doi:10.13582/j.cnki.1672-7835.2024.03.017

机器无心人有心:从心-机关系再思人工智能的发展

张亮,俞泉林
(南京大学 哲学系,江苏 南京 210008)

摘　要:人工智能的迅猛发展看似模糊了心灵与机器的界限,但是二者之间依旧存在着不可逾越的鸿沟。经验的主观性使得心灵与身体很难被彻底分隔开。物理主义还原论无法应对那些具有文化特色的心灵活动。虽然在哲学视阈中,人们并不能借助机器创造意识,但是在现实应用中,人工智能却表现出了人独有的"偏见"。正是数据库内容的不完善和算法设计中的不平等,让过去隐藏在认知主体中的偏见浮出水面。人工智能的发展仅仅依靠模仿是行不通的,在新一轮的科技狂潮中,要为我国的人工智能产品安上一颗"中国心"。

关键词:人工智能;身心问题;"中国心"
中图分类号:B82　　**文献标志码**:A　　**文章编号**:1672-7835(2024)03-0146-07

亚里士多德坚信,"人生而求知"①。这种求知不仅体现为对外在世界的探索,也体现为对人类内在精神的挖掘。什么是人类意识的本质?人的意识是否能够被创造?计算机的出现让关于人工智能的哲学追问从科幻世界来到现实生活,大型数据库、云计算、脑机互动、生成式人工智能等等,加速涌现的新奇科技成果不仅激发了市场的狂欢,更让许多局中人狂热地相信,意识的创造、智能的创造会在不远的将来变成现实。有超级银行家预言二十年内将出现超级人工智能,将比人类聪明一万倍!② 如何理解人与机器的鸿沟?人的意识是否可以在机器中被重构?人工智能给人类带来的是更平等的解放,还是被推入偏见与对立的深渊?历史表明,人工智能的发展也是螺旋式上升的。在新一波高潮正加速演进之际,我们有必要以史为鉴重新思考前述基本哲学问题,以便对不期而至的危险做出及时的回应。

一　似旧仍新的哲学结论:人与机器的鸿沟依旧存在

计算机出现之初,图灵(1912—1954)、纽曼(1897—1984)等人工智能先驱就曾乐观展望过人工智能的未来,相信智能机器可以像人一样在"经验中成长"③。2022 年以来,Chat GPT 的横空出世再次让社会大众沉浸在科技突破引发的巨大狂欢中,同时也再次激发了社会大众对人工智能发展的深深忧虑:未来有多少工作会被取代?人工智能该如何被法律监管?在社会层面的忧虑背后,大众内心深处有一个尚未被清晰意识到的终极忧虑:具有人类意识的人工智能何时会出现?换言之,社会大众所忧虑者,正是图灵等所预言和期待者,即人与机器、人的智能与人工智能之间并

收稿日期:2023-12-06
基金项目:南京大学"认知成像"揭榜挂帅项目(2023300320);南京大学中央高校基本科研业务费项目(2023300118)
作者简介:张亮(1973—),男,江苏徐州人,博士,教授,博士生导师,长江学者特聘教授,主要从事国外马克思主义哲学、当代西方左派思想史、历史唯物主义研究。

①Aristotle. *Complete Worsk*, Princeton, N.J: Princeton University Press, 1991, 980a22-980a27.

②Anton Bridge. *SoftBank CEO Son Says Artificial General Intelligence will Come Within 10 Year*, https://www.reuters.com/technology/softbank-ceo-masayoshi-son-says-artificial-general-intelligence-will-come-within-2023-10-04/.

③Copeland, Jack edited. *The Essential Turing: Seminal Writing in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life*, Oxford: Oxford University Press, 2004, pp.472-475.

不存在不可逾越的鸿沟。难道人与机器之间真的不存在不可逾越的鸿沟，“大脑就是一台计算机”？[①] 早在半个世纪前，布雷斯韦特（1900—1990）、戴维森（1917—2003）、内格尔（1937—）等哲学家已经对这一问题进行了深入讨论[②]，指出这一问题的本质是关涉身心关系的哲学问题：意识的本质是否是可以彻底还原的物理过程？他们进而将这一大问题分解为三个小问题，并逐一做出了明确的否定性论证。

首先，身心是否是可分的？心灵是否可以独立于身体而存在？在《人的问题》一书，内格尔深入思考了这个问题[③]，并将这个极其抽象的问题转化为一个极为感性的问题：人可以想象成为蝙蝠是什么样吗？内格尔认为不可以，因为经验是意识不可或缺的成分，而经验具有主观性，这导致意识难以超越主体经验的边界[④]。“子非鱼，安知鱼之乐？”尽管人类已经知道蝙蝠的感觉器官是如何发挥作用的，但因为人类的生理结构和蝙蝠完全不同，所以无法理解蝙蝠经验中的世界。内格尔进一步指出，这种经验上的隔阂不仅存在于不同生物之间，而且存在于同一物种之间。例如，视力正常的人和先天失明的人之间就无法理解彼此是如何感知外在世界的：即使闭眼体验很长一段时间的盲人生活，视力正常的人依旧可以想象出绚丽的晚霞、雄伟的高山，但因为从未用视觉去感知这个世界，先天失明的人则无法想象出颜色的差异以及山川河流的样子。也就是说，无论以何种方式，视力正常的人和先天失明的人都无法描绘出对方经验中的世界。既然有机物之间的经验尚不能互通，那么，假设人工智能也有感知外在世界的能力，我们又如何能够证明无机物的机器与有机物的人能够互通经验、彼此理解呢？此外，经验的主观性使得意识主体无法想象超越自身经验范围的东西，再充分的想象力也受制于经验的限制，例如，十年前科幻电影中对宇宙飞船内饰的想象，充满了大量的机械按键和仪表盘，如今科幻电影中宇宙飞船安装的则是一块块手势操作的全息投影屏。因此，我们可以得出结论，心灵和身体很难完全割裂开分析，心灵的边界受制于身体的边界。

其次，心理状态是否都可以被还原为身体状态？物理主义还原论并不排斥经验的主观性，也不否认心灵无法绕过身体感知世界，它坚信心灵能够被物理还原。在物理主义还原论那里，缤纷多彩的世界可以还原为无数小颗粒，然后用物理学语言达成对世界的客观描述，进而消解不同个体之间在物理层面之外的感知差异。这种观点让人有穿越感，仿佛回到了前苏格拉底时代，但却备受今天科学家的推崇，之所以会如此，归根结底是因为现代科学不仅构建出了复杂庞大且自洽的系统，而且在解释客观世界时获得了空前的成功。于是，人们就很自然地试图将这种还原论推广到意识层面，只不过，在此意义上意识作为一种心理状态被还原为了身体状态。生物学和物理学等学科的研究进展已经使人们可以清晰地描述身体的状态，因此，如何描述心理状态，并寻求心理状态和身体状态得以吻合的情况，就成为还原论能否成立的关键。戴维森曾经论证指出，如果心理事件具有物理的原因和结果，它们必定有物理的描述[⑤]。心理状态和身体状态的关系可分为三种类型：

类型 A：主体的心理状态与直接经验保持一致，身体受到了外在环境的直接影响，并将影响反映在心灵中。例如，被针扎感到疼痛，此时身心是完全一致的。虽然不同的人对疼痛的感知能力有所不同（有人大声尖叫、有人面不改色），但感知能力的不同并不影响心理状态和身体状态之间的同一性。因为对于感知能力正常的人而言，意识上感到疼可以直接还原为身体在被针扎后产生的一系列神经反应，这可以满足还原论的要求。

类型 B：身体没有受到外在环境的直接影响，心理状态依旧可以和身体状态保持一致。例如，有些人并没有亲身经历溺水，但在看到深渊时依旧会感到慌张。总的看来，类型 B 不像类型 A 那

①理查德·马斯兰：《我们如何看见，又如何思考？》，顾金涛译，中信出版社 2021 年版，第 174 页。

②Copeland, Jack edited. *The Essential Turing: Seminal Writing in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life*. Oxford: Oxford University Press, 2004, pp.494-506.

③托马斯·内格尔：《人的问题》，万以译，上海译文出版社 2004 年版，第 225—244 页。

④Nagel, T. “What is It Like to Be a Bat?” *In The Language and Thought Series*. Cambridge: Harvard University Press. 1980, pp.159-168.

⑤Davidson, D. “Mental Events”. *In Contemporary Materialism*. London: Routledge, 2002, pp.122-137.

么普遍,以恐高症为例,有的人天生恐高、有的人完全不恐高、有的人甚至喜欢从高空坠落的感觉,但恐高这种心理状态依旧可以被还原为身体状态,神经学家依旧可以通过生理学语言来解释恐高这种心理状态,只不过这种还原过程更为复杂,且具有特殊性。

类型 C:同样的感知对象在不同历史文化背景下会使主体产生完全不同的心理状态。例如,同样看到满月,中国人想到的可能是阖家团圆的欣喜,欧洲人则可能视之为精神失常的前兆。不仅如此,同样的感知对象在不同的情境下会给主体带去完全不同的心理状态,例如,漂泊他乡的游子看到满月更多会是乡愁而不是欣喜。历史文化的复杂性导致人们很难用物理语言去描述,毕竟我们不可能像科幻电影那样在实验室中构建一个社会,然后通过不断重启这个社会去重复观察其孕育出的文化特征。同时,某些文化现象就如同天才们的灵感,即使历史重演,这些天才般的灵感也不必然会再次出现。也就是说,历史文化对主体意识的影响是潜移默化且不可避免的,但历史文化的特殊性与难以复制性却在本质上与物理主义的诉求冲突,因此,类型 C 中将意识置于还原论之中必然会面临难以克服的困难。

最后,隐含着的第三个问题得以被提出:心理状态不可还原的深层原因是什么?回答这个问题的前提是定义"意识"是什么。单纯且独立的感觉经验显然不足以构成意识。意识至少需要在诸多独立的感觉经验中建立联系,这种联系可以是命题式的(例如将夜空中有圆缺变化的天体理解为月亮),也可以是因果式的(例如因为看到月亮而感到伤心),正是后者即意识构建出的因果性联系与物理主义还原论难以兼容。用生理学语言描述,睹月思乡就是月亮的图像引起了大脑某种激素的分泌,从而让人变得忧伤。不过,仅仅将某种情绪还原为某种激素的分泌,这对意识的还原来说是远远不够的,因为还需要还原出导致这一情绪的其他原因,不然就像用"Moon"和"Luna"表示月亮一样,只不过是换了一个表达体系而已。心理状态能否完全还原为身体状态关键是物理主义能否解释所有意识中的因果关系,显然,物理主义做不到这一点。意识中的很多因果联系并不单纯来源于外在世界的刺激,在其形成的漫长过程中,意识会受到不同的文化背景、历史进程和社会环境的影响。这意味着,分析意识的恰当方式不应是静态的、孤立的,而应是动态的、历史的。

我们无意抹杀、否认在物理主义还原论基础上人工智能所取得的诸多科技成果,只是在科技再一次高歌猛进的时候,我们认为,对其理论根基进行适当的哲学反思依旧是必要的,因为不管是从哲学上讲还是从科技发展的现实上看,人工智能如何产生意识至今仍然是一个难题①。

二　机器非人但内在具有人之偏见

康德曾说,"像从造就人类的那么曲折的材料中,是凿不出什么彻底笔直的东西的。"②对于绝大多数科学家来说,智能机器和人类之间是否存在无法逾越的意识鸿沟太过形而上学。真正激励他们投身人工智能行业发展的是一种看得见摸得着的美好愿景:人有偏见而机器没有偏见,较之于人,人工智能更高效、更准确,也更客观、更科学。大自然的运行有其客观规律,是不以人的意志、意识、价值观等为转移的,"天行有常,不为尧存,不为桀亡。"那么,作为客观的物质系统的人工智能机器能够像大自然那样排除人的意志、价值观等不偏不倚地客观运行吗?加拿大学者通过分析 ProQuest、IEEE Explore、PubMed 等 8 个数据库,发现人工智能技术同样具有价值负载性,至少存在输入偏见、算法偏见和认知偏见等三种偏见形式,在涉及种族、医疗和商业等问题时,人工智能的回答就像"人"一样!③

首先,人工智能存在数据输入的偏见。受到图灵构想的机器教育与机器学习方案的影响,现行人工智能的发展技术路线是将人工智能置于庞

①一个由神经科学家、计算机科学家和哲学家组成的研究团队,系统研究了当前神经科学意识理论中的六种代表性理论,并从中筛选了一系列意识指标,以作为人工智能有无意识的评判标准。尽管该团队认为"构建满足这些指标的 AI 系统并无明显的技术障碍",但"当前的各类 AI 系统并无意识"。

②康德:《历史理性批判文集》,何兆武译,商务印书馆 1990 年版,第 10 页。

③El Morr, Christo. *AI and Society: Tensions and Opportunities*, Leiden :CRC Press, 2023,pp.1.

大的数据库中以实现人工智能的自我学习和进化①。也就是说，人工智能最终产生的结果很大程度上取决于数据库提供的数据，如果数据库的电子病历信息不能基本准确地反映社会大众身体和心理健康数据的自然分布状况，人工智能的认知结果就不可能是不偏不倚的。研究显示，各种电子健康数据库似乎都嫌贫爱富、不“喜欢”不处于社会中心位置的人②。所谓不处于社会中心位置的人，通常都是生活在交通不便的偏远地区的人群、面临住房困难的移民人群和社会经济地位较低的人群，他们的基础医疗保障水平低，很难定期去医疗机构体检，也很难有较为固定的诊疗场所，自然无法将自身的健康状况实时同步到医疗数据库中。医疗数据库如果缺失这类人群的数据，其分析结果自然仅仅能反应处于社会中心位置的人群。但遗憾的是，在基于电子健康记录生成预测算法的研究中，仅有一半的研究有意解决缺失数据问题③。因此，我们不无尴尬地看到，各种健康数据库似乎都支持“种族主义”观点，认为白人更容易生病，除了癌症、心血管疾病、精神疾病、糖尿病等常见疾病，白人女性似乎更容易罹患疤痕性脱发（CCCA）和盘状红斑狼疮（DLE）等自身免疫性皮肤病——在现实生活中，这些疾病在非白人女性中才更为常见④。人工智能之所以会从数据库中得到“种族主义”偏见，主要是因为美国的白人比少数族裔拥有更好的医疗保障，更容易获得更好的医疗服务⑤。基于数据输入而导致的人工智能的种族偏见在日常社会生活中同样存在，例如，在由神经科学家、计算机科学家和哲学家组成的科研团队 ImageNet 推动的一项计算机视觉研究中，由于其数据库中美国人占 45%，中国人和印度人加起来占 3%，但现实中美国人只占世界人口的 4%，中国人和印度人加起来则占 36%，于是，AI 就既匪夷所思又很好理解地将身穿白色婚纱的白人女性标注为“新娘”，却给印度新娘打上了“戏服”和“表演”的标签⑥。

其次，人工智能存在算法上的偏见。人工智能在算法上的偏见有多种来源⑦，有的容易识别，有的难以察觉。编程者在设计之初的模型选择属于容易识别的算法偏见。这种模型选择上的偏见既可以通过优先某种选择取向实现，也可以通过简化模型来实现。前者并不仅仅意味着刻意地通过算法将某一特定年龄段或特定居住地的人群排除在外，也可能是通过增加更模糊的价值规则，后者则是通过减少模型中的数据偏差实现特定的输出预期。NBC10 波士顿频道（全球广播公司）首席调查策略师科伦在 2021 年的一份报告中就指出，医疗保健系统使用的算法模型将治疗保健成本作为疾病的因素，有效地将黑人患者排除在额外医疗服务和护理管理计划之外⑧。因为医疗服务的获取资格存在结构性的不平等，所以黑人患者即使比白人患者更有资格，也因其医疗保健支出较低而被算法排除在外。人工智能对自然语言的学习与运用则属于难以察觉的算法偏见。人类的自然语言的表达很容易将价值性描述与事实性描述混为一谈，或是将价值取向嵌入事实性表述中。以 1815 年拿破仑重返巴黎为例，当时巴黎的一份报纸最初报道称“科西嘉的怪物在儒安港登陆”，随后则称“波拿巴占领里昂”，最后变成“陛

①Copeland, Jack edited. *The Essential Turing: Seminal Writing in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life*, Oxford: Oxford University Press, 2004, pp.472-475.

②Parikh, R. B., Teeple, S., & Navathe, A. S. *Addressing Bias in Artificial Intelligence in Health Care*. Chicago: JAMA: Journal of the American Medical Association, 2019, pp.322(24):2377-2378.

③Goldstein B A, Navar A M, Pencina M J, Ioannidis J P. *Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data*, *J Am Med Inform Assoc*, 2017, 4(1):198-208.

④Lee, M. S., Guo, L. N., & Nambudiri, V. E. “Towards Gender Equity in Artificial Intelligence and Machine Learning Applications in Dermatology”, *Journal of the American Medical Informatics Association*, 2021, 29(2):400-403.

⑤Institute of Medicine Committee on Understanding Eliminating Racial Ethnic Disparities in Health Care. In Smedley, B. D., Stith, A. Y., & Nelson, A. R. (Eds.), *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, *National Academies Press*, 2003, 76(4): S1377-S1381.

⑥Zou, James, Londa Schiebinger. “AI can be Sexist and Racist—It's Time to Make It Fair”, *Nature*, 2018(559):324-326.

⑦Ploug, T., & Holm, S. “The Right to Refuse Diagnostics and Treatment Planning by Artificial Intelligence”, *Medicine, Health Care and Philosophy*, 2020, 23(1):107-114.

⑧Curran, K. “Health Equity-Artificial Intelligence in Health Care Needs Scrutiny to Eliminate Bias”, *Health Progress*, 2021, 102(4): 68-70.

下将于今日抵达忠实于他的巴黎"①。除了命题式的表达或者成段落的表述,对自然语言的价值嵌入同样会出现在某些单词中,如将某种性格特点和某些职业自然地联系在一起,"医生通常是善良的""商贩是斤斤计较的"等等。因为完全习惯了自然语言的表达,所以人们很难主动去反思日常使用的自然语言,难以察觉隐藏在自然语言中的那些价值取向。为了让 Chat GPT 这类生成式人工智能表现得像人类,设计者们努力让它们学习、模仿人类自然语言的表达方式并进行内容呈现,其结果就是将隐藏在自然语言中的价值偏见传播出去②,人工智能的能力越强,价值偏见的传播就越广。例如,当我们问 Chat GPT 与种族肤色、身份认同等有关的问题,它的回答就带有明显的"白左色彩",会非常刻意地指出"不要将肤色与品德或能力联系在一起",但让它给世界伟大人物做排名,它又会直接将西方的伟大人物置于更前列。就像 BBC 报道所言,Chat GPT 的出现让非英语使用者掉队了③。

最后,人工智能的偏见本质上来自主体认知的偏见。科学的价值负载性决定了科学研究的过程充斥着科学家们作出的价值判断④。人工智能研发更是如此,因为在这一过程中,人们要求机器尽可能表现得更像人类,最好像人类一样看到灰蒙蒙的景色就能写出忧郁的诗。这在客观上放大了人类主体的认知渗透性与价值负载性。所谓认知渗透性是指,作为认知主体的人所拥有的以往经验、背景知识、注意力指向以及所处的环境,都会渗透进认知过程并影响最终的认知结果。为什么白人看到与犯罪相关的图片或文字描述更容易捕捉到与黑人相关的信息?认知渗透性使然也⑤。科学技术专家倾向于认为扩大信息搜索范围、增强信息处理能力和知识存储能力,可以有效避免确认偏误。按照这个逻辑,拥有人类无法比拟的信息收集、处理、存储能力的计算机应当是确认偏误的完美终结者。但令人遗憾的是,越来越多的迹象显示,大数据的出现似乎不是明显减少而是进一步加剧了社会的认知偏见。因为认知主体在拥有大量信息时更容易产生知识上的自信,认为自己所掌握的知识已经足够帮助自己做出客观准确的判断⑥。

人工智能内在具有人之偏见这个事实让我们看到,"事实上,我们所知的人工智能完全依赖于更广泛政治和社会结构"⑦。政治和社会结构不同,人工智能内嵌的价值观也会有所不同,从而使得具体的人工智能产品会内在具有隐秘的不同的文化身份。这就意味着,人工智能技术或许无国界,但人工智能产品肯定有国家!

三　以"中国心"立机:中国人工智能的发展方向

人工智能是新一轮科技革命和产业变革的重要驱动力量。早在 2014 年,习近平总书记就强调指出:"抓住新一轮科技革命和产业变革的重大机遇,就是要在新赛场建设之初就加入其中,甚至主导一些赛场建设,从而使我们成为新的竞赛规则的重要制定者、新的竞赛场地的重要主导者。"⑧正是在习近平总书记科技创新重要论述的指引下,我国的人工智能研究迅猛发展,量质齐升,已经稳居世界第一方阵。较之于 Chat GPT,目前相应中国产品的技术水平无疑还有显著差距,但在它们以 Chat GPT 为对象比学赶超之际,我们想追问的是,模仿 Chat GPT,我们能创造出技术领先并符合中国需要的生成式人工智能产品吗?我们的答案是否定的!遵循人工智能研究的

①Georges Blond. *Les Cent-Jours*, *Napoléon Seul Contre Tous*, Julliard, 1815-3-2. Dominique de Villepin.Les Cent-Jours, ou, L'esprit de sacrifice, Perrin, 2001.

②Caliskan, A. *Detecting and Mitigating Bias in Natural Language Processing*,https://www.brookings.edu/research/detecting-and-mitigating-bias-in-natural-language-processing/2021.

③乔·提迪:《 ChatGPT 的语言偏见: 令非英语使用者"掉队"的三种方式》,https://www.bbc.com/zhongwen/simp/science-67270190.

④Elliott, K. C. *A Tapestry of Values: An Introduction to Values in Science*, Oxford University Press ,2017,pp.12-13.

⑤Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. "Seeing Black: Race, Crime, and Visual Processing", *Journal of Personality and Social Psychology*, 2004,87(6):876.

⑥Tversky, A., & Kahneman, D. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment",*Psychological Review*, 1983,90(4):293-315.

⑦Crawford, K. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*,Yale University Press. 2021, pp.8.

⑧中共中央文献研究室:《习近平关于科技创新论述摘编》,中央文献出版社 2016 年版,第 29 页。

一般规律，为中国的人工智能产品安上“中国心”，这才是中国人工智能的发展方向。

首先，人工智能的机器学习要有“中国心”。人工智能的本质是通过对人类活动的机器学习，模仿人脑的决策能力，从而使机器能够像人一样感知、识别、思考、学习和协作。机器学习有三个基本要素即数据、模型和算法，由于这三个要素都是在特定的社会关系体系中被创建、发展起来的，因此，机器学习不可避免地具有二重性：既是技术的、普遍的，又是文化的、特殊的。机器学习二重性的表达是不均衡分布的，绝大多数日常情况下，技术的、普遍的特性是显性的，文化的、特殊的特性是隐性的，但在极少数非日常的极端情况下，文化的、特殊的特性则会发挥决定性的作用，决定了机器学习之非此即彼的结局。例如，面对“电车难题”，人工智能的决策必然会彰显不同文化传统的根本价值观分野：自由主义的欧美会选择扔骰子，把选择权留给随机的命运；尊崇国家、集体的中国则会基于传统伦理进行选择，把选择权留给主体的道德修养。在此意义上，中国生成式人工智能的发展必须强化高质量中文语料库的建设、算法的伦理嵌入、决策的伦理指导等维度，才能最大限度将“中国心”植入人工智能之中。如果我们今天不能给机器学习安放一颗澎湃的“中国心”，怎么可能指望人工智能在危机时刻做出符合中国人根本利益的决策呢？

其次，人工智能的产品研发要有“中国心”。在科学技术的发展史上，确实存在过为科技而科技的纯真年代，但这种纯真年代不仅短暂而且已经被超越。今天，科学技术的发展已经和社会经济发展深度融合，市场化成为推动绝大部分新技术发展的最强劲动力。例如，太阳能和风能技术都源于欧美，但欧美长期把这两种清洁能源技术作为对发展中国家进行小院高墙式技术封锁的独门秘籍，反倒是中国以经世济用的情怀、以市场化的方式大力发展这两种清洁能源技术，最终让旧日“堂前燕”“飞入寻常百姓家”，成为造福人类的“白菜价”产品，并反向输出欧美。人工智能作为技术产品同样只有在市场化中才能获得源源不断的发展动力，否则就会沦为猎奇性的技术奇观，随着人们的关注度的下降而黯然退场。人工智能产品研发的“中国心”，一是指要突出经世致用的实用性，考虑产品研发要能便利生活、造福社会、促进文明进步，而不仅仅是具有新奇性的时髦电子玩具；二是指适应中国国情，会写美式笑话的 Chat GPT 在美国能够成功，但是中国需要的则是会讲中国段子的人工智能；三是指坚守人民立场，坚持以人民为中心、站稳人民立场是中国共产党领导科技事业发展的根本价值取向，因此，人工智能产品研发既要能够市场化，也要有利于满足广大人民群众的现实需要及其自由全面发展，不能让人工智能产品造成新的“奴役”。

最后，人工智能的社会监管要有“中国心”。2023 年 11 月，Chat GPT 所属 Open AI 的董事会发生了一场震惊科技圈的“宫斗大戏”①，尽管这场“宫斗大戏”很快就戏剧性结束，但再次暴露出人工智能领域的一个关乎人类未来的根本问题：人工智能，是商业化，还是安全性？对这个问题的不同回答，决定了不同的社会监管策略②。美国的回答显然是商业化。所以，我们看到，美国基于其新自由主义的政治经济立场，一向对人工智能技术的发展和其带来的挑战持自由放任态度，迄今为止都不急于对人工智能研发进行广泛监管，而以“无需批准式监管”和“轻触式”审慎监管放任人工智能新技术自由发展，以期充分释放人工智能的技术潜力，尽可能扩大美国在人工智能领域的技术霸权。欧洲的回答明显是安全性。因此，2023 年 12 月 8 日，欧洲议会、欧盟成员国和欧盟委员会三方经过漫长谈判，最终就《人工智能法案》达成协议，这一法案是全球首部人工智能领域的全面监管法规，其目的在于“促进以人为本和值得信赖的人工智能的采用，并确保在欧盟内高度保护健康、安全、基本权利、民主和法治以及环境免受人工智能系统的有害影响，同时支持创新”。我国学术界对人工智能立法进行了一系列有益的探索③，我国政府出台一系列行政法

①根据中国新闻网的报道，Open AI 于 2023 年 11 月 17 日宣布阿尔特曼不再担任公司 CEO 并将离开公司。随后，在 Open AI 约 770 名员工中，有超过 740 人连夜签署联名信，以集体辞职相要挟，要求阿尔特曼回归。微软作为投资超过 130 亿美元的最大投资方出面调停。最终，阿尔特曼于 11 月 22 日重回 CEO 之位，并重组董事会。

②和军，杨慧：《Chat GPT 类生成式人工智能监管的国际比较与借鉴》，《湖南科技大学学报（社会科学版）》2023 年第 6 期。

③王荣余：《在“功利”与“道义”之间：中国人工智能立法的科学性探析》，《西南交通大学学报（社会科学版）》2022 年第 2 期。

规,以负责任的大国担当推动人工智能科学监管,努力促进人工智能健康、规范、高质量发展,使科技创新和人工智能服务人民,助推经济社会高质量发展和中国式现代化建设,辩证地提出对待人工智能要“发展和安全并重”,鼓励其发展,规范其服务,明确其责任①,并及时将人工智能研究纳入伦理审查的范畴②。人工智能是一场全球性的技术变革,因此,中国也需要加大在全球人工智能治理中的话语权,中央网信办发布的《全球人工智能治理倡议》明确指出,“人工智能治理攸关全人类命运,是世界各国面临的共同课题”,因此,各国之间必须要凝聚共识,协同治理,“促进人工智能技术造福于人类,推动构建人类命运共同体”③。2023 年 11 月 1 日,中、美、英等 28 国和欧盟共同签署《布莱切利宣言》,这是全球第一份人工智能治理的国际性声明,既代表了国际共识,也代表了中国在人工智能治理领域的影响力。可以自豪地说,一种具有“中国心”的人工智能发展和管理模式已经基本形成,其世界历史意义和人类命运共同体价值终将被实践证明!

Humans Have Heart, But not Machines: Rethinking the Development of Artificial Intelligence from the Perspective of Mind-Machine Relationship

ZHANG Liang & YU Quanlin

(Department of Philosophy, Nanjing University, Nanjing 210008, China)

Abstract: The rapid development of artificial intelligence seems to have blurred the boundary between mind and machine, but there is still an insurmountable gap between the two. The subjective nature of experience makes it difficult to completely separate mind from body. Physicalist reductionism cannot cope with mental activities with cultural characteristics. Although from a philosophical perspective, people cannot create consciousness with the help of machines. However, in real-world applications, artificial intelligence exhibits “biases” that are unique to humans. It is the imperfection of database and the inequality in algorithm design that have caused the biases hidden in cognitive subjects in the past to surface. The development of artificial intelligence cannot rely solely on imitation. In the new round of technological craze, we must equip artificial intelligence products with a “Chinese heart”.

Key words: artificial intelligence; mind-body issues; “Chinese heart”

(责任校对　王小飞)

①国家互联网信息办公室等:《生成式人工智能服务管理暂行办法》,中国网信网,http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm。

②科技部等:《关于印发〈科技伦理审查办法(试行)〉的通知》,中国政府网,https://www.gov.cn/zhengce/zhengceku/202310/content_6908045.htm。

③中央网络安全和信息化委员会办公室:《全球人工智能治理倡议》,中国网信网,http://www.cac.gov.cn/2023-10/18/c_1699291032884978.htm。