*, vol.68, no.4, 2018.

⑳ 参见张中雷：《数字社交环境中的亲密关系：媒介技术、用户行为与内容生产的交互影响》，《长江师范学院学报》2025年第4期。

㉑ 参见张平、裴晓军：《准社会关系理论在传播研究中的价值、意义与方法》，《东南传播》2014年第7期。

㉒㉝ Horton D, Richard Wohl R, “Mass Communication and Para-social Interaction: Observations on Intimacy at a Distance,” *Psychiatry*, vol.19, no.3, 1956.

㉓㉔ Winnicott D W, *The Collected Works of DW Winnicott*, Oxford: Oxford University Press, 2016.

㉕ 参见晏青、阳书琴：《人机亲密关系的交互过程及机制研究》，《西南交通大学学报》（社会科学版）2025年第4期。

㉖ Ihde D, *Technology and the Lifeworld: From Garden to Earth*, Indiana: Indiana University Press, 1990.

㉗ Coeckelbergh M, “Robot Rights? Towards a Social-relational Justification of Moral Consideration,” *Ethics and Information Technology*, vol.12, no.3, 2010.

㉘ 参见龙松熊：《论企业数据保护的合理性基础》，《法理——法哲学、法学方法论与人工智能》2024年第2期。

㉙ 参见宋美杰、刘云：《交流的探险：人—AI的对话互动与亲密关系发展》，《新闻与写作》2023年第7期。

㉚ Ehsan U, Riedl M O，“Explainability pitfalls: Beyond Dark Patterns in Explainable AI,” *Patterns*, vol.5, no. 6, 2024.

㉜ 参见艾尚乐：《政府数据开放视域下人机交互应用的隐私伦理风险及其规制》，《湖湘论坛》2025年第3期。

㊱ 参见赵云泽、黄子婷：《从情感劳动到情感剥削：人工智能亲密关系中的资本逻辑与主体性危机》，《编辑之友》2025年第10期。

㊲ 参见张启德：《情感何以拜物：数字资本主义下的情感异化及生存突围》，《天府新论》2025年第6期。

㊳ 参见杨钊、仲佳：《青年群体“AI伴侣”走热的生成逻辑、伦理风险和引导方向》，《宁夏社会科学》2025年第3期。

㊴ 陈一帆、周思宇、李欢：《AI伴侣走热，如何看待争议与风险？》，新华网，https://www.news.cn/tech/20241031/ec869bb26273426fa6e0e94b7bf2e2b6/c.html，访问日期：2025年12月15日。

㊵ 参见常晋芳：《智能时代的人-机-人关系——基于马克思主义哲学的思考》，《东南学术》2019年第2期。

㊷ Latour B，“On Actor-network Theory: A Few Clarifications,” *Soziale welt*, vol.47, no.4, 1996.

㊸ Hayles N K，“How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics,” *Public Understanding of Science*, vol.9, no.4, 2000.

㊹ 解学芳：《数智时代文化产业高质量发展范式——基于“技术—文化—制度”模型》，《华中师范大学学报》（人文社会科学版）2025年第5期。

编辑　孙冠豪　张　蕾