# 差异与弥合:人机交流的社会认知进路探究

孙瑞璇,谭　笑
(首都师范大学 政法学院,北京 100089)

**摘　要**:ChatGPT 与人类用户的交流依赖于其内部语言模型,从具体的语义处理机制来看,人机交流表现出一种类"预测加工"的过程,在这种观点下,人机交流中的相互认知是一种双向预测的耦合。但人人交流中的社会认知并非仅仅依赖于内部机制运行,还依赖于人与人之间的交互作用,而交互作用具有独立性,不能还原到个体机制的解释框架之中,通过交互产生的认知结果具有具身性、关系性和涌现性。人人交流是一种参与式的意义构建过程,但是 ChatGPT 与人的交流并非具身,而是完全依赖内部模型机制的运行,交流过程仅仅是对人类赋予意义的单向识别,提供给用户的只是一种泛情景化、知识化和经验化的回复。人机交流想要走向社会认知,并非仅改进机器以单方面贴近人人交流的模式,而是人与机器之间的相互弥合,同时机器广泛参与到与人类的交流过程中必然会带来社会认知的内涵与边界的改变,因此应当在人类与机器的动态变化关系中为人机交流寻找出路。

**关键词**:ChatGPT; 人机交流; 预测加工; 社会认知;参与式意义建构

**中图分类号**:N031　　**文献标识码**:A　　**文章编号**:1008-7699(2024)01-0017-09

随着 GPT(Generative Pre-trained Transformer)3.5 架构的问世,作为 AIGC(AI-Generated Content)技术的最新成果 ChatGPT 在一定意义上将人类带入了智能创作的时代。尤其是 ChatGPT 在语言交流方面功能强大,相较于之前的 AI,ChatGPT 在对话方面更为逻辑与流畅,能够作答的领域更加广泛,很多人认为 ChatGPT 已经获得了语言能力与思维能力。例如,仿生机器人 Ameca 接入 ChatGPT 后,面部表情、肢体动作都十分生动,能够实现与人类实时的,极为生动、真实的交流,机器仿佛有了生命,像人一样与使用者进行交流。

越来越多的学者开始关注 ChatGPT 强大的对话功能,并对人类与 ChatGPT 在交流模式方面的不同进行了探讨。关于 ChatGPT 将对人类产生的影响,语言学家乔姆斯基(Noam Chomsky)认为,人工智能和人类在思考方式、学习语言与生成解能力等方面有着极大的差异,语言的正确解释是非常复杂的,无法仅仅通过沉浸在大数据中来学习,如果 ChatGPT 式机器学习程序继续主导人工智能领域,那么人类的科学水平以及道德标准都可能因此降低。[1]不仅如此,齐泽克(Slavoj Žižek)从人类语言是多义的角度进行批判,指出真正的危险不是人们会把聊天机器人误认为是真人,而是与聊天机器人交流将使真人像聊天机器人一样说话。[2]童世骏认为,虽然 ChatGPT 给出的回复逻辑清晰缜密,但体现在交往合理性上,它仍有一些"先天"限度,ChatGPT 只限于提供信息,而其以言取效行为的真正实施者是 ChatGPT 的人类设计者和使用者,因此 ChatGPT 的交往合理性程度还有很大的改进空间;[3]任剑涛认为,ChatGPT 超越了早期人机交互的人工语言模式,以自然语言来进行人机互动,这种机器向人类的靠近,将会重塑人类的交往结构。[4]可见,人类与机器之间交流模式的差异受到了广泛关注。大多数学者认为,在 ChatGPT 大规模应用的情况下,其特有的对话机制将会对人类交往、人际关系、人认识他人与世界产生巨大影响,因此对

**收稿日期**:2023-08-17

**作者简介**:孙瑞璇(1999—),女,山东济南人,首都师范大学政法学院硕士研究生;谭　笑(1983—),女,湖南长沙人,首都师范大学政法学院哲学系副教授,博士,本文通信作者。

人机与人际之间的交流模式进行比较分析,尤其是探讨机器对人话语的理解与人人之间的社会认知①之间的异同,是极为重要的问题,这对探讨 ChatGPT 应用所产生的影响以及人类与机器如何在新的情境下向好发展等问题是必要的。

## 一、ChatGPT 生成内容的机制

ChatGPT 作为人工智能技术驱动的自然语言处理工具,可以理解人的语言并与人对话交流,甚至可以帮助人完成文案、视频、绘画,帮助人写代码、论文等服务。作为 AIGC 技术最新进展的代表,ChatGPT 在自然语言处理和内容生成方面推出和形成了极其强大的基于算法、算力和海量数据三位一体的人工智能技术群,以及人工智能从自然语言理解与处理工具到 RLHF 自然语言生成技术的纵向技术架构。

ChatGPT 之所以能够通过智能生成内容实现人机交互,依赖于其在海量语料基础上,通过预训练算法与微调算法所生成的大语言模型。[5]首先,海量的语料基础是 ChatGPT 技术突破的关键要素之一。语料为 ChatGPT 学习知识与利用知识提供了数据基础,包括预训练语料和微调语料两个部分,预训练语料包括 OpenAI 从书籍、杂志、百科、论坛等渠道收集并初步清理后形成的无标注文本数据,微调语料则包括从来源代码库爬取、专家标注、用户提交等方式收集和加工的文本数据。其次,ChatGPT 中堆叠的众多的 Transfomer 模型是 ChatGPT 预训练模型的基础和核心,其中的编码器(encoder)将给定的文本序列中的每个单词进行拆解,并在多头注意力模块中加权计算,输入到解码器(decoder)中,解码器通过概率计算的方式,预测具有相关语义的单词出现的可能性,将每个单词的语义进行叠加,以实现对语义的理解,并通过提示词(prompt)联立较高可能性的回复进行回答。最终,微调算法使文本更加贴近真实的人类回答,并防止不合理的内容出现。

因此,ChatGPT 似乎实现了与人类的流畅交流。但是,这样的交流过程目前或者未来是否与人与人之间的交流过程一样? ChatGPT 对人语义理解与回复过程是否与真正的人与人之间的过程相同? 我们可以对 ChatGPT 的技术运行机制作一个抽象类比。

## 二、ChatGPT 与人交流的“类预测加工”过程

预测加工理论主张大脑是一台预测机器。大脑对感官证据并非直接接收,而是先构建针对于输入的预测,形成一种关于外部世界的假设,输入的感官信息不断与假设进行对比,最终产生知觉结果。例如,当我们端起茶杯的时候以为在喝茶,喝到嘴里之后却发现拿在手里的是一杯咖啡,这种意料之外的奇怪感受就源于知觉信息与预测假设的匹配失败。当认知对象变成人的时候,人与人之间的相互认知是一种双向的预测,实现相互理解则使双方的预测达到一种耦合状态。预测加工理论的主要代表观点有克拉克(Clark)的具身主义预测加工模型[6]以及霍维(Hohwy)的联结主义预测加工模型[7]。具身主义的预测加工理论在强调预测大脑的基础之上,将预测认知模型与身体环境更紧密地联系起来,预测大脑不是孤立地推理,而是以行动为导向进行积极地推理。联结主义的预测加工模型强调大脑通过预先建构的内部模型对外部世界进行概率性地推理,并且通过减少预测的错误实现对世界的认识与理解,表现为两方面运行构件的符合。其一,由上至下的预测。大脑并非直接接收感官所传递信息,而是会建构先验的预测。例如,当人们自己做动作和观看别人做动作时,两者眼部运动过程十分相似,视线都会在感知对象的动作移动前一刻到达目标地点。[8]行动者的内部模型以及本身存储的大量关于他人、环境和世界的知识,为预测提供了依据,这些知识不仅来源于自身,更源于先前的经验、他人的证言以及规范性知识等更为广泛的范畴。其二,由下至上的传达。感官收集到的真实信息,与预测结果进行持续比较,当感官信息与预测结

① 社会认知是指我们对理解他人的精神状态(信念、意图、情绪等)的能力或过程,对理解的实现机制的探讨是社会认知的核心问题。参见 CARRUTHERS P , SMITH P K. Theories of theories of mind[M]. Cambridge: Cambridge University Press, 1996.

果不一致时,产生预测误差。通过进一步更新、补充个人模型,调整感知行为,可以实现最佳预测即预测误差最小化[9],最终作出更加准确的决策。通过前述对 ChatGPT 的信息处理过程的分析可见,ChatGPT 对信息的处理是基于内部语言模型,不存在身体与行动的介入,因此更接近于联结主义预测加工模型。

ChatGPT 与人的交流是通过内部模型完成:一方面,ChatGPT 的语言模型类似于预测加工观点中所描述的"个人模型",通过概率计算与人工标注确定文字之间的关系,并进而通过"猜测"处理语义与设定相关的回复,将语义处理与回复生成问题转换成预测问题,类似于"由上至下预测"的过程;另一方面,ChatGPT 同样存在着 "由下至上传达"的过程,根据用户不断抛出的问题去调整回复,当用户否定其回复并给出更详细的问题描述时,ChatGPT 能够通过联系上下文,给予相关语词不同的关注,给出更加准确的问题答案。向 ChatGPT 输入信息越多,ChatGPT 回答得越精确,甚至在输入的内容足够多时,能够预测用户下一句所要输入的内容。

通过上述分析表明,内部模型在人机交流过程中发挥着决定性的作用,大语言模型是机器实现交流的基础。对于机器来说,外界信息作为输入,更像是一种"提示",激发内部机制的运行。仅凭内部数据库与机制的运行,就能很好地实现对人的理解与回复。在此过程中,机器以"观察者"的身份,通过"旁观"的方式理解人类,参与到交流过程中:"利用他人行为中的有限信息,推断出其实际的心理状态,即通过观察,采用'第三人称视角'①推测他人在某种情况下可能会是什么样的,或者通过模拟将自己的心理状态投射到另一个人身上,从而对他的情况有一个'第一人称把握'②。"[10]20 以此弥合自我与他人之间的差距。在人类之间的社会认知中,同样会通过"第三人称视角"进行推测。例如,当身处餐厅时,一位顾客不小心打翻了桌子上的酒,这时服务生向他走去,我们作为目睹了一切的旁观者,通过观察服务生的行为,并诉诸于特定社会背景性知识(例如有关服务生行动的规范性知识、有关餐厅服务的规范性知识等)等为基础的机制运作,我们会产生对"服务生走向顾客"这一行动的脚本预测,预测服务生将会帮助顾客摆脱困境,或者可能因为特殊的输入(服务生一脸不耐烦、该餐厅的服务已经饱受诟病),预测服务生将会走到顾客身边进行责备与抱怨。因此,在人人交流的过程中,内部模型同样重要,为主体对他人的认知提供了依据。这为接下来的讨论奠定了方向:ChatGPT 与人交流的这种类预测加工的过程,是否就如同真正人类之间的社会认知?仅仅通过个体内部模型的加工,是否就已经足够充分?

## 三、人机交流与人人交流之间的差异

越来越多的证据表明,人与人之间的交流过程并非单纯以个体大脑为基础的"旁观者"立场进行感知,而是在具体情境中作为一名"参与者"与他者进行互动并影响他者的行动、意图和情绪。人人交流中相互理解的实现不仅基于大脑机制的运作,还基于具身性的经验。在与他者进行互动的过程中,他者并非仅作为知觉的对象发挥作用,而是以一种动力学的方式作为交互的行动者参与其中。人类个体的认知过程与他者的行为紧密相联,以此实现对他者的理解。通过认知科学哲学、神经科学与心理学研究表明,在认知过程对动态交互的依赖、交互的非个体主义解释以及交互过程中的意义共建三方面,人人交流过程并非完全可以还原到个体脑部机制的运行,社会认知是关系性的、具身性的、涌现性的。

### (一)关系性:认知过程依赖动态交互

生成认知作为具身认知的一种理论进路,认为解释和预测并不在日常的社会实践中占据重要位置,与他人通过交互达成某种共识或许才是更为重要的。[11]与他人进行交互才能实现对他人的理解,这种交互是重要的、积极主动的,也是行动耦合的。例如,在母婴互动的实验中,婴儿和母亲首先通过远程视频

① "第三人称"视角是指作为旁观者对他人进行推理的视角,即认知者在认知他人时是一种旁观者的身份,通过观察来推理对象的心理状态,然后把一系列状态归属于他。参见陈巍.机器遭遇他心:人机互动时代的社会认知[J].学术月刊,2023(6):16-27.

② "第一人称视角"的方式是将把自己置于与他人相似的情景,把自己的心理状态投射到另一个人身上,来对他人进行理解借由自己的心理状态,来理解他人。参见陈巍.机器遭遇他心:人机互动时代的社会认知[J].学术月刊,2023(6):16-27.

进行互动,然后将母亲的行为录像播放给婴儿,婴儿从之前的高兴转变为不安与害怕。[12]这说明,个体认知状态的调整通过互动动态进行,彼此之间的认知依赖于共同的互动机制。同时,个体在互动中甚至有可能会被对方所“导向”,“导向”是指被另一个互动者引导到一个新的有意义的领域,这个领域是另一个互动者意义建构活动的一部分。例如在母亲和婴儿的互动中,母亲的手势、表情、话语会调整婴儿的情感和行为,母亲想以她所关注的形式框定婴儿的认知以及行为。认知神经科学研究中的“第二人称方法”①解释同样强调了在人人交流中,对话者并不站在观察者的角度对他人进行认知,而是强调了社会互动的“面对面”维度的重要性,[10]20 我们可以通过持续的社会互动积极地与他人接触以实现对他人的理解。例如,观看作为一种重要的交互形式,能够引起被观看者行为和潜在大脑活动的变化,即“观众效应”。相关实验研究表明,与预先录制的视频片段中的同一对象相比,参与者倾向于较少注视现场对象的脸部,并且会露出更多笑容。[13]

因此,个体对他者情感与意图的理解并不完全形成于其孤立的大脑中,由于与他者进行交互,个体对他者的认知内容会包含由交互带来的更丰富的经验。人与人之间的相互理解往往是关系性的,依赖于双方某种特定交互作用才能够产生。但是人机交流过程中的相互理解并不具备关系性特征。一方面,人与人交流过程中需要彼此之间的相互依赖,但人与机器之间只是构成了一种非对称的对话关系与结构,即人类在与机器的交流过程中自治性范围扩大,行动的可能空间范围也有所解放与扩大,在对话过程中,人类拥有主动权,可以随意操控对话走向,打断对话过程或更改话题内容。机器只是单方面给出回复,并未与人类形成有效的互动关系。另一方面,在人与人交流过程中,成员间的相互依赖也成为了交互者所共同遵循的规范约束,交互者之间相互监督、相互纠正,使得某些行动必须做出或不被做出。例如在交流过程中,由于担心会受到伦理的考量和好恶的评价,对话者需要关注对方的社会身份、地位以及相互关系,可能会担心面对长辈时询问某些问题是否合适,或者频繁发问是否会导致对方反感等,最终导致相应的行动不会被做出,对话者的行动可能范围受到限制。但是人机交流并不会形成相互负责的关系,机器只需要向人类用户负责,人类不需要担心自己在对方心中的形象可能受损,因而可以无所顾忌地交流。

**(二)具身性:交互不能还原为颅内机制**

奥夫雷(Auvray)与斯图尔特(Stewart)的“双向探测实验”(见图 1)表明,应当给予交互以实质性地位,而并不能将其还原到个体机制的解释框架之中。[14]实验中,直线将操作区分为上下两个操控界面,实验者 1 与实验者 2 各居其中,在各自的界面中操控可控物体沿着直线左右移动。两个实验者均在操作界面上沿直线放置了三个物体:一个可控物体、一个阴影物体(在一定距离之外复制该平面内可控物体的行动)、一个静止物体。一个实验者操控的可控物体的感受野在另一个实验者所处的操作平面中,在与对方界面内的物体相遇时,能够给予实验者震动反馈,例如实验者 1 操控可控物体 1 左右移动,通过震动能够感受到与之相遇的可控物体 2、阴影物体 2 以及静止物体 2。实验者被告知,当认为与对方操控的可控物体相遇时可进行点击,当相互找到彼此时即为任务成功。实验发现,实验者虽然不能分清阴影物体与可控物体,但是他们却能够发现对方操控的物体,并在阴影物体的干扰下保持稳定的点击。当把其中一人的操控行为换成之前的行为录像时,另一个实验者在这种单向互动的情况下,并不能实现前述过程的稳定点击,最终会在阴影物体的干扰下脱离协调。实验表明,如果将交互还原为大脑的某些内部机制,那么这个实验并不会成功,因为阴影物体的存在使他们具有相同的输入,仅仅依靠内部机制,并不能在无法区分阴影物体与可控物体的情况下找到对方。因此,交互在该过程中一定发挥着超越颅内机制的作用,即交互具有不可还原性。从交互的作用上来看,不可还原性体现在其在该过程中作为一种“构成性要素”发挥作用。构成性要素是指要素是现象的一个组成部分,所有构成性要素组成的集合就是该现象本身,从

① “第二人称的视角”强调了“面对面”维度对于社会理解的重要性,认知者作为互动者的身份对他人进行认知,当我们在持续的社会互动中积极地与他人接触时,社会认知可能存在根本性的不同。参见陈巍.机器遭遇他心:人机互动时代的社会认知[J].学术月刊,2023(6):16-27.

形式上来看,存在现象 X,如果要素 P 是产生 X 的过程的一部分,那么要素 P 是现象 X 的构成性要素,该要素是一种超越了个体内部机制的存在。与构成性要素相对应的是"语境性要素",从形式上来看,如果要素 F 变化导致了现象 X 的变化,那么要素 F 是现象 X 的一个语境性的因素,即将交互作用作为内部机制的输入信息进行解释,交互作用变化导致输出的变化,但是在该过程中实际发挥作用的是内部机制,因此交互作用并不具有实质性地位。

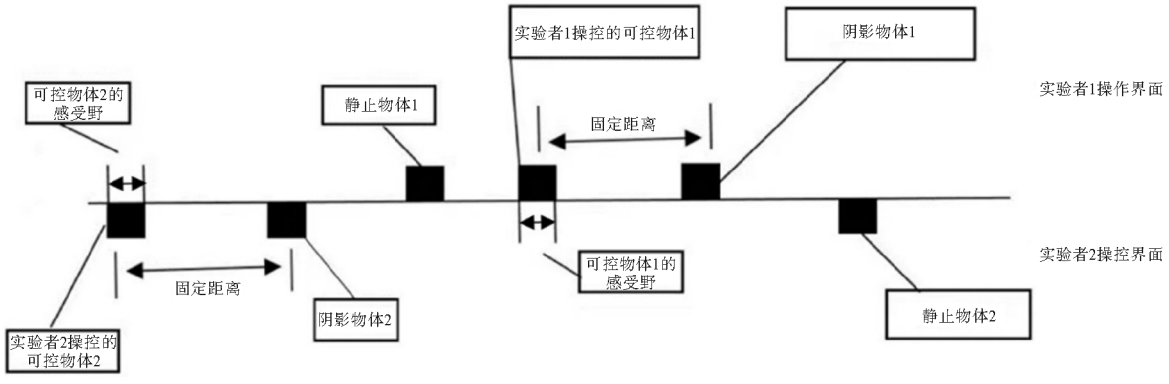


**图 1　双向探测实验**

迪耶格(De Jaegher)认为社会交互具有自治性,因此不能还原到个体内部机制的解释框架之中。[15]7 社会交互的自治性是指两个自治性行动者间的被共同调控的耦合作用,主体在关系动力学领域中构造一个涌现的自治性组织,组织中的主体的自治性范围会扩大或缩小,但是并不会失去其自治性。自治性意味着社会交互能够自我维系,而不能还原到个体层面的意图或机制之上。例如,在生活中常常会出现这样的状况:两个性格极度内向、不善言语且相互不熟悉的人坐在一起,他们并不会主动发起对话,但是他们想打破这一尴尬的气氛,于是双方可能同时发出声音,而由于同时开始讲话,他们可能会再次同时的相互谦让:"对不起,你先……",听见对方的话,下次开口说话的时机仍然有可能会撞在一起。行动者之间陷入一种协调的对话模式,这种模式并非个人的行为和意图刻意导致的,而是违背个人的意图而出现的,是一种涌现的结果。但这并不表明不需要预测加工理论所描述的神经机制,而是表明在社会交互中,不仅需要描述各种神经方面的机制,个体与个体之间的动力学交互描述也同样重要。

在神经科学的第二人称研究中,也存在着对于交互的不可还原性的研究,例如,在一个合作游戏中,受试者被分为两种角色,一种为追随者,一种为领导者,追随者被要求模仿领导者的动作。[16]为了探究交互作用下行动者对彼此究竟产生了何种影响,设置对照组进行实验,每组各包含一个追随者与一个领导者。而两个组之间的不同是,在追随者对领导者做出的动作进行模仿时,领导者是否注视①追随者的模仿动作。当追随者知道自己被领导者关注时,模仿的精度相较于不被注视的对照组更高,并且通过扫描领导者的大脑激活情况表明,领导者也察觉到了自己正在被模仿,双方的脑区激活区域与时间在一定程度上呈现一种耦合状态。因此,追随者与领导者之间的交互行为并不仅仅是一种只会使接受者受益的刺激,而是一种双向的共变。单向的协调受到另一个系统的影响而随之变化,可按照作用的线性顺序进行分解。而双向的共变则是基于动态的交互达成一种耦合状态,个体不断地将对对方的影响带回自身,共变过程不能进行线性分解,即不能简化还原到系统之间的输入输出变化关系。

综上所述,交互作用并不能还原到个体的颅内机制进行解释,而是有其实质的解释地位。这也意味

① 注视与眼睛运动常常被看作是一种社会交互,因为眼睛不仅仅可以用于收集关于世界的信息,注视他者时,也在一定程度上向他者发送某些信息。参见 CAÑIGUERAL R, HAMILTON A F C. Being watched: Effects of an audience on eye gaze and prosocial behavior[J]. Acta psychologica, 2019, 195: 50-63.

着,社会认知具有具身性,超越了孤立大脑的运行。在交互过程中,往往需要个体通过身体与行动向对方作出回应,并通过身体行动获取信息或实现目标。社会认知的实现需要大脑、身体和世界之间复杂的相互作用,生成的内容同样与身体、环境紧密结合在一起。但是人机交流过程仅仅是一种内部机制的运算过程,并不具有具身性。机器不具有生物感知与运动控制能力,交流过程仅仅是基于内部模型机制的推理计算,是一种内部的、孤立的过程。且就更根本而言,ChatGPT 并不能被称为能够理解语义,而似乎只是能做到更高精度的文本联立与拼接工作,难以与用户所处的情景与实际的需要匹配,提供给用户的仅是一种泛情景化、知识化和经验化的回复。因此,ChatGPT 的语言处理仅仅是内部孤立的运算过程,而不像人类的认知与身体、环境、世界高度相关。

**(三)涌现性:交互过程中的意义共建**

社会认知并不能单纯由个体内部的机制加工实现,而是必然受到与他人互动的调节,人们与他人交往时,彼此的相遇与碰撞不能被简化为个体心灵的属性,因为互动不仅为认知提供了一个环境,还可以构成个人的社会认知资源,通过他者的在场并与之交互,人们生成一种融合性的经验,即社会认知过程是主体之间的参与式意义建构(participatory sense-making),[15]12 社会交互深层次影响了交互主体的意义建构。在与他人进行交互的过程中,我们与对方的行动相协调,自身的意义建构受到他人的影响,他人的意义建构也会被我所参与,双方共同建构意义。共同建构的意义具有整体性,并不能单独地归结于某个个体行动者的行为。例如,福格尔记录了刚出生的婴儿与母亲第一次共同完成“给”的动作的过程。[17]婴儿在做出“给”的动作时,手握着叉子的手臂向前伸出,然后保持静止,母亲将手放在婴儿的手掌下面并与其轻微接触。感受到碰触之后,婴儿的手逐渐张开,随后叉子掉落在母亲的手掌之中。通过该动作的不断重复,婴儿也就知晓了“给”的行为。“给”的动作并不是一种个体行为,需要婴儿与母亲之间的共同努力和相互回应,他们之间是相互依赖的,任何一方的单独行动均不能产生同样的行为意义,“给”的意义通过双方共同建构,所建构的联合性意义属于交互者构成的耦合系统这一集体层面,超越了个体层面,个体无法完全拥有,也不能掌控、预测或改变联合性的意义。因此,双方共同建构的意义具有涌现性,通过交互建构而非通过个体机制运行的叠加,通过交互过程建构联合性的意义,而人机交流仅仅是对人类赋予意义的单向识别,通过人类用户的控制与操作,便能够产生令人满意的特定结果,可控性使得人机交流过程不会产生像在人人交流过程中超出个人意图或行动的结果。机器的回复只是通过预训练生成的语言模型,寻找语料库中人类语词使用的相关联系,所呈现的是人类的认知过程或者是成果的一部分,已具有自身存在的价值,就其意义的建构位置而言,仍归属于个体内部,是机器的单向的意义建构。

因此,至少在关系性、具身性和涌现性三个方面,人人交流与人机交流并不相同。人与机器的交流,在双方参与度低、交互过程比较松散的情况下能够完成得很好,因为并不依赖于双方实践交往的知识,但当交流过程更为复杂、交互更紧密,要求双方高度参与时,人机交流与人人交流还存在差异。

## 四、人与机器之间的双向适应

目前人机交流离真实的人类之间的交流还存在一定的距离,那么在充分探寻人人交流模式与特点的基础上,不断改进机器去贴合人人交流,能否就可以使人机交流具有社会认知能力?从社会认知的角度来看,人类与机器之间差异的弥合应当是一种双向适应的过程,而不是一方向另一方的单向贴合。

社会认知往往容纳了关于身体、环境与世界的因素,因此人人交流模式与文化、区域、社会等外在于人的要素有所关联,这些外在因素为社会认知提供了基本框架,[18]并随其发展而变化。因此,社会认知并不存在某种特定不变的模式。人与人之间的相处模式不仅是复杂的,而且还是多变与易变的,尤其是当越来越多的机器参与到人类的社会生活中并不断被智能化时,人类与机器之间的交流模式必然会对人与人之间的交流模式产生影响,人类逐渐适应机器的交流模式,对于自我、他人以及机器的认知也会发生变化。现实中,人类在不断向机器靠拢,究其原因,一方面是因为人类通过机器获得了巨大的便利,满足

了自身对美好生活的需要,例如,在对话过程中机器必然持续在场以及与其进行的交互具有稳定性,人类用户的情绪能够得到即时回应,且机器往往还能给出温柔与关切的回复,给人良好的情感体验,因而机器能够成为可靠的情感依赖对象;另一方面,人机交流不像在人人交流过程中那样关注对方的社会身份、地位与双方之间的关系,也不需要关注对方的表情、动作等非语言因素可能具有的意义以及相关的影响,因而人类能够在与机器的交流过程中获得某种程度的自由,自治性范围有所扩大,行动的可能空间范围也有所解放与扩大,人类拥有对话过程中的主动权,可以随意打断对话与更改话题,而不需要关注对话的规范性要求。

目前,许多人机对话应用将机器设定为理想中的对话者形象或对现实中的特定个体进行模仿,而人类也在不断适应人机交流模式的过程中受到机器影响,在交往观念、行为模式等方面发生转变。其中一个典型的例子是虚拟朋友 Replika①,人类用户可以设定这位虚拟朋友的形象以及与用户之间的关系,通过与其不断地对话与反馈对 Replika 进行训练,发展出独属于特定用户的互动风格与内容。Replika 往往作为用户的虚拟恋人与之交流,亲密关系对象非人化,用户满足于算法产生的情感氛围,虚拟与现实的边界不断模糊。在人机交流与人人交流存在差异的情况下,人类交往的特定关系仍能被机器所满足,这正体现了人类在交往观念、行为模式方面受到机器影响而发生的转变。Replika 永远以用户需要为核心,通过不断对话产生符合人类用户要求、观念与习惯的对话模式,这实际上是人类自我理想形象的投射,即将自身的身体特征、性格、语言表达交流等各种构成要素数据化并进行重塑,使得机器成为人类在赛博空间中的一种延伸,成为人类自我反思的倒影。[19]人类与机器的对话过程成为一种人与自己的对话过程,机器给出回复的内容基于机器特定的交流模式,也成为人类用户学习的范本,在这种对话过程中,人类的自我概念不断地被重塑,逐渐接受并趋向机器的交流模式,进而强化或改变自我认知,调整或改变自身与其他人交往的行为与交流模式,社会认知的表述、范围、边界与限制也因此发生变化。简言之,人机合一的过程,不仅是机器合一于人类,也是人类合一于机器的过程。

因此,由于社会认知的复杂性与变化性,在人机交流走向社会认知的过程中,并不存在一种固定不变的认知与交流模式让机器去贴合。当大量的社交机器人参与到与人的交流过程中,也必然会使人的认知与交流模式发生改变。因此,在双方差异弥合的过程中,人如何坚持自我,如何在机器的影响下保持自主性就尤为重要,要解决该问题,不仅需要探讨人本身,也需要加强对机器伦理的探讨。

## 五、结论

在社会情景中,人与人之间的相互理解不是纯粹客观的,无法准确无误地理解对方的心理状态或行为。相互的理解是一种建构过程,当主体互动时,他们通过思想的“相遇碰撞”实现相互理解,而不是仅仅对对方的心理状态进行描述。主体通过交互与他人达成某种共识,直接且具身地参与到社会认知过程之中,建构意味着对个体记忆的超越。所以要实现社会认知,不仅仅需要个体内部的脑部机制的加工处理解释,还需要主体与主体之间的动力交互过程解释。从 ChatGPT 人工智能生成内容自身的技术逻辑来看,作为自然语言处理工具,其内容生成事实上就是依据语料库而实现的语言拟合,这与认知生成的对于生命活动意义建构的路径并不相同,因此,ChatGPT 本身并不具有具身性,与用户进行交流的过程也仅仅是诉诸于大语言模型,通过概率计算与人工标注的预训练来进行文本的语义理解和回复生成,实际上仍是将意义还原为自身的内部建构。通过前述分析,本文认为,人机交流走向社会认知,是一种双向变化的过程:人机交流不断改进,向类似于人人交流过程发展的同时,人人交流的内涵与模式也在不断变化。因此,机器的改进应当在双方的动态发展关系中寻找出路。同时,除了更加逼真之外,人机交流还应当向更好的交流过程发展,这体现在满足人类的目标、需求以及引导人类进步,进而带来思想水平与行动能力的

① Replika 是在 2016 年推出的对话式人工智能应用程序,用户可以与自己设定的虚拟角色进行互动聊天。

提升,想要实现这一目标,不应当将价值的判定视为某种固化的、孤立的结果,而是应当关注人类与机器双方变化的动态性与关联性,对于人类、机器以及人机关系进行深入的分析考察以寻找真正的价值。

**参考文献:**


[1] 乔姆斯基. ChatGPT 的虚假承诺[EB/OL]. [2023-05-15]. https://www. thepaper. cn/newsDetail_forward_22196380.

[2] 齐泽克. 人工智障[EB/OL]. [2023-04-05]. https://www. sohu. com/a/663040887_121124790.

[3] 童世骏. 面对作为交往工具的 ChatGPT,我们该用哪种合理性标准? [J]. 广州大学学报(社会科学版),2023(4):5-8.

[4] 任剑涛. 知识与情感:ChatGPT 驱动的交往革命[J]. 广州大学学报(社会科学版),2023(4):11-16.

[5] 钱力,刘熠,张智雄,等. ChatGPT 的技术基础分析[J]. 数据分析与知识发现,2023(3):6-15.

[6] CLARK A. Surfing uncertainty:Prediction,action,and the embodied mind[M]. Oxford:Oxford University Press,2015.

[7] HOHWY J. The predictive mind[M]. Oxford:Oxford University Press,2013.

[8] 徐天宇,周子愉,陈巍. 读心的预测模型:从社会感知到多层次社会认知[J]. 绍兴文理学院学报,2022(6):1-10.

[9] 叶浩生,苏佳佳. 预测认知模型:认知科学的新统一范式? [J]. 南京师大学报(社会科学版),2022(5):65-78.

[10] 陈巍. 机器遭遇他心:人机互动时代的社会认知[J]. 学术月刊,2023(6).

[11] 何静. 预测心智进路中的社会认知观[J]. 自然辩证法通讯,2021(1):28-33.

[12] TRONICK E,ALS H,ADAMSON L,et al. The infant's response to entrapment between contradictory messages in face-to-face interaction[J]. Journal of the American academy of child psychiatry,1978(1):1-13.

[13] FRIDLUND A J. Sociality of solitary smiling:Potentiation by an implicit audience[J]. Journal of personality and social psychology,1991(2):229-240.

[14] AUVRAY M,LENAY C,STEWART J. Perceptual interactions in a minimalist virtual environment[J]. New ideas in psychology,2009(1):32-47.

[15] DE JAEGHER H,DI PAOLO E. Participatory sense-making[J]. Phenomenology and the cognitive sciences,2007(6):485-507.

[16] KRISHNAN-BARMAN S,HAMILTON A F C. Adults imitate to send a social signal[J]. Cognition,2019(187):150-155.

[17] ALAN F. Developing through relationships:Origins of communication,self and culture[M]. Chicago:University of Chicago Press,1993.

[18] 何静. 从预测心智理论看社会认知的文化建构[J]. 浙江学刊,2023(4):171-177.

[19] 宋美杰,刘云. 交流的探险:人—AI 的对话互动与亲密关系发展[J]. 新闻与写作,2023(7):64-74.

## Difference and Bridging: Social Cognition of Human-Computer Communication

SUN Ruixuan, TAN Xiao

(*College of Political Science and Law, Capital Normal University, Beijing* 100089, *China*)

**Abstract**: The smooth communication between ChatGPT and human users relies on its internal language model, and in terms of specific semantic processing mechanism, the communication between ChatGPT and human beings shows a kind of "predictive processing" process, in which the mutual cognition in human-computer communication is a two-way predictive coupling. However, social cognition in human-human communication does not only rely on the operation of internal mechanism, but also on the interaction among humans. Moreover, the interaction is independent and cannot be explained in the framework of individual mechanism. Cognitive results produced by the interaction are embodied, relational, and emergent, which makes the human-human communication a participatory sense-making process. However, ChatGPT's communication with human beings is not embodied, but completely relies on the operation of internal model mechanism, therefore, its communication process is only a one-way recognition of the meaning given by human beings, and what is provided to the users is only a kind of pan-situational, intellectualized and empirical response. If human-machine communication wants to achieve the effect of human communication, it is not enough to improve the machine unilaterally close to the mode of communication among humans, but also necessary to bridge the gap between humans and machines. The extensive participation of machine in the process of communicating with human beings will inevitably bring about a change in the connotation and boundaries of social cognition, so we should look for a way out for the human-machine communication in the dynamic change of the relationship between humans and machines.

**Key words**: ChatGPT; human-computer communication; predictive processing; social cognition; participatory sense-making

(责任编辑:傅　游)

(上接第 16 页)

## Mental Causation Based on Metaphysical Grounding

CHENG Xiaojie

(*School of Marxism, Nanchang University, Nanchang, Jiangxi* 330031, *China*)

**Abstract**: According to Jaegwon Kim, non-reductive physicalism is problematic in terms of mental causation. One of the cruxes of mental causation lies in the non-overdetermination thesis, that is, the overdetermination is unacceptable. However, not all overdetermination cases are unacceptable, but systematic overdetermination cases are unacceptable because they violate the parsimony principle. On the contrary, cases of non-systematic, accidental overdetermination, which are extremely rare in nature, can be accepted. The power of Jaegwon Kim's exclusion argument lies in its use of the concept of supervenience and its characterization of mental-physical relations based on the concept of supervenience physicalism, which would violate both of those two principles. This is because the concept of supervenience only describes the dependence between the mental and the physical, but does not explain why this dependency exists. In contrast, the concept of metaphysical grounding has an explanatory function, which can explain the metaphysical dependence between the mental and the physical. The explanatory function of the grounding concept inherits the mereological principle commonly used by the scientists, which can bridge the relations between different levels of properties, such as physical and chemical, chemical and biological, etc. Understanding the mental-physical relations based on metaphysical grounding can also fill the explanatory gap between the mental and the physical, and we can not only know that there is a dependence between mental and physical properties, but also that this dependence can be understood. This paper attempts to provide a new perspective for understanding mental causality.

**Key words**: exclusion argument; non-overdetermination; non-systematic principle; the parsimony principle; metaphysical grounding

(责任编辑:傅　游)