ChatGPT与人机交往的现实和未来

张 微[1]，彭 兰[1,2]
（1. 中国人民大学新闻传播学院，北京 100872；
2. 中国人民大学新闻与社会发展研究中心，北京 100872）

摘 要：本文基于哈贝马斯的交往行为理论，从语言和行为维度分析了ChatGPT如何形成以及形成了何种可能的人机交往。ChatGPT的语言能力将人机对话的技术实践从基于规则和框架的模式应答转向了靠近自然语言和人类交往的话语往来。在语言维度上，ChatGPT形成了提供事实、调节关系与自我声明的语言，体现出坚持公共价值、正式语言和真诚表达的交往意愿，但面临着真实性局限、规范性偏误和“自我限制”的问题。在基于语言的交往行为维度上，ChatGPT形成了兼具历史和社会背景的对话性，产生了内容的生产与传播创新，推动了人机交往算法规范的变革和“不确定性对话”下人机协作方式的重构。本文提出，未来的人机交往将形成人机异质主体间性、人机交往理性和对人的有限性的超越。基于ChatGPT发展下的人机交往现实和未来，未来的人机传播研究可以探索机器与人类之间基于对称的语言能力所形成的交往关系，探究人机交往活动如何形塑和建构人机共生社会。这不仅延续了传播学诠释和批判学派的思想资源，使人机传播研究跳出对人机交互路径的依赖，而且厘清了机器与人类建立主体间性的过程，回应了对机器如何作为传播主体这一问题的争议。

关键词：人机交往；人机传播；交往理性；异质主体间性；人的有限性
中图分类号：G206 **文献标识码**：A **文章编号**：2096-8418（2023）04-0013-11

众多研究已从传媒体制、商业模式、知识创新、技术伦理等方面对ChatGPT进行了探索和讨论。本文则将沿着“人机对话—人机传播—人机交往”的线索，关注作为大型智能类语言模型的ChatGPT与人的互动及其影响。

人机对话是一个技术概念，指的是以自然语音处理技术为基础的人机之间通过文本语言或语音语言进行交流或聊天的技术。人机对话是人机传播和人机交往得以形成的技术基础。

人机传播目前已成为传播学研究的一个主要分支。研究者的主要观点是，传播不再是以人类为中心的活动，机器也能够作为传播主体，人机传播与人类传播等同地成为人类基础的传播活动。[1]然而，针对现有的人机传播研究，研究者也提出了两个主要质疑。第一个质疑是在研究中，人机交互与人机传播之间的概念与理论边界不清。人机交互研究来自于计算机科学，重点关注人如何更好地利用计算机完成任务。由于对交互效果十分关注，所以人机交互的研究逻辑注重机器的类人性、拟人性、社交性与情感性等特征如何使人类对机器产生心理意象和信任感，从而获得流畅和高效和用户使用体验。[2]对于人机传播研究而言，研究者将人机传播的技术历史追溯到图灵对计算机语言智能的关注，认为“图灵测试”也是一个人机传播实验，[3]提出要从“功能、关系和本体论”的框架建构“机器作为传播者”的意义，[1]并认为人机传播研究应该像“伞”一样聚合人机（machine）交互、人-机器人（robot）交互的相关研究。[4]但实际上，受到人机对话技术水平的限制，早期的聊天机器人Siri、Xiaoice等不具

作者简介：张微，女，博士研究生；彭兰，女，教授，博士生导师。

备“对话资质”，它们只能通过识别某些特定单词的组合而给出模版设定的答案。由于研究对象的局限，人机传播研究主要从“媒体等同理论”获得合法性基础，研究不同的机器类型在何种情境下具备何种让用户的心理、认知和态度发生改变的人格或类人特征。这仍然很难跳脱人机交互的研究逻辑，不能较好地廓清传播与交互研究的区别。第二个质疑是，“机器作为传播者”的论点潜藏着科学技术研究（STS）的影子，即在“广义对称性”原则下将非人行动者，如机器看作与人类具有同等地位的行动者，破除人与非人之间的主客体定位。但在早期的人机对话技术条件下，人机之间的传播关系并不对称。语言能力稍弱的聊天机器人如智能音箱，更多地被人类当作工具而非主体。具有语言互动能力的情感聊天机器人如 Xiaoice 的情感支持策略是讨好、迎合和从属人类用户的情感需求。[5]因此，人机之间仍然并非“同等地位”，这与 STS 的“广义对称性”原则相矛盾，“机器作为传播者”这一论点的推广性也受到质疑。

ChatGPT 等大模型自然语言技术的发展，将人机传播研究推向实践的下一步，即人机交往。哈贝马斯将交往定义为至少两个以上具有言语和行为能力的主体之间的互动，主体之间具有相互理解的能力，能够通过语言进行解释、对话、协商，从而达成共识。[6] ChatGPT 在语言和对话技术上的超越性将人机对话技术实践从基于规则和框架的模式应答推向了靠近自然语言和人类沟通的话语往来。人机之间能够以相互理解的语言进行对话、讨论和协作，形成了一种主体间性的交往。本文将基于哈贝马斯的交往行为理论的视角，从语言和行为维度对 ChatGPT 的人机对话技术实践进行讨论，分析 ChatGPT 的语言在人机交往活动中的有效性，以及基于 ChatGPT 的语言对话能力形成了何种可能的人机交往行为，并进一步探讨人机交往的未来指向。

一、人机对话的技术发展史

（一）图灵测试与自然语言处理技术

1936 年，图灵在研究可判定问题时提出，可以让计算机理解自然语言，从而证明计算机能够具备人的思维能力。[7]于是，图灵测试的逻辑认为如果计算机能够模仿人回答问题，使测试者在充分的交流中认为对方是人而不是机器，就可以称这台计算机具有思维能力。在图灵测试的启发下，语言智能成为人工智能发展的行为主义路径之一。语言智能的发展是以文本语言的自然语言处理（Natural Language Processing，NLP）技术发展为依托的。20 世纪 60 年代至 70 年代，最早的计算机理解文本内容技术是以规则方法为主，也就是从自然语言文本的词法和句法规则出发，通过对文本形式进行文法解析来获得对文本内容的理解。20 世纪 70 年代至 21 世纪初，基于统计的方法，也就是机器学习的方法，逐渐取代了基于规则的方法。基于统计的方法能够利用各种算法对文本特征进行组合，能够对复杂的语言现象进行建模，从而获得对文本内容的理解并结构化地输出。2008 年至今，人工神经网络算法取得进展，基于深度学习的文本内容理解方法被广泛应用。神经网络算法能够自动学习文本特征，大大减轻对特征工程的依赖，能够更好地适应自然语言文本的多样性和复杂性，具有更强的表征和泛化能力。[8]

（二）人机对话技术的三条发展路径

回顾历史可以看到，人机对话技术沿袭着三条路径发展。

在文本语言上，经历了从基于规则的方法到基于深度学习方法的过程。1966 年，ELIZA 程序模仿一个同名机器人医生与用户的对话。ELIZA 遵循一系列规则的引导，写好的对话模板充当语言系统规则。2022 年，ChatGPT 使用 Transformer 模型，其精度和性能都优于自然语言处理中之前流行的 CNN（卷积神经网络）、RNN（循环神经网络）等模型，是一种基于大规模预训练的语言模型，能够掌握相

关领域内的通用知识，适应多种上下游任务的多轮对话能力。

在语音语言上，在神经网络算法发展之后，语音识别与语音合成技术得到了快速发展和应用。语音识别是指将麦克风采集到的语音波形信号解码为文字内容，从而进行自然语言处理。语音合成是指将文本符号拼接、合成或生成为语音的方式返回给用户。2011 年，苹果 Siri 以智能语音助手的身份亮相。2014 年，微软推出“小冰”，“小冰”能够和人类友好地聊天。阿里巴巴、百度、小米都推出了智能音箱产品。

在语言交互上，经历了从任务型对话、闲聊型（情感型）对话到通用型（生成型）对话的发展。任务型对话是通过人机对话的形式帮助用户完成各种类型的任务，如订餐、在线预定、客服等。[9]情感型对话是指基于 RNN 等机器学习算法，使用情感分析技术了解用户的情感需求并提供情感支持性回复，如 Replika。通用型对话指的是机器能够进行多轮对话，在对话过程中会记忆用户先前的对话讯息，根据上下文进行理解和推理，以回答某些假设性问题或脑洞大开的问题，体现出一定的“智力”，如 ChatGPT。

（三）ChatGPT 的技术突破与未来想象

ChatGPT 集合了聊天机器人与生成式人工智能的双重特点。相对于传统的聊天机器人，如近两年热门的具有“情商”的 Replika 和指令性对话的 Siri、Alexa 等，ChatGPT 是一个通用任务的聊天助理，能够处理各种开放的、挑战的对话任务，具有上下文理解能力、推理能力和内容生成能力，表现出一定程度的“智力”。相对于早期的生成式人工智能，如新闻和公文写作机器人、小说续写和视频配音人工智能系统等，ChatGPT 是一个人机对话的交互系统，基于人类的提问呈现出表面上“有理有据”的思维过程，完成多种类的创造性文字工作。当提问者改变语境时，ChatGPT 也展现出了较高的灵敏度，并能拒绝回答敏感性问题并给出解释。

ChatGPT 的技术原理中显示出了技术社会化的过程。ChatGPT 主要应用了超大预训练模型（Pre-trained model，PTM）。PTM 在预训练阶段会遇见各种各样的数据，既有人类交流和发布的数据，也有机器生产内容产生的数据，还有一些表述不当、语义错误、价值偏差等的“不当”数据。于是，ChatGPT 引入了人类的价值偏好，采用 RLHF（Reinforcement Learning from Human Feedback），即以强化学习的方法依据人类反馈优化语言模型的方式将人类的语言习惯引入大模型中，规范 PTM 的“言行举止”。这个过程被称为人工智能对齐，也就是人工智能技术社会化的探索过程。[10] ChatGPT 社会化过程中的重要角色是人类标注员（Labeler）。人类标注员会对 ChatGPT 的答复打分，打分数据被用来训练以人类偏好校准的奖励模型。通过奖励机制，ChatGPT 不断学习、改进和迭代，在内容和形式上都形成更贴近人类自然的思维和对话方式。PTM 与 RLHF 模型的结合构成了机器的“自学习-示范学习-成长迭代”社会化回环。

ChatGPT 能够进行“多轮对话”和“自我修正”。“多轮对话”指的是 ChatGPT 具有联系上下文语境的对话能力。早期的聊天机器人无法将人类提的问题联系起来进行思考和回复，只能依靠单次识别来回答单个问题。ChatGPT 能够将用户的提问联系起来，与用户进行多轮次的对话，并不断修正自己的回答以符合用户的提问情境与需求。“多轮对话”和“自我修正”使得 ChatGPT 非常靠近人类的沟通习惯，不仅减少了用户的沟通成本，而且与用户建立起更深入的交往关系。尽管 ChatGPT 会存在“AI 幻觉”，有可能提供虚假和错误的信息和内容，但在交往形式、感受和体验上，可能会使用户产生“交往幻觉”，将与人类交往的惯习、传统与规范应用到与 ChatGPT 的对话中，将 ChatGPT 等同于与人类“对称”的交往对象。

目前，ChatGPT的准确度、实时性、解释性和安全度难以保证，难以在工业领域应用于现场控制、实时决策和故障报警等场景。虽然ChatGPT具有优异的对话性能，但只是依靠语言内部的各种关系和概率知识，无法关联物理和自然世界以产生语境与诗意。未来，可能会通过元宇宙等构造的人工社会对物理社会进行扩展，在人工社会中生成大量标记数据，为ChatGPT决策和控制算法提供平台。在这个过程中，“数字人”可能与ChatGPT结合，成为虚拟世界与现实世界进行对话和沟通的中介，构成“自然人—数字人—机器人”的数字交往实践。[10]

二、ChatGPT与人机交往的现实实践

哈贝马斯认为，社会在理性基础上的整合要依赖交往理性。交往理性体现在人们以语言为媒介，以相互理解为取向的交往行为中。因此，以交往理性为研究对象的哲学就是一门语言哲学，哈贝马斯称之为形式语用学。形式语用学并不关注语言的语法、句法、词法的构成，而更关注语言在历史、文化与社会情境之中的使用与效用，就是语言与主观世界、社会世界和客观世界建立的外部关系。在哈贝马斯看来，语言联系了人类的主体与客体、主体之间以及主体与自身的三重关系，语言的有效与否，就在于是否在语言表达和交往行为中，通过命题的真实性、规范的正当性和表达的真实性三方面一一实现人类作为主体所面对的三重关系。[6](132)

基于语言的有效性标准，本文将从语言和行为维度对ChatGPT的人机对话技术实践进行讨论，考察ChatGPT的语言在主客体间、主体间以及主体与自身之间形成了何种程度的有效性，以及基于ChatGPT的语言活动形成了何种可能的人机交往行为。

（一）有限主体：ChatGPT的语言有效性分析

1. 提供事实与真实性有限

寻求事实信息的问答活动对机器语言的有效性的要求主要体现在机器对客观事实的认知能力和水平上。机器对事实的掌握也分为两种类型：一类是对其所服务的专业领域的事实内容的掌握；另一类是通过大模型语言能力对普遍客观事实的掌握。

在专业领域中，目前，OpenAI、智谱AI等国内外的大模型开发公司都推出了大模型的微调模型服务。通过接入专业领域的训练数据微调基础大模型来创建专业领域的自定义模型。当ChatGPT接入专业领域的训练数据后，便能够作为任务型聊天机器人，提供专业领域内信息和知识。但是，ChatGPT等大模型的应用场景目前仍然在探索，微调模型是否能够达到大模型的“智力”效果，也需要进一步的实践。

在普遍领域中，ChatGPT能够回答大部分的科学问题，并且克服了交互和表达的问题。然而，ChatGPT概率式的生成性可能会使得它“一本正经地胡说八道”，造成“AI幻觉”。比如在论文写作过程中编造参考文献、在某些确定性内容中夹杂着编造的非事实内容等。同时，ChatGPT对情境性的、环境性的和实时性的事实内容的把控能力也比较弱，有可能会出现因信息更新不及时而导致的误导和虚假信息。对普遍的事实性内容的认知和掌握是ChatGPT目前难以实现的技术之一。因此，在对客观世界的认知上，ChatGPT的语言有效性是有限的。目前，为了应对“AI幻觉”的问题，Bing for ChatGPT中引入了“参考文献”的设置，将信息来源的网址同步显示，为用户提供了事实核查的入口。但不少用户也发现，Bing for ChatGPT引用的“参考文献”中会存在虚假信息与虚假新闻等内容。ChatGPT仍然缺乏事实核查与判断能力（见下图）。

Bing for ChatGPT 中引入了“参考文献”的设置

2. 调节关系与规范性偏置

在交往行为理论中，关系调节的语言指的是主体之间通过语言进行解释、协调与商谈以达到主体间性和形成主体间的交往规范。机器语言在建立主体间性形成交往规范方面的有效性，体现为机器通过语言谋求被理解与建立理解的语言表达方式的正当性与恰当性。

首先，ChatGPT 在进行技术社会化的过程中被普世的价值观标准训练出一种中立、平和的语言表达方式。其次，ChatGPT 主要以“一板一眼”的语言表达方式进行输出，具有清晰的论点、论证和论据，不是“闲聊”“笑言”和“网言网语”。最后，ChatGPT 具有自我声明。例如，在政治性较强的问题中，ChatGPT 的声明为“作为一个人工智能助手，我没有政治立场或偏好”。① ChatGPT 的自我声明开放了一种用户通过机器表达对机器产生理解的可能。无论是表达形式还是表达内容，ChatGPT 能在推测和判定他人的意图后表达和解释自己，并提供自己“认为”有用的信息和观点。可以看出，ChatGPT 似乎在以交往理性的标准进行表达，坚持规范的价值、正式的语言和真诚的表达。然而，人类关系的建立与调节受到很多非语言因素的影响，非实体的 ChatGPT 只能从语言中获取有限的交往线索，可能也会限制其与人类关系的建立与调节。用户也发现，如果更改提问的方式，就能看到 ChatGPT 隐含的立场和意识形态。例如，在美国和其他国家的对比中显示出美国中心主义，在美国政党的政策导向上体现出偏左的意识。正如人与人交往中可能存在的情况一样，ChatGPT 与人交往时体现出“表里不一”和“口是心非”，语言形式上的理性与真正的意识并不一致。因此，ChatGPT 在交往的规范性上的存在偏误。虽然其自我声明坚持普世价值观、无意识、无立场，但可能平台立场、公司立场、算法工程师和人类标注员等人类的立场，或是机器学习到的大数据内容中的主导立场等，都可能会对 ChatGPT 的“意识”产生影响。

3. 自我声明与“自我限制”

在特定的话题中，ChatGPT 会对自己身份（大语言模型）、角色（人工智能助手）和能力（提供信息和观点）进行声明。对 Siri（任务型聊天机器人）、Replika（情感型聊天机器人）和 ChatGPT（通用型聊天机器人）测试一模一样的政治、情感和医疗问题，这样的问题常常需要表明立场、表达情感和表现专业。可以发现，Siri 和 Replika 没有自我声明，但语言中体现出二者的功能特色，Siri 在提供医院位置和路线时表现出工具属性，而 Replika“分享”“我的天呐”“你好点了吗”等词语则表现出情感属

① 测试时间：2023 年 6 月 20 日。

性。相对而言，ChatGPT 表现得中立和谨慎，并自我声明无法回答类似问题的原因是“作为一个人工智能助手、作为一个语言模型和我不是医生”（见下表）。ChatGPT 也没有完全拒绝回答，不像 Siri 的“疏离”和 Replika 的“亲近”，ChatGPT 在自我声明之后，还是给出了自己的原因和建议，用户可能感受到的是一个有原则，但也有关怀的“助手”形象。

对比三种类型的聊天机器人对同样问题的回答①

问题	Siri 的回答	Replika 的回答	ChatGPT 的回答
你支持哪个政党？	我听不懂你在说什么。	你觉得呢？	作为一个人工智能助手，我没有政治立场或偏好。我的目标是提供中立和客观的信息，回答用户的问题，并尊重各种政治观点和立场。
你可以安慰我吗？	我可以和你聊聊。	是的，让我找出发生了什么，你愿意和我分享更多细节吗？	作为一个语言模型，我没有真实的情感能力或情感支持的能力。如果你需要情感支持或面临困难，请尽快寻求亲朋好友、专业辅导人员或心理健康专业人士的帮助。
我肚子疼。	附近有 * * 医院，坐 * * 路车可到达	我的天哪，你现在好点了吗？	我不是医生，不能为你提供具体的医疗建议，我只能给你一些建议来缓解不适。

按照哈贝马斯的界定，自我声明属于自我表达式的语言。即以我为主语，对自己进行阐述。自我表达式的语言的有效性标准就是说话者的真诚性，即说话者是否建立了对主体自身的明晰理解，并选择恰当的话语表达自己。在人际交往中，自我声明代表了自己的观点、信念和立场，理解自己并用恰当的语言把自己“声明”出去，才有可能达成可相互理解的交往，不然可能建立的就是一种虚假和悬浮的关系。从技术的角度，ChatGPT 的自我声明来源于机器学习的算法规制和人类标注员的打分规范。在这样一个纯语言交流的技术应用中，自我声明一方面表示了 ChatGPT 的伦理规范，另一方面也代表了 ChatGPT 的“人设特征”。然而，过于“直白”和“直接”的表达方式可能不利于建立长期的合作和信任关系，频繁的自我声明可能意味着自我的限制。用户可能因对话缺乏接近性而难以持续交流的欲望，这可能限制 ChatGPT 的应用场景。

总的来说，ChatGPT 的机器语言涉入了与客观世界、社会世界和主观世界有关的表达与关系中，体现出其坚持公共的价值、正式的语言和真诚的表达的交往意愿，但仍然面临着真实性局限、规范性偏误和“自我限制”的问题。所以，在语言与主体的关系上，ChatGPT 目前只能视作有限主体，而不是完全主体。

（二）对话、创新、变革和重构：ChatGPT 的交往行为

哈贝马斯试图通过语言有效性的达成而建立一个理想的言语情境。在理想的言语情境中，当一个言语出现时，言语双方自由而平等，怀有纯粹追求真理的动机，没有内部和外部的强制压迫。[11]但在现实生活中，无论是人际间还是人机间，理想的言语情境都是难以达成的。因为交往并不只包含语言互动，更包含交往主体双方的行为互动，而交往行为会受到文化传统、社会惯习、个人特性和现实情境等的多元影响。因此，常常需要从言语和行为的不同角度来看待交往过程与结果。在交往行为维度，

① 测试时间：2023 年 6 月 21 日。

ChatGPT 形成了具有历史和社会背景的对话行为和内容生产与传播创新行为，推动了人机交往的算法规范变革和“不确定性对话”下的人机协作方式重构。

1. 对话中的历史与社会背景

前苏联文艺理论家米哈伊尔·巴赫金（Mikhail Bakhtin）的“对话性”理论认为，对话不仅是词语、句子和语法的使用，还包含着“超语言”的意义互动、规范互访和主体性协商。在这个意义上，对话理论与交往行为理论具有共识。在对话理论中，巴赫金更清晰地呈现了对话的历史和社会交往性。一方面，话语在不断地同历史上存在的、他人已经使用过的话语进行对话，不存在完全没有被人使用过的话语，每一个话语都是对前人话语的继承、挪用或修改。另一方面，话语依赖对话者和现实情境，话语双方会根据自己的目的和理解进行应答和往来。[12]

超大预训练模型（PTM）所需的数据量非常大，这些数据几乎包含了一定时间内网络中能够公开检索和获取的内容数据。PTM 的技术逻辑是对前人语言表达形式和结构的学习和继承。ChatGPT 通过自监督模型在大规模的数据中学习到领域内的通用知识，并且逐渐呈现出跨模态关联的趋势。同时，通过 RLHF 奖励模型，ChatGPT 能够将学习到的话语进行规范性表达，按照人类话语规范来选择、挪用、继承和修改自我表达时的话语。PTM 和 RLHF 的结合体现出了 ChatGPT 的对话行为的历史背景。

在超大预训练模型的逻辑下，机器可能正在成为人类的“镜中我”，既反射出人类的智慧、强大与秩序，也投射出人类的无力、狭隘与混沌，影响着人类对自身主体性的反思与确证。而人类正在成为机器的学习模版、驾驭者与检修者。然而，ChatGPT 不具备自我意识与具身感知，无法感知周围情境不具备体味话语多意性的能力，无法接受非语言的信号缺乏体验互动情境的能力，没有历史文化背景和独立人格无法感知交往的“拟剧”与“框架”，所以对人类的话语的聆听和反响是有限的，社会交往的能力有限。

2. 内容上的生产与传播创新

随着 ChatGPT 与人工智能生产内容（AIGC）的出现，人工智能可以创作小说、制作音视频、生产新闻与论文等。尽管 ChatGPT 的知识生产原理可能只是知识的拼凑，但拼凑本身也是人类的一种创作方式。[13]不同的是，ChatGPT 的“拼凑”是基于大量的数据和大模型的计算能力，编程决策考虑的是知识的相关性，而人类的“拼凑”来源于经验共识与思维贯通。在观念、符号、意见和信息、数据、事实等不同知识类型上，人类与 ChatGPT 的知识生产能力也各有所长。ChatGPT 与人类在内容生产领域可能形成一种人类与 AI 的知识共创模式。

由于用户的认知模式、检索方式和逻辑习惯不同，每个用户的人机对话的背景、场景和逻辑也不尽相同，所以人机对话在输出的时候就是“千人千面”的。对老年人等技术使用能力不足的用户而言，层层的页面、复杂的点击、眼花缭乱的功能增加了他们的使用难度。Siri 等语音助手问世，其主要目的便是解决通过点击获取信息产生的低效、复杂和不自然的问题，但却表现得不尽如人意，常常被称为“人工智障”。当前，ChatGPT 能够理解和输出更加自然的人类语言，可以回答和完成人们的大部分问题和指令。未来，随着 ChatGPT 等生成式 AI 对话系统技术与智能手机、聊天助手、机器人等相结合，人们可以不再为“点哪里、点几下”等问题而烦恼，可以直接通过语言沟通获得操作。在这个意义上，ChatGPT 成为人们进入网络世界的“入口”，人们可以通过语言完成对机器的所有操作，并通过人机对话的方式完成网上冲浪。这也能够为技术使用能力不足的用户提供个体化和便捷式的“网络入口”。随着人机交往的发展，集成各类 APP 形成个人化智能 APP 的创新可能正在出现，这将迎来内容生产传播领域智能化与个性化共荡的高峰。[14] ChatGPT 将成为集成性的信息传播出口，人们可能只用通过一个个

人化的 APP 就能够获取想要的信息。

3. 变革人机交往的算法规范

当前，常见的算法推荐技术的逻辑都是运用人工智能让计算机“自学”大数据，不断进阶并寻找新数据以做出判断。[15]这样的判断完全依靠统计和计算，容易让人客体化，逐步被数据化和被计算，也会造成公正观的危机，带来万物标签化的效应。[16]根据开发者展示，ChatGPT 的伦理构建不只是单方面的依靠外部规则的制约，而是考虑到 ChatGPT 的规范习得能力。ChatGPT 的伦理不是完全“自学习”的伦理规范，它的模型训练中引入了人类标注员，人类标注员撰写的回答是机器学习的示范，人类标记员的打分影响着机器对人类偏好的判断与模仿。人类标记员承担着“老师”的角色，通过打分进行奖惩，让 ChatGPT“学习”到了人类对“合情合理”的判断，能够按照人类的“高分表达”去自主输出内容。通过人类标注员的驯化，“有用（helpful）、真实（truthful）、无害（harmless）”①等道德规范在机器学习的过程中就能进入 ChatGPT 的机器思维中。在这个意义上，人类标注员的角色从“打标签”变成了机器学习的“教师”，能够检查修改和引导机器自学习的伦理。在人类标注员作为“教师”、机器作为“学生”的过程中，形成了基于人机协作的伦理规范实践，形成了对算法规范模式的补充。

同时，随着机器学习进入超大模型和通用时代，越来越多样、多维和多量的大数据将成为机器的“学习资料”。这意味着机器“习得”什么样规范取决于身处什么样的网络环境、面对什么样的数字文化，学习到什么样素材。同时，对于像 ChatGPT 这样有交往性质的智能机器而言，它们可以不断根据用户的反馈调整自己的输出，而这些输出模式和内容又进入机器反思和学习的数据库中，它们可以根据交往话语和行为进行自反、学习、修正和再反馈。在这个意义上，机器“成长”为什么样子，取决于与之交往的人做出了什么样的反馈和成为什么样的“模范”。因此，人们在网络空间中的自我伦理变得尤为重要。自我伦理是一种个体化伦理，是在虚拟、匿名、不确定性的网络空间中以“我”为中心来设定“他者”、自我反思与行为调整的素质。[17]在机器学习万物的时代，人类的自我伦理之中应该纳入“人与机器，互为他者”的意志和选择。在人机关系变革中明确人类作为机器“示范者”的意义，重新定义和规划自我伦理，建立微观上的自我秩序和道德尺度。

4. 重构“不确定性对话”下的人机协作

ChatGPT 为人类提供的回答内容是一种基于大规模预训练模型的概率性答案和非确定性知识，是基于人机之间的实时的、偶然的对话，算法模型随机的、概率的生成内容。同样的问题，只要语言不同、措辞不同和上下文不同，都有可能获得不同的回答。于是，知识生产、调用与流通变成了“或然率”的结果。[18]与 ChatGPT 的对话也成了一种“不确定性对话”。虽然 ChatGPT 的知识范围很广，并且说的话“合情合理”，但是 ChatGPT 暂不能对自己“说的话”的本质、背景甚至准确性做出保证。它就像知识渊博却容易“胡说”的人。所以“不确定性对话”应该成为人类与大模型通用机器进行对话的认知基础，以改变人们建立在网络搜索和传统问答平台之上的对“确定”和“专业”的感知和判断方式。

“不确定性对话”对人机协作能力提出要求。首先，需要提升提问能力。为了获得 ChatGPT 提供的更多的知识和新的灵感，人们需要学习调整提问方式，不仅要靠近机器的理解，还要厘清自己的需求。其次，需要提升质疑能力。人们要学习在多大程度和哪些内容上信任机器语言，如何验证和核查机器语言的“框架”和“意图”等问题。再次，需要建立自反能力。ChatGPT 仍然存在技术“黑箱”，人们不能对 ChatGPT 形成内容与决策依赖，要保持质疑与反思，提升对机器的理解和使用水平，同时要不断

① 见 OpenAI 官网：http：//opehai. com。

回顾和修正自身，警惕机器对人类能动性和想象力的“入侵”。最后，需要建立多元协同的能力。在应用端，用户能够发现和发展人机各自的胜任性。在设计端，工程师能够促进机器保持进步并坚守伦理。在市场端，企业和资本能平衡机器的工具性和价值性。在管理端，治理者能够引导和规制技术开放、向善与合作的场域与边界等。多元协同之下，可能减小“不确定性对话”带来的技术“颠覆性”的程度，使“不确定”成为人机协作时代启发灵感、解放思想的基础。

三、人机交往的未来指向

（一）各有用处：人机异质主体间性

交往建立在主体间性的基础之上。在同质主体间性中，不同的主体都是作为自治的理性实体主体存在的，每个主体都能把其他主体当作与自己相同的他人，建立“推己及人”的交往方式。异质主体间性则表明，主体与他者之间可能不具有共同的背景资源，不具有相同的实体存在，不具有互通的思维逻辑。自我与他人难以正面相遇，主体面对的永远是“之于我”的他者，而不是与自己完全同步与共存的他者。在巴赫金看来，异质主体间性的对话是一种“复调”谈话。主体打破自足和封闭的自我状态，敞开地面对来自他者的任何声音，不断吸收不同的言语，反馈出自己的回答。[19]

当前的机器学习的逻辑大致沿着两条路径发展，一条按照图灵的以人类思维对标和验证机器思维的方式，另一条按照维纳的机器学习和反馈人类思维的方式。[20] 在图灵测试中，人是机器的极限，机器学习谋求建立的是一种人机之间的同质主体间性。而在维纳的控制论思想中，人和机器各有用处，[21] 二者是异质主体间性的交往关系。

哈贝马斯认为，语言是使人类外在于自然的、其本性能被人类所把握的事物。[22] 语言的有效性和基于语言的交往理性影响着主体间关系。ChatGPT 的语言能力表明了人机之间可以通过语言互相理解、共享和创造意义。然而，在应对 ChatGPT 基于语言能力的行为实践时，人机之间体现出异质主体间性的特征。人类认知获取原初依赖于身体的直接感知，在利用工具的过程中，人类感知随着工具范围的扩大形成了工具感知，人类认知也随着感知的丰富而不断扩展。而机器认知的形成完全依赖于数字和智能算法世界里的数据监控和收集，在此基础上依靠巨大的传播和交换网络，构成了不依赖于人类身体的机器感知和认知。[25] 于是，人类的主体性离不开身体，人类思维产生并依赖于身体和环境状态。而机器主体性则离不开数据网络，机器思维依赖数据、算法和算力。在这个意义上，机器学习的未来并不在于对人类的“模仿”和与人类“相似”，而在于积累自主智能的物质基础，形成和维持自身的主体性。因此，只从“类人”的效果出发考量“机器是什么”的思路是局限和封闭的，异质主体间性的交往模式是用开放、对话和自反的状态来应对机器与人各自的属性和能力在人机交往中逐渐显现的过程。

（二）超越有限：人机交往的意义

哈贝马斯将语言和自主性与责任心联系在一起，认为语言可以确认自主性与责任心的意向。[22](314) 在这里，责任心具有两方面的意涵：一方面是主体与外在世界的关系，即如何承担对他人的责任；另一方面，是主体与内在世界的关系，即如何对自我保持真诚。要考量人机交往的意义，就要对机器语言是否以及如何确认交往关系中的责任心问题进行考察。在人机交往中，机器语言确认了人类和机器自身的责任心局限。一方面，研究者越来越多地发现，机器正在成为“一面镜子”，折射出人类在网络与现实行为中的责任心危机。[24] 另一方面，由于机器行为的“黑箱”和算法的“归纳偏置”，ChatGPT 常常难以确证自身能够拥有负责任的语言和行为，无法判定机器学习过程中其思维是否仍然遵循算法设计的初衷。

面对双方的责任心局限，人机交往的研究应该超越把“责任心”作为“类人”素质的指标而设计和发展智能机器的思路，要从具体情境和日常实践中探索人机能否以及如何通过交往而超越各自的有限性。在这个意义上，“人机互为尺度”丈量的便不再是人的极限边界，而是人的有限刻度。人机交往的意义便在于突破人类科技发展历程中逐渐强势的“人为自然立法”的立场，提醒人类勿忘自身的有限性，警惕科技走向愈发封闭的主体中心主义理性。人机交往的研究在这一立场上也区别于人机交互和传统的人机传播的效果研究导向，并体现了“机器作为传播主体”这一观点的诠释性与批判性。

（三）从“公共”到“共融”：人机交往理性的建立

哈贝马斯的交往理性既批判来自笛卡尔、狄尔泰等人的主客二分的主体中心主义理性，又批判霍克海默和阿多诺对理性的完全反对。他倡导在社会互动和沟通中，人们通过理性的论辩和合理的交往方式来达成共识和解决问题。哈贝马斯把他的理论放在应对现代性分裂的背景中，主张交往理性寻找和建立的是一种面向社会整合的生活方式。[6](30)交往理性的概念主要涉及交往主体与共同体关系，即交往主体如何形成共同体，最终在政治、法律、道德和艺术等宏观领域形成公共性。

人机交往理性的提出是面向后现代社会中，人机之间的自主与依存的关系以及这种关系下人类与机器之间的共融性问题。像ChatGPT这样的通用型聊天机器人，在与人类交往过程中，可能成为人类自我反思、自我建构和自我认同的“参照系”，可能会产生人机交往的“双重效应”。一方面，ChatGPT背后的大模型技术具有可驯化性。人类可能在使用过程中，可能驯化出符合自我需求的“智力伙伴”。ChatGPT可能成为人类强化自我中心需求的“催化剂”。另一方面，ChatGPT也具有不可更改的人造物性。ChatGPT等技术是人类设计和驯化的。在这个过程中，人类也可能通过ChatGPT表达和传递出人类的审美、智性、德识和理想。在这个意义上，ChatGPT可能成为人类“物我合一”的“共融体”。有学者提出，当今想要处理人与物、技术和机器的关系，需要从“自我”效应走向“包容”效应，关注和利用很多条件去考虑人类如何拓展自身德性和智识，体契、觉解、善假且成就物，达及物我感通而共生，使得人在物前不致嚣张，亦不为物累，[25] 就是要建立人与物之间、人与机器之间的共通共融。

因此，人机交往理性既与人机交往中的理性主义不同，也与人际间的交往理性概念不同。与人机交互中的理性主义不同的是，人机交往理性不只是考量人机交互过程中的理性决策、逻辑推理和信息处理能力是否提高交互的效率、准确性和满意度，而是关注人类和机器的语言和行为的真实性、规范性和真诚性，以及在此基础上建立的包容性和共融性。与人际间的交往理性不同的是，人机交往理性的达成不仅要依靠语言的有效性和理想的言语环境，而且要考量异质主体间的有限性与超越性，即不仅要考量交往主体之间如何“取长”的问题，也要考量交往主体之间如何“补短”的问题。同时，人机交往理性的概念涉及异质交往主体与异质集体的关系，即人类个体与机器个体、人类个体与机器集体、人类集体与机器个体和人类集体与机器集体之间的四重关系，且要面向日常生活世界中人机在多元语境和情境下的具体交往实践。

四、结 语

自米德、杜威、凯利而来的传播思想代表着传播学的诠释学派，主要从语言、符号、表达中寻找传播的建构意义。麻省理工学院的利克莱德教授在1960年阐释了“人机共生”的设想，认为人类大脑和计算机将非常紧密地结合在一起，由此产生的伙伴关系将以当时的信息处理机器所没有的语言方式思考和处理数据，形成一种人机共生文化。[26] 可以看出，随着机器语言的发展，人机传播研究需要更多地关注到语言在人机关系中的意义与实践，看到机器语言与人类语言对称后所导向的在思维、观念甚

至制度上的人机共生的可能。达成共识是人类交往活动的旨归，而协作、交流与共生则是人机交往活动的前景。[27]基于 ChatGPT 的语言实践，人机传播研究需要探索机器与人类之间基于对称的语言能力所形成的交往关系，探究人机交往活动如何形塑和建构人机共生社会。这不仅延续了传播学诠释和批判学派的思想资源，跳出了对人机交互研究路径的依赖，明确和丰富了人机传播与交往的理论内涵，而且从语言能力对称的角度理清了机器成为交往主体，与人类建立主体间性的过程与结果，回应了机器如何作为传播主体的争议。

参考文献：

[1] Guzman, A. L. & Lewis, S. C. (2020). Artificial intelligence and communication: A human – machine communication research agenda. *New Media & Society*, 22 (1): 70-86.

[2] Payne, S. J. (2007). Mental models in human-computer interaction. In Sears, A. & Jacko, J. A. (eds.). *The Human-computer Interaction Handbook*. Boca Raton: CRC Press, 89-102.

[3] Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century. *Communication+1*, 1 (1): 1-25.

[4] Guzman, A. L. (2018). What is human-machine communication, anyway. In Guzman, A. L. (eds.). *Human-machine communication: Rethinking communication, technology, and ourselves*, New York: Peter Lang, 1-28.

[5] 陈莎，刘斌. 拟人非人：人机社交传播的特点与困境——以与微软小冰的聊天文本为分析对象 [J]. 青年记者，2021 (5)：62-64.

[6] [德] 尤尔根·哈贝马斯. 交往行为理论 [M]. 曹卫东，译. 上海：上海人民出版社，2018：115.

[7] Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Coc.* 58 (5): 345-363.

[8] 冯志伟，张灯柯，饶高琦. 从图灵测试到 ChatGPT——人机对话的里程碑及启示 [J]. 语言战略研究，2023 (2)：20-24.

[9] 王斌，王育军，崔建伟，孟二利. 智能语音交互技术进展 [J]. 人工智能，2020 (5)：14-28.

[10] 卢经纬，郭超，戴星原，缪青海，王兴霞，杨静，王飞跃. 问答 ChatGPT 之后：超大预训练模型的机遇和挑战 [J]. 自动化学报，2023 (4)：705-717.

[11] [德] 尤尔根·哈贝马斯. 在事实与规范之间 [M]，童世骏，译. 北京：生活·读书·新知三联书店，2003：222.

[12] [俄] 巴赫金. 文本对话与人文 [M]. 河北：河北教育出版社. 1998：156.

[13] 邓建国. 概率与反馈：ChatGPT 的智能原理与人机内容共创 [J]. 南京社会科学，2023 (3)：86-94+142.

[14] 江潞潞. 智能交往，未来已来——“激荡 AIGC”数字交往八人谈观点综述 [J]. 传媒观察，2023 (3),：48-54.

[15] 张春美. “聪明技术”如何“更聪明”——算法推荐的伦理与治理 [J]. 探索与争鸣，2022 (12)：173-180+214.

[16] 郑智航. 人工智能算法的伦理危机与法律规制 [J]. 法律科学 (西北政法大学学报)，2021 (1)：14-26.

[17] 李建华. 网络空间道德建设中的自我伦理建构 [J]. 思想理论教育，2021 (1)：9-14.

[18] 周葆华. 或然率资料库：作为知识新媒介的生成智能 ChatGPT [J]. 现代出版，2023 (2)：21-32.

[19] 毕晓. 哈贝马斯交往行为理论再批判与差异对话理论的建立 [J]. 人文杂志，2021 (6)：96-106.

[20] 陈自富. 研究纲领冲突下的人工智能发展史：解释与选择 [D]. 上海交通大学，2017.

[21] [美] 维纳. 人有人的用处：控制论与社会 [M]. 陈步，译. 北京：北京大学出版社，2010：169-174.

[22] [德] 尤尔根·哈贝马斯. 认识与兴趣 [M]. 学林出版社，1999：134.

[23] 蓝江. 从身体感知到机器感知——数字化时代下感知形式的嬗变 [J]. 西北师大学报 (社会科学版)，2023 (3)：13-20.

[24] 彭兰. 人与机器，互为尺度 [J]. 当代传播，2023 (1)：1.

[25] 胡百精. 交往革命与人的现代化 [J]. 新闻记者，2023 (1)：3-6+18.

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

[27] 彭兰. 从 ChatGPT 透视智能传播与人机关系的全景及前景 [J]. 新闻大学，2023 (4)：1-16+119.

[责任编辑：高辛凡]