



doi: 10.3969/j.issn.1674-8425(s).2021.12.007

逻辑与科学方法论

主持人语:

中国逻辑学会名誉会长 邹崇理 研究员

人工智能是人类认知主体的延伸,而虚拟世界是人类认知客体的延伸。对人类及现实世界的摹本,即延伸的主体和客体及其关系进行探讨是当今哲学思考的重大问题。元宇宙是科学家们设想的一类与现实世界并行的虚拟世界。《人工智能、缸中之脑与虚拟人生——对元宇宙问题的跨学科研究》一文围绕这个思想界的重大话题进行讨论和分析,这样的工作是有意义的。

蒯因是哲学界和逻辑学界的大师,他对说谎者悖论消除“自我指称”的传统方案提出挑战。虽然蒯因具有较大的影响力,但《如何理解蒯因的说谎者》一文的作者对此仍然提出自己的看法,认为蒯因的说谎者悖论思考也暗含了自我指称的特征,因而蒯因的挑战也是可以质疑的。

人工智能、缸中之脑与虚拟人生

——对元宇宙问题的跨学科研究

张昌盛^{1,2}

(1. 中国社会科学院大学 哲学院, 北京 102488;

2. 中国社会科学院 哲学研究所, 北京 100732)

摘要:元宇宙是科学家们设想的一类与现实世界并行的虚拟世界。为了有助于科学认识元宇宙,简要梳理了元宇宙的思想史源流以及目前的发展现状;然后借助哲学家普特南的“缸中之脑”的思想实验问题的讨论,从科学和哲学两个角度对元宇宙问题进行理论分析,主要涉及大脑、意识与智能等的关系的探讨,同时涉及对强人工智能科技的可能性的讨论。在此基础上,从脑机接口和人工智能的技术条件和哲学基础两方面分析未来元宇宙可能实现的形态。最后,分析元宇宙对未来人类生活和文明的导向性和可能的负面影响。

关键词:元宇宙;“缸中之脑”;人工智能;脑机接口;导向性

中图分类号:B152;B842.1;B81 **文献标识码:**A **文章编号:**1674-8425(2021)12-0052-12

近年来,元宇宙(Metaverse)成为信息科技领域的一个新的技术投资风口,也是近期学界关注的热点之一。元宇宙简单地说是借助数字科技构造虚拟世界的技术,具体来说主要是通过整合脑机接口、互联网、大数据、VR、AR、人工智能、区块链等技术来建构虚拟世界的科技领域^①。

收稿日期:2021-11-06

基金项目:中国社会科学院创新工程项目“人工智能重大哲学问题研究”(2021ZXSCXB04)

作者简介:张昌盛,副研究员,博士,主要从事科学哲学研究。

本文引用格式:张昌盛. 人工智能、缸中之脑与虚拟人生——对元宇宙问题的跨学科研究[J]. 重庆理工大学学报(社会科学), 2021(12): 52-63.

Citation format:ZHANG Changsheng. Artificial intelligence, brain in a vat and virtual life——An interdisciplinary study of meta verse[J]. Journal of Chongqing University of Technology(Social Science), 2021(12): 52-63.

① 参见许英博、陈俊云.《元宇宙:人类的数字化生存,进入雏形探索期》,中信证券研究部,2021年9月12日。

元宇宙这个名称最早出现在20世纪90年代的科幻小说《雪崩》中,而元宇宙的理念实质上是肇始于普特南的“缸中之脑”的思想实验,后续的电影《黑客帝国》以及新近的电影《头号玩家》以科幻的方式体现了类似的观念。

按照现在各大科技企业对元宇宙的建设目标的基本叙述,元宇宙被理解为是未来的互联网的终极形态,是一种与现实物理世界并行的虚拟世界,但又部分地与现实世界深度融合。元宇宙的“元”含有超越的意思,不仅可以超越现实的空间区隔,而且能够构造超越于现实世界的虚拟景观和场景,提供超越现实独特体验。

元宇宙在未来会深刻而全面地改变人类生活的方式,以及整个社会的运作模式和历史走向。元宇宙将提供给人类极其丰富、新奇而又逼真的体验,又具有超越空间的便利,未来人类的生活、工作乃至生产活动将有很大一部分时间是在虚拟世界中进行的。随之而来,未来人类的生活形态也将发生根本改变。进而,人类关于虚拟世界与物理世界乃至虚拟生活与真实生活等的观念,也会随着元宇宙的到来而发生根本的改变。如果从当前主流的自然主义立场设想,未来元宇宙的极致形态将会是“黑客帝国”的“矩阵”(The Matrix),很多人将生活在数字化的虚拟世界中。当然,这种极致形态的“虚拟人生”会不会出现,需要我们结合科学进展和哲学思想进行深入分析和预判。

因此,需要结合相关的科技与哲学,从跨学科的视野对元宇宙的未来趋势、社会价值以及战略意义进行理论本质的深度思考 and 实践趋势的综合预判,为我们的元宇宙科技战略提供思想支持。

一、元宇宙的理念渊源和前景

元宇宙的名称虽然出现得很晚,但相关理念早在人类思想的早期就已经触及。什么是真实世界?在不同的神话框架和思想体系中有不同的答案。在未来元宇宙技术高度发展的时候,随着虚拟世界与现实世界的交融共生,虚拟人生在人类生活中扮演越来越重的角色时,元宇宙将会重新激发人们对真实与虚幻的深度思考。

(一) 什么是真实的世界?

回看历来人类对于世界的观念,最主要的两个问题是起源和结构/性质。关于世界的起源问题,中国的神话中是盘古开天辟地,这是一种发生学角度的神话叙事。在传统西方主流文化中,世界起源也是神创论,基督教的神创论是最为主流的一种版本。中国传统的世界观念虽然也会涉及到天上与人间、人间与地府等二元区分,但主要是在神话和宗教信仰中,而在社会主流的世俗的思想、哲学中反而着重强调现实世界,更像是一种一元论世界观。在西方神学和哲学中,对世界的描述往往是在一种理想与现实的二元结构中展开的。早在柏拉图的理念论中,认为真实的世界是一种由作为本源的理念构成的世界,现实世界反而是这个本源性的理念世界的模拟产物;基督教神学从神创论的角度,认为耶和华的伊甸园才是更为理想的世界,而现实世界则是神为人类创造的;莱布尼兹的单子论延续了神创论的思维,认为宇宙是神的造物,而现实世界则是神所能创造的可能世界中最好的一个世界。

近年来,思想界有一种虚拟宇宙论的学说,认为我们人类、我们所在的世界乃至所处的宇宙都是一种更高级的智慧在计算机中虚拟建构的,这是神创论的一种现代版本。

而在元宇宙的设想中,人类自己成为造物主,创造出足以与物理世界匹敌的虚拟世界。如果从自然主义的角度看,只有物理世界才是真实的世界,关于可能世界的设想都是基于现实世界这个原版的想象

和发挥。但是,如果从观念论或者宗教/神话看,则那种理念的世界才是纯粹的、完美的,现实世界反而是拙劣的、粗糙的模仿版本。

而元宇宙的设想,则超出了以上这两种模式,它既有模仿现实的因素,但又有超越现实的可能世界的因素,是与现实世界部分融合但又截然不同的新的世界。元宇宙会是一个复数概念,会有模仿现实的版本,也会有基于想象、超越现实世界规则桎梏的新奇版本,后者能够带给人类在现实世界中无法实现的全新体验。这种新奇、创造和超越的因素,是元宇宙的魅力所在。可以说,通过元宇宙的建构,人类终于摆脱了现实世界的模型,可以充分发挥自己的想象力,塑造理想化的世界。在这种可以创造现实世界缺乏的全新体验的特定意义上,可以说人类就是新的神,虚拟世界与现实世界可以并驾齐驱。

(二) 元宇宙的前景展望

有些研究按照虚拟与现实的关系,把未来元宇宙的建构依据分为3个阶段:数字孪生、虚拟原生、虚实融生^①。第一个阶段的数字孪生就是用数字技术仿真现实世界的某些方面建立平行的数字化模型,主要是模拟、复制现实世界,以此虚拟的数字化孪生模型作为辅助管理和改进现实世界的技术工具。第二阶段是模拟出具有高度逼真性的虚拟世界,部分内容模仿现实世界,另外一部分内容又是原创性的、超越现实世界的,与现实世界并行。第三阶段是虚拟世界足够宏大和发达,把现实世界作为一部分融入了元宇宙之中,虚实界限逐渐模糊,虚拟世界和现实世界融合为一体。

1. 虚拟游戏

元宇宙的最成熟版本就是现在各大互联网公司的虚拟游戏。借助于VR设备和配套的全体外传感装置,游戏玩家可以在游戏中有3D效果的游戏场景体验。这种虚拟技术主要是通过VR眼镜模拟出沉浸式的视觉效果,借助于AR设备,可以产生出增强的触觉体感体验;但由于这种非侵入式的设备本身的局限性,无法具有超高逼真度的真实体验感。

2. 数字孪生城市

数字孪生城市就是用智能化数字技术建构一个与现实城市平行的数字化版本,可以看作是之前设想的智慧城市的升级版,也是未来元宇宙的一种初级版。通过在一个城市中的整体规划,借助人工智能技术,建立一个超大规模的系统、高度发达的基建以及硬件配套,实现物理城市与虚拟现实的虚实融合,使用户产生身临其境的互动体验。

深圳市已经把数字孪生城市写进政府新的城市规划中^[1]。现在深圳市与华为云联手共建了一个基于5G、云、AI的鹏城智能体,深圳的万万千千个生活场景正在被重构,水文气象可以被感知、机场港口会思考、车流可以被读懂。借助于强大的数字技术尤其是人工智能,深圳已经在数字孪生城市和真实物理城市之间实现了虚实映射、虚实交互。

3. 英伟达的元宇宙:数字孪生地球与数字孪生工厂

近期,英伟达公司也宣布了自己的元宇宙技术Omniverse(用于3D工作流程的虚拟世界模拟与写作平台),在11月9日GTC大会(GPU Technology Conference)的主题演讲中,黄仁勋宣称:“我们现在已经拥有了能够创建新的3D世界,或者对真实物理世界进行建模的技术。”^[2]“Omniverse将在仓储、车间与厂房、物理及生物系统、5G边缘场景、机器人、自动驾驶汽车乃至化身的数字孪生等层面提供模拟功能。”^[2]这种目标之一是建立孪生数字地球“地球2号”(Earth Two),通过对地球气候进行足够详细的建

① 参见王儒西、向安玲主笔:《2020—2021年元宇宙发展研究报告》,清华大学新媒体中心,2021年9月16日。

模,以实现全球气候变化的模拟、监控和预测,目标是准确预测未来10年、20年或30年的气候,从而为人类在农业、工业、交通尤其是环境治理、气候问题解决等方面提供气候背景参考。这种建模是通过物理机器学习模型来解决气候长期预测中的种种复杂问题。

其次,他们认为这项技术也可以用于制造业。在黄仁勋的设想中,未来可以在元宇宙中设计数字产品如汽车等,然后才在现实世界中去完成设计。同样可以在英伟达的元宇宙(Omniverse)中建立虚拟工厂,并使用虚拟机器人进行运营,虚拟工厂和机器人是其物理复制品的数字孪生,实体版是数字版的复制品。建立与实体的工厂平行的数字孪生工厂,生产流水线和工艺可以首先在元宇宙工厂中进行设计、测试和运作,通过在虚拟世界中的生产,实现对生产流水线各个环节的生产工艺和管理的迭代改进、不断优化,实现生产效率和质量并重的最优解。再通过与智能化的实体生产线的并网,把这种虚拟生产转变为实体工厂生产线的生产过程。这种虚实融合的智能化生产能极大地提高生产效率并降低相应的投资成本。

4. 未来虚拟世界

未来虚拟世界的主流方向应该是创造原生虚拟的平行数字世界,不断地开拓虚拟世界的边界,最终趋向于把现实世界容纳于元宇宙的世界,虚实交融,实现所谓虚实融生的元宇宙。

二、元宇宙的思想来源、理论假说和哲学基础

与元宇宙相关的思想渊源早在几千年前就产生了,但直接思想来源应该是哲学家普特南的“缸中之脑”的思想实验。普特南的这个思想实验涉及到对人的大脑/意识等的基本看法以及对人工智能的基本观念,直指哲学的核心问题。而且这个思想实验设想了虚拟世界或者说今天所说的元宇宙的极致形态。因此,普特南的“缸中之脑”思想实验对元宇宙的深度技术分析和哲学反思都是一个很好的思想范例。

(一) 普特南的“缸中之脑”

元宇宙的思想来源可以追溯到哲学家普特南的“缸中之脑”的思想实验。在1981年出版的专著《理性、真理与历史》中,普特南提出一个“缸中之脑”的思想实验^{[3]7-9},想象一个人的大脑被某些恶意的科学家们用手术从他的身体中分离出来,放入一个充满营养液的缸中,缸中的营养液和设备可以让这个大脑维持长期生存;并把他大脑的神经末梢和一台超级计算机相连,这台计算机非常聪明,可以精准地提供各种电子脉冲信号给神经末端,让大脑主人保持一切完整的幻觉,让他觉得依然处于真实的世界中,一切经验和行为都与原先没有区别。普特南进一步设想,这个人的所有器官都泡在缸中,而外部世界是一台大的自动机。在普特南看来,在理想实验条件下,上述的缸中人无法区分自己的脑是颅中脑还是缸中脑,自己是在真实世界还是虚拟世界中。

普特南的“缸中之脑”的原理相当于通过中间的计算机设备把人脑和虚拟世界关联起来,这套装置通过电信号,实现双向的信息通讯。一方面,计算机可以提供类似于人面对世界时的体验和认知的模拟信号;另一方面,人类大脑的命令可以通过脑电信号传输给计算机,通过后者控制传动装置进行活动,完成任务。在这个过程中,计算机反馈回来的虚拟信号在大脑中枢得到处理,形成类似于人在真实世界中操作身体或工具进行的活动时的体验效果。由于计算机及虚拟计算产生的模拟信号过于逼真,乃至大脑无法区分自己是处于真实世界还是虚拟世界。

普特南的这个思想实验启发了后续关于虚拟现实的思想灵感。上述的小说《雪崩》,电影《黑客帝

国》《盗梦空间》及《头号玩家》等都延续了这个思想实验的灵感。近年来,随着认知科学和人工智能技术的发展,普特南的这个思想实验不再停留在科幻作品阶段,而是成为人工智能和信息科技等科技探索的前沿主题,现在的“脑机接口”、VR、AR、元宇宙等都与此相关。

元宇宙的设想就是类似普特南的这个思想实验,通过虚拟和增强技术,让元宇宙的用户产生一种具有“身临其境”效果的沉浸式体验,类似于把人脑和虚拟世界“并网”。不仅用户的视觉、触觉及体感等无异于现实世界,而且用户的意图、指令等可以在虚拟世界中通过代表虚拟自我的ID执行,产生一种自身自由驰骋于虚拟世界中的效果。当然,伴随着这种具有高度逼真性的体验,元宇宙对用户的反向影响也是深度的,不仅影响用户的身体体验,而且还可以反向影响用户的意识活动、认知、记忆等。

(二)“缸中之脑”隐含的理论假说

普特南的“缸中之脑”的思想实验涉及几个基本假设:

第1个假设是基于脑科学的假设,即人类的一切意识、智能活动及关于世界的经验都是基于大脑神经中枢对于输入信号的信息加工、处理而在大脑中形成的,也就是人类感知到的世界、他人及自我,都是大脑基于生物电信号传入的信息而建构的模型。这种信号是生物电信号以及化学信号,依托的平台是人脑的神经元组织及其活动,这个信息处理过程对应的是脑神经元组织内的生物电信号、化学信号借助于神经递质而在神经元间的传递、处理、储存等过程。也就是说,在理论上,人类的一切认知、意识活动和行为都对应于大脑神经元组织中的动态变化着的生物电信号集中的某些特定的信号,因此可以由大脑生物电信号及其变化来识别人类的不同意识活动状态,包括对应思维、意图、语言及行为的特征意识状态。由此,可以得出一个推论,掌握了大脑的电信号,就等于掌握了人类的智能和意识活动。

第2个假设是世界中的一切事物包括人的活动及体验可以归结为广义的物理的或生物的算法,而且可以被形式化描述。还有一个数字化假设,即这些关于世界、生命的形式化表述是可以数字化或者近似地可以数字化的。因为计算机的二进制决定了数字化表达是离散的信息,而自然中的很多事物虽然可以形式化,但在这种数学描述中却仍然是连续的。可数字化表达意味着对原有的信息可以用离散的数据进行近似表达。

第3个假设是人的智能可以看作是基于脑神经元组织的广义的生物算法。如第1条假设所主张的,人的意识、认知等都可以看作是大脑神经元组织活动的产物,而后的活动可以看作是基于生物电信号和化学信号的计算。这种特殊的生物算法是基于脑神经元组织的特殊平台的,这种智能乃至意识跟神经元组织的特殊结构、规模及运作方式相关。

第4个假设是假设类脑的人工智能可以实现。如果假设这种大脑运作的生物算法可以被人工智能完全或者部分模拟,那么就可以建立类脑人工智能。即:通过类脑的人工智能来实现对世界及人的数字化表述,进而可以通过类脑算法实现用类似人类的视角去认知世界,并且也可以虚构类似的认知数据,从而可以给人脑提供类似于人对世界的真实感知的信号。

第5个假设是基于前4条假设的一个技术条件假设,即我们可以对这些由人脑提供的电信号和由计算机所提供的数字信号实施双向编码和解码,从而实现基于信号的双向传输的脑-机通讯。

要实现这个思想实验,需要满足上述5条假设,尤其是要满足第5条假设。第5条假设涉及两个方面:一是预设了脑科学足够成熟而且发达,可以通过设备充分提取大脑的电信号,并且可以解码这些信号,尔后由计算机重新编码形成操作外在物理世界中的设备或虚拟世界中的对象的指令,完成相应的活

动或任务;二是可以对由计算机传向大脑的信号进行反向解码,生成可以为大脑所处理的生物电信号,从而让大脑可以认知外在的物理世界或者虚拟世界。通过这种认知和操作的配合,可以实现让大脑处在真实世界中的幻觉的效果。

(三)“缸中之脑”蕴含的哲学问题

上述5条假设涉及到一系列相关的哲学假设:

上述第1条假设隐含了意识问题的一种自然主义、生物主义的哲学假设,这条假设是一种比较强的意识理论,即一切意识活动都是生物神经元组织的活动的产物,或者说人的意识活动是一种生物平台即大脑神经元组织的信息处理过程。

这种意识假说对应的意识理论有几种;其中,最强的立场是生物主义的还原论,即意识现象可以还原为构成神经元组织的每一个神经元的功能及性质的集合;较弱的立场是生物主义的非还原主义,即意识是由神经元组织的复杂结构及非线性系统的突现的整体属性,不能还原到单个神经元的活动的线性组合。

还有一种弱的二元论也可以兼容脑科学的工作假说。这种二元论的学说把大脑神经元组织比作集成电路硬件模块,把意识比作电流,二者合在一起才产生智能和认知等功能。因为脑科学所需要的实质性的工作假说仅仅是认为人的意识活动可以一一对应于大脑神经元组织的生物电信号变化(借助于这种严格的映射关系,可以对后者进行提取,进而借由解码技术来识别前者的内容;再通过计算机编码输出人的意识内容尤其是意志指令)。至于这些意识活动是由神经元组织的活动产生的,还是意识的产生另有来源,神经元组织只是作为产生意识的必要条件之一,对于脑科学的工作假说和方法而言并无差别。

上述第2条的可形式化的假设,首先,隐含了一种类似毕达哥拉斯主义的宇宙观和生命观,即预设一切事物包括生命的本质特征都在于其形式,一切事物的运作过程也包含着某种形式的机制;其次,这些事物及过程的存在依赖于一种确定的运作机制,可称之为物理算法或生物算法;再次,这些事物及过程的形式方面的特征还可以用数学语言来描述,类似于伽利略的自然数学化,而且生命本身也可数学化,乃至人的心理状态乃至整个意识也可以被数学化,即一切事物的特性可以被形式化表达;最后,还需要假设,这些事物及过程可以被数字化地描述,并且可以用计算机的算法来表达。

上述第3条假设是一种生物自然主义的哲学立场,即认为人的智能乃至意识是人脑的产物,具体而言是大脑神经元组织的活动产生的现象。这种立场认为智能乃至意识是基于自然因素和规律的产物,但需要特定的生物条件;只有像人类大脑神经元组织这样的结构足够复杂、规模足够庞大的生物系统才能产生智能乃至意识。这种立场有别于物理主义,后者认为智能乃至意识实现条件是具有足够的复杂度和规模的物理组织,生物神经元组织并非智能及意识的必要条件,而基芯片加算法也能实现类人智能;这种立场也有别于二元论,后者认为智能及意识是心灵实体或属性的产物,仅凭神经元组织不足以产生类似于人的智能,遑论意识了。

当然,这里同样对应有一种弱化立场,即认为人的一切智能和认知等活动跟这种生物平台的特殊算法及计算相关,但并不假设这种大脑神经元组织的活动是实现类人智能的充分条件;同样,可以承认人类的一切意识活动一直伴随着这种生物神经元组织层面的计算过程,但并不主张意识产生于这种生物算法及计算过程。

上述第4条假设是关于人工智能的假设。在人工智能哲学中,对应着联结主义的仿生的人工智能范式,即计算机的算法可以借鉴乃至模拟人脑的生物算法,从而实现类脑智能;按照普特南的“缸中之脑”

的设想的类脑智能应该是强人工智能,因为该思想实验中的计算机用数字模拟的外界过于逼真,以至于缸中的大脑无法分辨真假。

对于元宇宙技术而言,也是把具有高度逼真性的体验作为追求的主要目标之一,但并不一定需要达到“缸中之脑”的那种逼真程度,因此也可以预设弱人工智能立场,即类脑智能虽然无法完全像人脑一样认知外界以及模拟外界,但可以通过技术进步逐渐达到逼真度很高的模拟结果。

上述第5条假设是一个基于脑科学假说的技术性假设。如果通过脑机接口可以实现大脑的生物信号和元宇宙的数字信号的高度保真的有效通讯,那就意味着大脑生物网络和数字网络之间并没有不可跨越的天堑,某种程度上可以把关于智能乃至意识现象的生物现象还原为数字网络的物理过程,这其实预设了生物自然主义立场可以还原为物理主义立场的潜在可能。

三、元宇宙的技术基础及哲学问题

元宇宙的核心技术包括虚拟技术如人工智能、脑机接口、VR、AR、5G、云计算、物联网、区块链等,但最核心的信息技术部分则是脑机接口这样的人脑-机器之间信息通道以及人工智能作为构造虚拟世界内容的引擎和主要生产力。因此,我们这里需要结合科学和哲学考察元宇宙的这两部分技术基础问题。

按照物理主义或生物主义的假设,要实现元宇宙的沉浸式体验,尤其是要实现类似普特南的“缸中之脑”式的身临其境的虚拟体验,核心技术就是对大脑的输入、输出信号问题的解决;概而言之,就是要以虚拟的信号输入代替人的视觉、听觉、触觉等感官,以机器信息通道和外接的设备取代人的身体对外界的操作和控制行为;并且,这种对大脑的信息输入和输出要形成双向交流的闭环,这样信息输出产生的操控会让游戏参与者在游戏中产生一种自身投入其中的代入感,同时信息的输入和反馈,会让游戏参与者产生一种感同身受,这两种体验的同步、结合会使游戏参与者产生一种身处真实世界中的幻觉。在元宇宙的技术体系中,这种解决信息输入、输出的双向传输通道的关键技术是脑机接口,而向大脑提供虚拟世界内容的关键在于人工智能的生产活动。

(一) 脑机接口

对于元宇宙而言,可以通过VR、AR以及脑机接口等不同方式实现与虚拟世界的交互,但要实现一种“身临其境”的沉浸式体验,最彻底、最高效的方式是通过脑机接口把大脑和电脑直接关联起来。

脑机接口(Brain Computer Interface)^①是指在大脑和外部设备之间建立的信息通讯通道。在脑机接口的设想中,预设了意识是大脑神经元的活动的结果,意识的内容可以由大脑电信号来解码。简单来说是通过植入大脑皮层的芯片,可以实现对大脑电信号的读取;通过对大脑信息的解码、编码,可以控制对大脑信号的输出、输入;因此,可以通过脑机接口读取大脑电信号,实现大脑信息的读取、输出、复制、下载,也可以反向地输入、上传、修改,甚至可以改变大脑的记忆、思维和认知,进而实现信号在大脑和电脑之间的双向闭环传输,这样就可以在大脑和外部设备之间建立起一种直接的通讯和控制通道,从而实现人通过大脑的意识活动与外部设备的交互作用。

脑机接口技术与人工智能及智能控制等技术结合,可以实现外部世界与虚拟世界的互动。具体而言,脑机接口的技术流程可分为4个环节:信息采集、信息处理、设备控制和信息反馈。第1个信息采集

^① 参见量子位微信自媒体发布的《量子位脑机接口白皮书》。

环节中,脑机接口通过植入大脑皮层中的微电极采集脑部神经元活动的信号;第2个信息处理环节中,对采集的信号进行解码,再对这些信号进行编码,转化为机器能读懂的指令信号;第3个设备控制环节中,通过计算机发出的指令对外部智能设备进行控制;第4个信息反馈环节是对控制过程以信号形式向大脑反馈。通过这样一个闭环的双向信息通讯过程,完成人机互动过程。

目前的脑机接口中有两个最大的困难:一是信号处理环节中对神经元信息的解码过程,二是从外界对大脑的信息反馈如何实现。在脑机接口这一部分先讨论第1个问题,第2个问题关系到人工智能技术,将放到下一部分去讨论。

信息处理环节对于脑机接口技术的实现非常关键。基于人工智能的解码和编码,使用者的设备能读取大脑的“想法”并传输给设备。使用者可以通过“意念”在对外在的设备进行控制,从而实现大脑的意图;也可以通过反向的编码、解码,实现对外部对象或活动产生的信号向大脑的传输,从而实现大脑的指令信息输出和反馈信号输入大脑的信息闭环。这种双向互动还有一个基本的前提:大脑可以通过视听感官或者传感器和计算机等中间设备“感知”外部物理对象或者虚拟对象,才能展开输出、操控、反馈、调整等一系列的互动。

脑机接口技术的目标是建立人脑和电脑之间的双向信息通道,对应于大脑的输入和输出两个任务,脑机接口的设想基于两个方面的技术前提:

第1个方面是读取和输出大脑信息:(1)初级任务:意念控制外物;(2)高级任务:解读大脑的意识如思维、情绪等。

如前所述,这里的基本理论假设是大脑的一切意识活动都对应于神经元的活动,越是能精确地监控各个神经元的活动,越能精准地把握意识活动的内容。但这里的理论难点在于,意识活动并不是与单一的大脑神经元区域的活动对应,大脑神经元的多线程是并行的、复杂的,同一个意识活动可能对应多处神经元的协作;同时,一处的神经元会参与到多种意识活动中,因此即使把握了每一个神经元的活动,也不意味着可以把握大脑神经元活动与意识内容的精确对应关系。尽管如此,但总体而言,越是能够提高对大脑神经元的监控精度,越是能区分不同的意识内容。这里的技术难点在于如何尽量高精度地监控大脑神经元的信号变化,并且能够成功解码。

要想高精度地提取大脑神经元信号,就需要在大脑皮层中植入大量的微型电极。通过植入微电极,通过对大脑皮层里的各个区域的局部场电位的测量,就可以实现监控大脑神经元活动状态的任务。理论上,植入的微电极越多,分布区域越广,就可以对大脑皮层的神经元活动实现更全面和精准的监控。按照目前神经科学的认知,人的大脑大约有860亿个神经元。目前,脑机接口获取大脑信息的水平还比较低。2020年,马斯克的脑机接口公司(Neuralink)进行的脑机接口实验^[4],植入小猪大脑中的脑机接口有1024个信息采集通道,即植入了1024个微电极。打个比方,这就相当于在一个有860亿人的广场上,有1024个麦克风来采集人群发出的声音,最后采集到的是那些最大音量的声音,对不同人的声音的分辨率还很低。

从终极理想而言,最好能实现对每一个大脑神经元大脑的电信号的监控。理论上,人类大脑皮层神经元组织可以有860亿个微电极;在比较充分的脑电信号采集的基础上,通过未来比较成熟的脑科学的理论建立的大脑神经元组织模型来分析这些数据,可以实现对大脑神经元信息处理机制的比较完整的把握。

而解码的关键是建立检测到的大脑神经元信号与大脑的意识内容之间的对应关系,可以凭神经元信

号解读出大脑意识的内容。目前所能采取的办法还很原始:通过监控大脑神经元信号与测试者的思维、语言和行为活动的对应关系,归纳式地收集解码数据库。因此,只有那些纳入解码信息库的大脑神经元信号才能被解读。未来要建立高精度的、全面的神经元信号解码信息库,还需要相应的脑机接口的神经元信号读取技术取得长足的进展。

第2个方面是对大脑的信息输入:把数字信息转变为脑生物电信号,输入大脑。包括机器传感器代替人的感官传感器输入信息。这种写入信息的技术比之前的脑电信号提取更困难:技术上的困难在于植入足够的微电极和微芯片。更大的困难在于脑科学对大脑的认知还很初级,对于输入大脑神经元组织中的电信号的信号类别、信息的传递通路、信息处理方式、信息处理的功能区域及不同区域间的分工协作等都缺乏足够全面和系统的了解,因此难以实现对数字信号的正确编码和精准输入。另外,某些信号的输入是否意味着对相应内容的输入,使主体可以获得相应的信息、知识,还停留在假说层面,缺乏实验的验证。

这里所设想的信息上载、知识输入、记忆修改等,依然是一种基于物理主义的假想,即假设意识等同于大脑神经元的活动,认知和体验通过神经元电信号为载体。但事实上,大脑的结构及活动非常复杂,而意识尤其是内容是否可以等同于神经元活动以及相应的电信号传输及处理,目前为止尚未得到实验的验证。因此,人的认知、记忆、知识乃至情感等的精确操作是否可行?脑机接口的技术边界在哪里?诸如此类的很多问题都需要等待未来真正成熟的脑科学来回答。

目前,所设想的脑机接口技术应用主要分3类:第1种主要应用是人机交互领域;第2种主要应用是通过监控脑电信号来进行评估治疗;第3种应用是借鉴科研。

第1种获得进展的领域是通过对大脑运动皮质部分的监控实现对机体外骨骼的控制。具体而言,因为运动皮质的可塑性,可以先通过反复训练意识的运动想象,捕捉相应的脑电信号,建立意识的运动想象和脑电信号之间的稳定关系;实际应用时通过运动皮质做运动想象,再通过脑机接口对相应的脑电信号的监控、解码,再重新编码、输出指令来控制外部设备。这个领域的应用前景很广,包括意念控制鼠标、机械臂、机器人、玩游戏等。这种技术人机交互技术尤其适用于医学治疗:协助残疾人通过机械肢来控制身体。

第2种获得进展的领域是通过对大脑体觉皮质的监控、解码为基础,修复视觉、听觉等。例如人工耳蜗就是麦克风传感器接收到外部声音信号,借助耳蜗中的电极,再通过电感应的方式通过皮肤把脉冲传到耳蜗,耳蜗里的电极再对这些信息经过处理刺激听觉神经,人就能够听到外界的声音。不过要使这种电子音的听觉升级到比较逼真的程度,就需要把电极数由最初的16个增加到3500个。另外一款利用体觉皮质原理的脑机接口设备人工视网膜的复杂程度则要高几个量级,如2011年FDA批准上市的第一款人工视网膜有60多个传感器^[5],戴上它之后可以让失明的人看到物体的边缘,努力增加到1000多个电极就可以阅读和识别人了;而人类的真视网膜有100多万个神经元,这相当于拥有100多万个传感器。

脑机接口的第2类应用是评估治疗。这里的评估包括对睡眠评估、学习的注意力评估,可以辅助教学。这里治疗是指通过音乐、电刺激进行治疗释放特定频率的电波来激发大脑皮层特定区域,增强专注力,提高学习效率;通过特定音乐舒缓情绪、放松精神;植入芯片来治疗治疗心理疾病如抑郁症等;刺激海马体协助恢复记忆。

脑机接口的第3类应用是借鉴科研,可以通过植入的微电极直接观察到脑部活动,对于掌握脑部神经组织的运作机制、脑部的疾病,对于研究认知和智能的机理、类脑芯片等,都具有重大的意义。

脑机接口技术预设的编码和解码涉及到人工智能,而信息反馈环节则更需要借助人工智能科技才能实现。人工智能是推动元宇宙建设的基本动力引擎。

(二) 人工智能科技

如前面所述,脑机接口的技术进展对元宇宙的发展水平有至关重要的作用,但对大脑的信息输入/反馈这个关键环节的完成却需要依赖于神经科学的进展以及人工智能的发展。首先需要人工智能生产出类似于人的感官所感受到的数字信息,由计算机把这种携带模拟信息的数字信号传输给脑机接口,再由后者把数字信号转换为大脑电信号,最终完成对大脑皮层输入反馈信息。这个输入环节中,无论是数字信息的产生,还是由数字信号向大脑电信号的转化,都需要人工智能的参与。

而如何产生出类似于人的感官感知外界的信息,这对于目前的人工智能而言是困难的问题。因为这意味着人工智能也可以以类人的方式感知和认知外部世界,模拟出类似的输入信息,乃至可以由算法虚拟地创造这种信息。只有神经科学彻底搞清楚了人的视觉、听觉、触觉、体感乃至嗅觉、味觉等的特性、输入方式以及处理机制,同时也需要人工智能技术足够发达,才可以由算法模拟高度接近于人的真实的感官信息。借助于这种类脑智能产生各种类人的逼真信息以及虚拟事物,并把这种虚拟信息通过脑机接口输入大脑皮层,最终让使用者产生一种身处于虚拟世界之中的真实体验。

而人工智能制造高度逼真的虚拟现实内容的能力,也需要借助于神经科学的充分发达,能够破译人体的感官信息以及大脑处理这些信息的方式,才能针对性地探索和建立高度发达的类脑算法,实现这种虚拟现实信息的生产和虚拟世界的建构。

另外,神经科学的研究也需要借助于脑机接口等神经元信息监测技术、人工智能算法的辅助科研,才能获得长足的发展。

由此可见,神经科学、人工智能、脑机接口代表的神经元信号检测这3个方面的技术具有相互依赖和相互促进的关系,需要相互支撑和循环迭代,才能最终走向高度成熟阶段。而元宇宙技术需要集成这几方面的进展才能进阶到高级阶段。

(三) 元宇宙核心技术实现前景

如前所述,这种身临其境的元宇宙体验是否能够实现,理论认知上依赖于脑科学的进展,在技术层面上依赖于脑机接口、传感器,以及人工智能的发展程度。但这些理论认知和技术是否最终能够实现,依赖于两个关键的假设能够成立:一是生物主义的假设,即人的体验、认知都与大脑神经元信号直接相关,或者就是大脑神经元活动的产物;二是类脑人工智能可以模拟出类似于人的感官产生的体验信息。

如果这两条假设从原则上说有任何一条不能成立,即人的意识内容不能还原为大脑神经元的信号,或者人工智能不能模拟出高度逼真的体验信息,那么低级版本的元宇宙虽可以实现,但终究无法接近或者达到普特南的“缸中之脑”思想实验所设想的高度逼真的元宇宙。

反过来看,普特南的“缸中之脑”思想实验在意识和认知问题上预设了物理主义/生物主义的立场,在人工智能问题上预设了强人工智能立场,这两个问题内在相关。人类的意识及认知是否可以完全还原为大脑的功能,这在科学和哲学上一直存在争论。彭罗斯、查尔莫斯等人一直坚持意识不等于大脑神经元组织的活动,而人的智能很大程度上是依赖于意识的活动^{[6] 10-20}。胡塞尔的现象学也认为,人的精神层面的活动具有一种遵循动机引发的自由的因果机制,并不能完全还原为身心之间关系。因此,按照这种立场,脑机接口即使可以监控所有神经元的活动,也并不能彻底揭示意识状态以及内容。另一方面,强

人工智能预设了功能主义的立场,认为智能依赖于算法和硬件,并不是人类大脑所独有的,也消除了意识的独立性,认为意识现象既可以由大脑产生,也可以由硅基的芯片加算法实现。但生物自然主义、二元论及现象学都认为意识具有本体论上的独立性,不能被还原为物理或生物的本体或属性,因此强人工智能的前提预设是错的,人工智能不能产生类人的智能,也不会产生意识。因此,按照这些哲学观点,人工智能不能代替人类的意识,因此也无法模拟出人类第一人称视角的体验和认知,虚拟世界的逼真性永远无法达到人类在真实物理世界中的体验和认知的水平^{[7]141-142}。

因此,脑机接口和人工智能的前景很大程度上决定了元宇宙的进度以及可以达到的程度。而这两个领域的发展前景又在很大程度上取决于脑科学的进展,或者更根本地说取决于认知科学和哲学对于意识现象的终极奥秘的揭示。

四、余论

虽然基于理论上的不确定性以及技术实现的困难,普特南设想的“缸中之脑”式的虚拟世界至少在短期内不一定能实现,但考量元宇宙未来可能的发展趋势以及对人类和整个世界历史进程的影响时,需要把这种极致的情况作为一种元宇宙未来发展的极端情况来参照。对于元宇宙的未来,我们需要预判未来的可能趋势,抉择我们要走的道路,制定可行的计划,规避可能的风险。

从科技战略和产业前景来看,现在就高度重视并理性规划元宇宙的建设是必要的,因为元宇宙技术的发展需要集成一系列相关的前沿技术。元宇宙的概念打通了人工智能、虚拟技术、增强现实、脑机接口、神经科学、物联网、虚拟货币、区块链技术、机器人、无人驾驶、信息安全、智慧城市等等众多的领域,可以在一个整体的视野中对这些领域的研究和应用进行整体规划,协同发展,相互促进,可能会形成产业集群的良性循环,促成一系列前沿尖端技术的大爆发,推动很多行业的发展,引领整个国家的工业产业的革新和转型,甚至可以推动新的技术革命和产业革命,搭上国际产业新的分工的顺风车。

关于元宇宙问题,2021年以来还有一场颇有意思的争论。国内外众多互联网大厂、软硬件企业都押重于元宇宙。但据媒体报道,我国著名科普作家刘慈欣近日在公开演讲中却严厉批评元宇宙:“人类的未来,要么是走向星际文明,要么就是常年沉迷在VR的虚拟世界中。如果人类在走向太空文明以前就实现了高度逼真的VR世界,这将是一场灾难。”^[8]不同于一些批评者认为元宇宙是概念炒作和资本骗局,刘慈欣在这里担心元宇宙的高度的诱惑性、致幻性会导致人们沉迷于虚拟世界中,不思进取,从而将引导人类走向歧途。在他看来,人类应该积极发展探索太空和宇宙的技术,才能为人类未来的拓展更大的生存空间,太空文明才是人类文明的希望所在。

应该说,刘慈欣的担忧是有一定道理的,元宇宙的根本问题在于它建构的虚拟世界越逼真,则致幻效果就越强。就目前的脑科学和认知心理学研究看,人类的进步遵循一种学习机制,努力得到回报,会形成一种正向的努力与激励的良性循环,推动人们更努力去创造、发明。但元宇宙却可以让人们省略在真实世界中的奋斗过程和激励路径,直接通过脑机接口的信息输入刺激大脑神经而满足消费者的大多数需要。这种激励的方式与精神致幻剂有类的效果,容易造成对元宇宙的高度依赖和成瘾。这种导向的激励会导致人们越发沉迷于虚拟世界,会极大地消解人类在现实世界中奋斗的意志。这种做法虽然可以暂时地安慰心灵、减少痛苦和焦虑,但毕竟治标不治本。毕竟,元宇宙的“矩阵”式虚拟世界和基于营养仓的“缸中之脑”般的虚拟人生不应该成为人类文明的主要方向。毋庸置疑,那种元宇宙导向的人类文明会严重延缓人类探索未知和发展文明的进程。

参考文献:

- [1] 《深圳市人民政府关于加快智慧城市和数字政府建设的若干意见》解读[EB/OL]. (2021-01-05)[2021-11-08]. http://www.sz.gov.cn/szsj/gkmlpt/content/8/8394/post_8394067.html.
- [2] 祝涛. 英伟达黄仁勋的元宇宙梦想:虚拟世界将如同互联网站那样不断涌现[EB/OL]. (2021-11-10)[2021-11-12]. <https://new.qq.com/omn/20211110/20211110A0AQ9D00.html>.
- [3] 普特南, 希拉里. 理性、真理与历史[M]. 童世骏, 李光程, 译. 上海: 上海译文出版社, 2016.
- [4] 付静. 刚刚, 马斯克宣布脑机接口重大进展! 请来三只小猪现场展示[EB/OL]. (2020-08-29)[2021-11-06]. <https://www.leiphone.com/category/industrynews/wcmsOwUIZZEK4l0l.html>.
- [5] 佚名. 美国发明人造视网膜技术 可助盲人复明[EB/OL]. (2011-02-15)[2021-11-06]. <https://www.chinanews.com.cn/gj/2011/02-15/2844650.shtml>.
- [6] 张昌盛. 论人工智能哲学的现象学进路[J]. 重庆理工大学学报(社会科学), 2020(12): 139-148.
- [7] 张昌盛. 人工理性批判: 对德雷福斯的人工智能哲学的现象学反思[J]. 重庆理工大学学报(社会科学), 2018(12): 9-20.
- [8] 邢思远. 刘慈欣怒批元宇宙[EB/OL]. (2021-11-12)[2021-11-12]. https://www.thepaper.cn/newsDetail_forward_15327743.

Artificial intelligence, brain in a vat and virtual life

——An interdisciplinary study of meta verse

ZHANG Changsheng

(1. School of Philosophy, University of Chinese Academy of Social Sciences, Beijing 102488, China;

2. Institute of Philosophy, Chinese Academy of Social Sciences, Beijing 100732, China)

Abstract: The metaverse is a kind of virtual world that conceived by scientists in parallel with the real world. For the purpose of scientifically knowing metaverse, this paper briefly combs the ideological history and current development status of the metaverse. Then with the help of the discussion on the ideological experiment of philosopher Putnam's "brain in a VAT", this paper makes a theoretical analysis on the metaverse from the perspectives of science and philosophy, mainly involving the discussion on the relationship between brain, consciousness and intelligence, and on the other hand, the discussion on the possibility of strong artificial intelligence technology. On this basis, the possible forms of the future metaverse are analyzed from the technical conditions and philosophical basis of brain computer interface and artificial intelligence. Finally, it analyzes the guidance and possible negative impact of the metaverse on human life and civilization in the future.

Key words: metaverse; brain in a vat; artificial intelligence; brain computer interface; guidance

(责任编辑 张佑法)