

生命进化与人工智能

——对生命 3.0 的质疑

蔡曙山

摘要：2016 年 AlphaGo 战胜李世石以后，人工智能战胜人类的神话已成为现实。2017 年泰格马克的奇书《生命 3.0》对人工智能赋予生命。2018 年霍金对人工智能的发展提出警告：彻底开发人工智能可能导致人类灭亡。这些“大问题”应该从生命进化、人类心智和人工智能的发展以及人与自然的关系中寻找答案。在达尔文物种进化论和现代综合进化论（基因进化论）的基础上，根据人类认知五层级理论，本文创立心智进化论，对生命 3.0 提出质疑，最后得出结论：生命是自然进化的产物，人造生命违背了自然法则；心智和意识是生命的唯一标准，不存在其他的标准和定义；迄今为止，人工智能只是人类智能的一种表现形式；人类永远不能也不可能充当上帝，更不可能是宇宙万物的主宰。

关键词：人工智能；生命 3.0；生命进化；人类心智；心智进化论

中图分类号：TP18 文献标识码：A 文章编号：1004-8634(2020)03-0083-(17)

DOI: 10.13852/J.CNKI.JSHNU.2020.03.008

人类历史进入倒计时？

当前的人工智能神话，最强的告诫莫过于“霍金警告”。

霍金多次表示，“彻底开发人工智能可能导致人类灭亡”。他说“到目前为止开发的原始形式的人工智能被证明非常有用，我却害怕创造出匹配或超越人类的某种东西的后果。我担心的是，AI 会自己起飞并不断加速重新设计自己。人类受到缓慢的生物进化的限制，无法竞争，将会被超越。”^①科学界把人工智能超越人类智慧的转折点命名为“奇点”——正如宇宙大爆炸也起源于这样一个奇点。^②

在人生最后的时间里，霍金频繁地回答科学家、科技企业、高级商业人士、政治领袖及公众问及的一些“大问题”。这些回答形成了一份巨大的个人档案。霍金去世后，他的学术同事、家人和遗产管理机构合作，根据这份档案出版了《十问：霍金沉思录》（*Brief Answers to the Big Questions by Stephen Hawk-*

基金项目：国家社会科学基金重大项目“语言、思维、文化层级的高阶认知研究”（15ZDB017）；国家社会科学基金重大项目“汉语非字面大脑加工的神经机制研究”（14ZDB154）

作者简介：蔡曙山，贵州民族大学民族文化与认知科学学院教授，清华大学心理学与认知科学研究中心主任，教授，博士生导师（北京 100091）。

① 史蒂芬·霍金《十问：霍金沉思录》，吴忠超译，湖南科学技术出版社 2019 年版，第 158 页。

② 奇点（Singularity）是宇宙大爆炸之前宇宙存在的一种形式。它具有一系列奇异的性质，无限大的物质密度、无限弯曲的时空和无限趋近于 0 的熵值等。

ing)。在该书“人工智能是否会以计谋打败我们”(Will Artificial Intelligence Outsmart Us)一章中,霍金最后警告说“计算机能力正在增长,量子计算机正在迅速实现。这将以指数方式的更快速度革新人工智能。量子电脑将改变一切,甚至人类生物学。”^①

智能机器是否会成为人类“终结者”?人类历史是否已经进入倒计时?

对于这些“大问题”,本文将从生命进化、人类心智的进化和人工智能的发展以及人与自然的关系中寻找答案。

人类心智的进化

生命是宇宙和地球演化以及生命进化的产物。基因和心智是地球演化和生命进化的两个重要结果,心智的进化依次产生了神经、心理、语言、思维与文化五个层级的心智和认知形式。

宇宙的历史150亿年,地球的历史46亿年,生命的历史34亿年。在生命的漫长进化中,直立行走、火的使用和语言的发明三大事件最终使人类进化为人。

关于生命的进化,迄今有两种基本的理论和解释:达尔文进化论(Darwin's Theory of Evolution)和现代综合进化论(Modern Comprehensive Evolution)。在此基础上,我们建立认知科学的心智进化论(Theory of Mind Evolution)。

1. 达尔文进化论(物种进化论)

1859年,达尔文(Charles Robert Darwin, 1809-1882)发表《物种起源》,标志着进化论的诞生。达尔文认为,生物之间存在着生存竞争,适应者会生存下来,不适者则会被淘汰,这就是自然的选择。生物通过遗传、变异和自然选择,从低级到高级,从简单到复杂,种类由少到多地进化和发展。达尔文进化论的四个成分一是进化论,即物种是可变的,现有的物种是从别的物种变来的,一个物种可以变成新的物种;二是共同祖先学说,即认为所有的生物都来自共同的祖先。分子生物学发现了所有的生物都使用同一套遗传密码,生物化学揭示了所有生物在分子水平上有高度的一致性;三是自然选择,即优胜劣汰是进化的主要机制;四是渐变论,即认为生物进化的步调是渐变式的,而不是跃变式的,它是一个在自然选择作用下,累积微小的优势变异的演化过程。

2. 现代综合进化论(基因进化论)

综合进化论是对达尔文进化论的发展,它应用现代基因科学的理论和研究成果,充分重视基因的变异在生命进化中的作用,以此建立了新的生命进化理论,其主要之点是:第一,基因突变、染色体畸变和通过有性杂交实现的基因重组是生物进化的表现形式;第二,进化的基本单位是群体而不是个体,进化是由于群体中基因频率发生了重大的变化;第三,自然选择决定进化的方向,生物对环境的适应性是长期自然选择的结果;第四,隔离导致新种的形成。长期的地理隔离常使一个种群分成许多亚种,亚种在各自不同的环境条件下进一步发生变异就可能出现生殖隔离,形成新种。

现代综合进化论彻底否定了获得性的遗传,强调进化的渐进性,认为进化现象是群体现象并重新肯定了自然选择压倒一切的重要性。

3. 认知科学进化论(心智进化论)

以上两种关于生命与进化的理论和解释,共同的缺陷是只看到生命的表现形态和基因的表现形态,而没有看到心智(Mind)的表现形态,没有看到心智在生命进化中的作用。

20世纪70年代以来,认知科学的发展取得了一系列重要的研究成果,根据认知科学的理论,我们可以对生命的进化提出新的解释,这就是认知科学的进化论,即心智进化论。认知科学将心智看作是生

^① 史蒂芬·霍金《十问:霍金沉思录》,第158页。

命进化的依据,并以心智为研究对象。认知科学是研究心智与认知现象及规律的科学。^① 认知科学的进化论或心智进化论,是根据认知科学的理论、方法和研究成果,对生命进化提出新的解释。根据认知科学的基本原理,将生命的进化过程看作是心智的进化过程。由于心智决定认知,因此,心智和认知从低级到高级的发展决定了生命从低级到高级的发展。

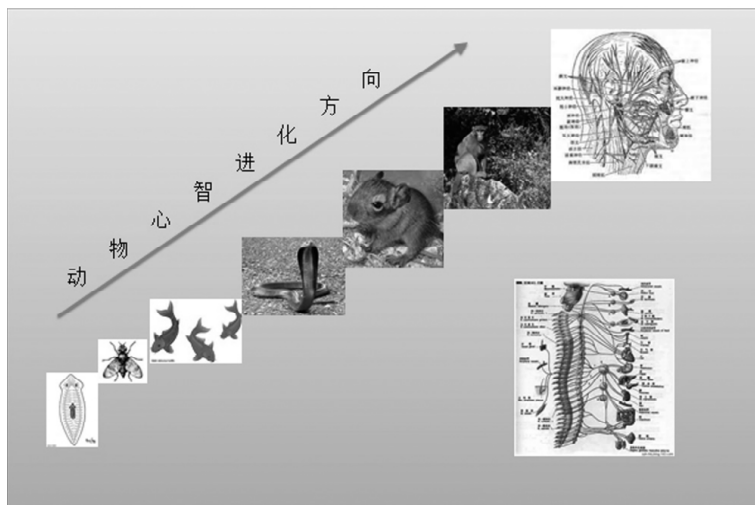


图1 动物心智进化图(蔡曙山《认知科学基础》2015)

根据心智进化论,我们将生命 35 亿年的演化过程,看作是一个心智的进化过程,心智的进化水平决定了动物的种性和形态。动物心智的进化可以简略地用图 1 来表示。

综上所述,可以看出,达尔文进化论看到内在条件(遗传和变异)和外在条件(自然选择)对物种进化的作用,但看不到基因在生命进化中的作用。综合进化论看到基因在生命进化中的作用,但却没有看到心智在生命进化中的作用。心智进化论重视心智在生命进化中的作用,超越了达尔文的物种进化论和基因进化论,对生命的进化提出了新的解释。

这里涉及古老的“心身问题”(Mind-body Problem),即心智和身体究竟谁是第一性的大问题,我们将这个问题留待后文讨论。

人类认知的五个层级

认知科学的科学和学科特征可以归纳为两句话:五层级贯通、多学科综合。五层级是科学结构,“6+1”是学科框架。

1. 人类认知的五个层级

按照心智进化论,动物心智的发展和进化经历了神经、心理、语言、思维和文化五个发展阶段,形成五个层级的心智能力。人类继承了全部五个层级的心智和认知能力,而非人类的动物只具有神经和心理两个层级的心智和认知。对心智的进化进行结构的分析,我们建立了人类认知五层级理论。^②

在认知科学中,认知是用心智来定义的。心智是认知主体对环境信息进行加工,得出有用信息,并用以指导自身行为的一种能力。人类在进化中获得的五种心智使人类具备五种认知能力,这就是神经认知能力、心理认知能力、语言认知能力、思维认知能力和文化认知能力。

五种心智和认知是具有层次结构的,我们称之为“人类认知的五个层级”,五层级认知的关系如下:

^① 蔡曙山《认知科学框架下心理学、逻辑学的交叉融合与发展》,《中国社会科学》2009年第2期。

^② 蔡曙山《论人类认知的五个层级》,《学术界》2015年第12期。

(1) 五个层级的认知首先被区分为非人类动物的低阶认知和人类特有的高阶认知。非人类动物仅具有的神经认知和心理认知称为低阶认知,人类所特有的语言认知、思维认知和文化认知称为高阶认知。人类认知涵盖所有五个层级,包括高阶认知和低阶认知。从神经认知、心理认知、语言认知、思维认知到文化认知的发展,是动物和人类认知进化方向的体现;人类认知的五个层级的存在,是心智和认知进化各阶段能力的遗留。

(2) 每一种初级认知依次成为高级认知的基础。神经认知是心理认知的基础;心理认知是语言认知的基础;语言认知是思维认知的基础;思维认知是文化认知的基础。当然我们也可以说,神经认知和心理认知是语言认知的基础;神经认知、心理认知和语言认知是思维认知的基础;神经认知、心理认知、语言认知和思维认知是文化认知的基础。

(3) 由于高级认知向下包含了较初级的认知,所以较高级的认知形式会对它所包含的初级认知形式产生影响。文化认知会对思维认知、语言认知、心理认知和神经认知产生影响;思维认知会对语言认知、心理认知和神经认知产生影响;语言认知会对心理认知和神经认知产生影响,等等。

2. 人类心智的进化与高阶认知

在生命进化的漫长进程中,三件大事最终完成了人类的进化,这三件大事是:直立行走、火的使用和语言的发明。直立行走使猿的一支人猿走出丛林,走向平原,扩大了活动范围,并腾出前肢进行采摘和捕猎。火的使用使人类能够加工熟食,增强体魄,并能大量摄入异体蛋白,脑容量爆发式增大。^①语言的发明使人类通过特有的抽象符号语言(概念语言)进行思想交流和行为交际,协调更大范围的群体行为,使人类从相对弱小的个体形成强大的群体。在抽象符号语言(概念语言)的基础上,人类能够进行抽象思维,语言和思维建构出理性逻辑的知识体系,知识积淀为文化。从此,人类的进化不再是或者主要地不是基因层次的进化,而是语言、思维和文化的进化,即人类心智的进化。^②

高阶认知的三种形式语言、思维和文化,是人类特有的三种认知能力。由语言认知、思维认知和文化认知构成的高阶认知是人类特有的认知形式,非人类的动物并不具有这种认知形式。因此,高阶认知也就是人类认知。

(1) 语言认知

600 万年至 200 万年前,南方古猿首先发明出抽象表意的符号语言,据此,他们协调了更大范围的群体行为,战胜比他们更强大的猿类,走出非洲,把自己的基因撒播到全世界。在人类认知的五个层级中,语言认知是核心。在高阶认知中,语言认知是基础。语言的发明,使人类脱离非人类动物而最终进化为人。人类的存在,不过就是语言的存在。我言,故我在。^③

20 世纪的语言学革命最终导致认知科学的建立。首先是 20 世纪 50 年代以乔姆斯基句法结构理论为标志的语言学革命,这一革命成为认知科学革命的先声绝非偶然。从人类认知五个层级我们知道,语言是全部人类认知即高阶认知的基础,所以,认知科学革命从语言领域爆发是必然的。然后是相继发生的以蒙太格(R. Montague)为代表的语义学革命和奥斯汀(J. L. Austin)、塞尔(J. R. Searle)为代表的语用学革命。语言学的这些成就引发了语言与心智研究、^④先天语言能力的研究、^⑤言语行为的研究、^⑥

① 考古证据表明,人类脑容量的爆发式增大与火的使用处于同一进化年代。

② 媒母(meme)亦称文化基因,是衡量人类心智的一个根本指标。基因(gene)则是衡量生理状态的指标,两者关系是心身问题的现代版本。

③ 蔡曙山《论语言在人类认知中的地位 and 作用》,《北京大学学报(哲学社会科学版)》2020 年第 1 期。

④ N. Chomsky, *Language and Mind*, New York: Harcourt, Brace & World, 1968.

⑤ N. Chomsky, *Language and Mind*.

⑥ J. R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, London: Cambridge University Press, 1969.

语用学的研究、^①意向性的研究、^②语言哲学和心智哲学的研究,^③这些研究指向一个共同的目标: 人类心智, 这些发展最终导致 20 世纪 70 年代中期认知科学的建立。

(2) 思维认知

在抽象的概念语言的基础上, 人类产生了抽象思维。这是人类进化中重要的一步, 语言和思维建构了全部的人类知识体系。所以说, 人类认知是以语言为基础, 以思维为特征的。笛卡尔尽人皆知的名言是: 我思, 故我在。笔者著文指出“人类的存在, 包括作为认知主体的存在, 作为思维主体的存在和作为文化主体的存在, 不过是语言的存在。‘我思, 故我在’这个经典的哲学命题, 在认知科学发展的今天, 应该被‘我言, 故我在’这个更深刻的命题所取代。”^④

20 世纪初, 为解决第三次数学危机即数学基础的危机而建立的数学逻辑(Mathematic Logic), 在形式语言的基础上建立逻辑分析的形式系统, 即一阶逻辑, 将数学的基础奠定在形式语言和形式逻辑系统之上。1930 年, 哥德尔证明这个系统是一致的, 即系统内所有可证的公式(定理) 都是真的。这样, 罗素悖论所引起的数学危机就解决了。同时, 他还证明这个系统是完全的, 即系统内所有的真公式都是可证的。尽管在历史上也曾经产生过如欧氏几何和亚氏三段论那样的公理系统, 但在形式语言的基础上建立形式系统并证明系统的一致性和完全性, 这在人类历史上还是第一次, 意义重大。1931 年, 哥德尔证明了一个更加伟大的定理: 形式系统的不完全性定理, 这个非凡的定理后来以他的名字命名: 哥德尔定理。^⑤ 哥德尔定理分为两部分: 第一不完全性定理和第二不完全性定理。第一不完全性定理说, 如果一个至少包含算术系统的形式系统是一致的, 那么, 它就是不完全的; 第二不完全性定理说, 如果这个形式系统是一致的, 那么, 它的一致性在系统内是不能证明的。哥德尔定理在一致性和完全性之间建立关系: 在一个形式系统内, 一致性和完全性是不可得兼的。一致性是一个逻辑系统必须满足的, 如果这样的话, 则这个系统是不完全的, 即系统内存在真而不可证的命题。并且, 系统的一致性系统本身无法证明的。哥德尔定理粉碎了人类可以全知全能梦想, 也粉碎了人类试图穷尽一切真理的梦想。哥德尔定理告诉我们, 人类的理性是有限的。哥德尔定理是人类理智结出的最灿烂的花朵。

(3) 文化认知

语言和思维形成知识, 知识积累为文化。人类发明语言、建立思维、形成知识、产生文化, 此后, 人类的发展不再是或者说主要的不是基因层级的进化, 而是语言 and 思维的发展以及文化的进化, 这是 600 万年前人类发明语言以后的发展历程。

文化是人类认知的最高形式, 文化认知是人类认知的最高层级。文化的内涵丰富, 文化就是人化, 是人所创造的一切存在物, 包括物质的存在和精神的存在。文化的城邦化就是文明。另一方面, 还要万分注意自然与文化的关系, 人类文化连同人类自身归根到底是自然进化的产物, 离开自然, 人类和文化都将不复存在。近代以来, 文化和文明越来越成为自然的对立物和异己的力量, 正如马克思和恩格斯所警告的, 人类正在遭受到大自然越来越严重的报复。^⑥

文化的内涵丰富、结构复杂。我们可以将其分为三个层级: 科学、哲学和宗教。^⑦ 还可以把文化按

① N. Kadmon, *Formal Pragmatics: Semantics, Pragmatics, Presupposition and Focus*, Blackwell Publishers, 2001.

② J. R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, London: Cambridge University Press, 1983.

③ J. R. Searle, *Speech Acts: An Essay in the Philosophy of Language*; J. R. Searle, *Mind: A Brief Introduction*, New York: Oxford University Press, 2004.

④ 蔡曙山《论语言在人类认知中的地位和作用》。

⑤ Gödel, Kurt. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, translated by B. Meltzer, introduction by R. B. Braithwaite, New York: Dover Publication, Inc., 1962.

⑥ 恩格斯《自然辩证法》, 参见《马克思恩格斯选集》(第 4 卷), 人民出版社 1972 年版, 第 284 页; 马克思《资本论》(第 3 卷), 人民出版社 2004 年版, 第 251 页。

⑦ 蔡曙山《自然与文化——认知科学三个层次的自然文化观》, 《学术界》2016 年第 4 期。

照七要素进行分类,即经济、政治、科学、宗教、道德、文学和艺术。^①文化的每一个方面都包含十分丰富的内容,都是人类知识的积淀和结晶。人类文化是自然进化的结果,所以,由联合国发起、联合国教育科学文化组织负责执行的世界遗产分为世界文化遗产、世界文化与自然双重遗产、世界自然遗产3类,足以彰显自然的价值、文化的价值、自然与文化的共同价值及自然与文化的关系。

文化最重要的特征是创造。科学、哲学、宗教、经济、政治、道德、文学、艺术无一不是人类的创造,而且是人类以语言为基础、以思维为工具的创造。哥德尔从形式语言和形式系统创造了不完全性定理,揭示人类理性的有限性,揭示系统内存在真而不可证的命题,揭示系统无法证明自身的一致性。爱因斯坦从光速有限和坐标表权两个前提创造了狭义相对论,推导出爱因斯坦公式,释放出原子能。20世纪的科学家从谱线红移这个物理现象,推导出宇宙膨胀的结论,建立了宇宙大爆炸论,等等。所有这些,说到底都是自然进化和人类心智发展的结果。

那么,是否存在自然进化系统之外的生命?机器会不会进化?人工智能是否具有生命?人工智能是否最终会超过人类智能?人工智能会不会危及人类的生存?对于这些“大问题”,我们将通过对《生命3.0》一书的质疑做出回答。

对“生命3.0”的质疑

近年,“语不惊人死不休”的一部奇书恐怕要数泰格马克(Max Tegmark)著的《生命3.0》,此书一出,轰动天下。

泰格马克,毕业于斯德哥尔摩经济学院与瑞典皇家理工学院物理系,获美国加州大学伯克利分校物理系博士学位,现任麻省理工学院物理系终身教授,在物理学上创立了平行宇宙论和数学宇宙假说。他认为,我们所看到的物理实在其实都是数学结构——我们不仅可以用数学描述所处的宇宙,甚至可以说宇宙本身就是数学。他关于星团的研究获得《科学》杂志“2003年度突破奖”第一名,他被誉为“当今最具原创力的物理学家之一”。顶着这样的光环,2014年他与Skype创始人杨塔里安(Jaan Tallinn)共同创立了“未来生命研究所”(Future of Life Institute, FLI)。FLI的目标很简单:保证生命有未来,会继续存在下去,并尽可能地兴旺发达。^②泰格马克关于未来生命的设想要点如下:

生命1.0发源于约40亿年前的生物阶段,在它的有生之年都无法重新设计自己的硬件和软件,二者皆由它的DNA决定,只有进化才能带来改变,而进化则需要许多世代才会发生。生命2.0大约产生于10万年前,也就是人类诞生以后的生命形式。生命2.0虽然不能更新硬件,但可以更新软件,即可以重新设计自身软件的一大部分:人类可以学习复杂的新技能,如语言、运动和职业技能,并且能够从根本上更新自己的世界观和目标。生命3.0是一种预言,目前在地球上尚不存在,它不仅能最大限度地重新设计自己的软件,还能够重新设计自己的硬件,而不用等诸多世代的缓慢进化。^③

对于泰格马克的这个奇思妙想,并不是所有人都同意。特别是对于“生命3.0何时出现”这个问题,泰格马克本人也承认“这个问题极富争议,而且争议得十分精彩”。“人工智能何时会超过人类”“如果超人类水平的人工智能出现了,会是一件好事吗”这两个问题的不同意见见图2。

在是非对错的判断上(横轴),卢德主义者(Luddites)^④相信,人工智能的发展结果一定是坏的,所以反对人工智能。数字乌托邦主义者(Digital Utopian)不仅认为超过人类水平的人工智能肯定会发生,而

① 钱穆《文化学大义》,九州出版社2011年版,第33—57页。

② M.泰格马克《生命3.0》,汪婕舒译,浙江教育出版社2018年版,第v页。

③ M.泰格马克《生命3.0》,第33页。

④ 卢德主义出现于工业革命初期,那时候的手工业工人对于大机器生产的出现非常愤恨,他们认为机械化和自动化使他们的工资降到了很低的水平,新技术迫使劣质产品进入市场。于是憎恨大机器,开始破坏这些新出现的机器设备,以换取就业。卢德主义者用来指代那些反对技术进步的人。

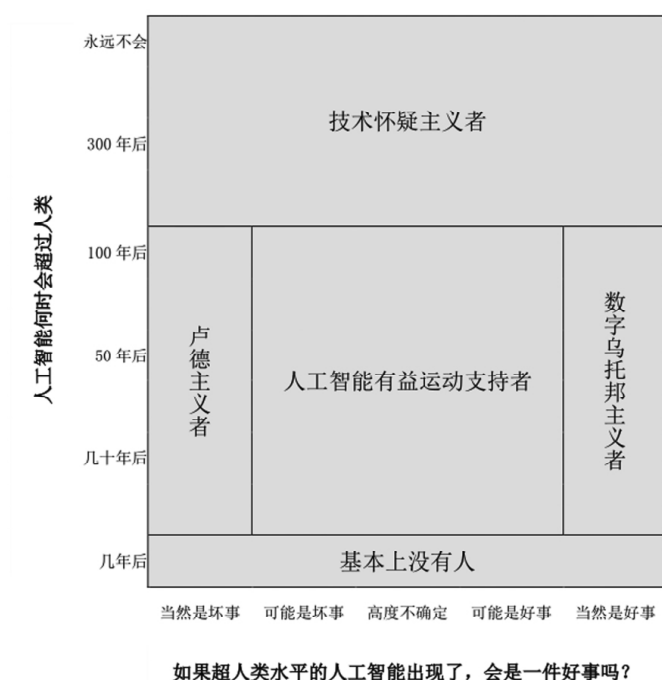


图2 各个学派关于强人工智能的争议

且认为那绝对是一件好事。大多数人是人工智能有益运动的支持者，他们对人工智能的态度在卢德主义和数字乌托邦主义之间，对于超人类人工智能的出现，他们的态度是高度不确定，认为可能是好事，也可能是坏事。

在时间早晚的判断上（纵轴），几年后出现超过人类水平的人工智能基本上没有人相信。但卢德主义者、人工智能有益运动的支持者和数字乌托邦主义者这三类人尽管对超级人工智能的价值判断有差异，但却一致认为超过人类水平的人工智能在今后一百年内一定会实现。只有技术怀疑主义者认为，建造超过人类水平的人工智能实在是太困难了，没有几百年时间，根本无法实现，因此，没有必要杞人忧天。泰格马克说，他曾当面向机器人制造专家罗德尼·布鲁克斯（Rodney Brooks）求证，“奇点”有无可可能发生？^① 布鲁克斯回答“百分之百地肯定，奇点不会发生在我的有生之年”。泰格马克追问“你确定你的意思不是99%？”布鲁克斯回答说“不是没用的99%，就是100%。奇点根本不会发生。”^②

泰格马克是一位严肃的科学家，他提出的思想、理论和观点值得严肃对待。下面是几个值得严肃认真思考的大问题。

1. 有无进化系统之外的生命

对这个问题，虽然泰格马克没有给出明确的是或否的回答，但从《生命3.0》一书的论述仍然可以清晰地看出，作者对这个问题的回答是肯定的。

首先来看作者对生命的定义。泰格马克说“生命最早是在何时何地、以何种方式出现在我们宇宙中的呢？这个问题依然没有答案。不过，有力的证据表明，地球上的生命最早出现在大约40亿年前。不久之后，我们的地球上就充满了各种各样的生命形态。那些最成功的生活很快便从中胜出，并具备了某种与环境共生的能力。具体而言，它们就是被计算机科学家称为‘智能体’（Intelligent Agent）的东西：这种实体用感应部件收集关于环境的信息，然后对这些信息进行处理，以决定如何对环境做出回应。可

① 此处借指通用人工智能出现的转折点。

② M.泰格马克《生命3.0》第42页。

以包括高度复杂的信息的处理过程,例如,你能用眼睛和耳朵收集信息,并用这些信息来决定在一段对话中要说什么;不过,它也可以只包括非常简单的硬件和软件。”^①

这里采用了简单粗暴的方法,将生命定义为硬件和软件系统。为什么要用“硬件”和“软件”这种计算机术语来定义生命?就是为了推出计算机生命——生命 3.0 的存在。定义的公式如下:

$$\begin{aligned} \text{生命} &=_{\text{df}} \text{硬件} + \text{软件} \\ &=_{\text{df}} \begin{cases} \text{生命 1.0 硬件} - \text{软件} - \\ \text{生命 2.0 硬件} - \text{软件} + \\ \text{生命 3.0 硬件} + \text{软件} + \end{cases} \end{aligned}$$

以上定义公式中,“ $=_{\text{df}}$ ”表示“定义为”,左边是被定义项,右边是定义项。“+”表示“能更新”,“-”表示“不能更新”。按照这样的定义,泰格马克得出“生命三个阶段”的结论。

泰格马克还有另一个生命的定义:将生命定义为一个能够“保持自己的复杂性,并进行复制的过程”。^②在这种定义下,宇宙万物和技术进步都可以看作某种生命系统。他说“我们的婴儿宇宙也像你一样,经历过指数型增长,以固定的周期,规律地将自己的尺寸翻倍,从最初的那一团比原子还小还轻的物质迅速膨胀,一直到超过我们用望远镜可以看到的所有星系。在这个过程中,每次翻倍都会引发下一次翻倍。技术进步的过程也同样如此:当一项技术的能力变成过去的两倍时,通常情况下,它又可以用来设计和建造能力翻番的技术,引发不停歇的能力翻番,这就是摩尔定律的精髓。”^③他论证了生命与它的物质形态无关。他说“硬件就是物质,软件就是形态。计算的‘物质’层面的独立性暗示着我们,人工智能是可能实现的:智能的出现并不一定需要血肉或碳原子。”^④他还说“一团物质想要学习,必须对自己进行重新排列,以获取越来越强的能力,好计算它想要的函数,只要它遵守物理定律就行。”^⑤他明确说“我们宇宙中的生命的最终极限取决于物理定律,而不取决于智能。”^⑥(以上引文以字体区别来做强调,为原文所有,下同——笔者注)如此这般,泰格马克便赋予宇宙万物以生命。为了保持他的这个理论的一致性,他给出了关于生命定义的一些名词术语的备忘录(表 1)。^⑦

表 1 生命与人工智能名词术语备忘录

名词术语	定义
生命(Life)	能保持自己的复杂性,并进行复制的过程
生命 1.0(Life 1.0)	靠进化获得硬件和软件的生命(生物阶段)
生命 2.0(Life 2.0)	靠进化获得硬件,但自己能设计软件的生命(文化阶段)
生命 3.0(Life 3.0)	自己设计硬件和软件的生命(科技阶段)
智能(Intelligence)	完成复杂目标的能力
人工智能(AI)	非生物的智能
专用智能(Narrow Intelligence)	可完成一个较狭义的目标组(例如下棋或开车)的能力
通用智能(General Intelligence)	可完成几乎所有目标(包括学习)的能力
普遍智能(Universal Intelligence)	在拥有数据的资源的情况下,可获得通用智能的能力
通用人工智能(AGI)	可完成任何认知任务,并且完成得至少和人类一样好的能力

① M.泰格马克《生命 3.0》,第 32 页。

② M.泰格马克《生命 3.0》,第 50 页。

③ M.泰格马克《生命 3.0》,第 90 页。

④ M.泰格马克《生命 3.0》,第 88 页。

⑤ M.泰格马克《生命 3.0》,第 93 页。

⑥ M.泰格马克《生命 3.0》,第 59 页。

⑦ M.泰格马克《生命 3.0》,第 50—51 页。

(续表)

名词术语	定义
人类水平的人工智能(Human-level AI)	其能力同通用人工智能的能力
强人工智能(Strong AI)	其能力同通用人工智能的能力
超级智能(Super Intelligence)	远超人类智能水平的通用智能
文明(Civilization)	一组相互影响的智能生命形式
意识(Consciousness)	主观体验
感质(Qualia)	主观体验的单个实例
伦理(Ethics)	制约我们应该如何行为的原则
目的论(Teleology)	用目标或意志而不是原因来解释行为
目标导向行为(Goal-oriented behavior)	更容易用目标或意志而不是原因来解释行为
拥有目标(Having a goal)	展现出目标导向行为
拥有意志(Having purpose)	服务于自己或其他实体的目标
友好的人工智能(Friendly AI)	目标与我们一致的超级智能
赛博格(Cyborg)	人与机器的混合体
智能爆炸(Intelligence Explosion)	能迅速导致超级智能的迭代式自我改进的过程
奇点(Singularity)	智能爆炸
宇宙(University)	在自宇宙大爆炸以来的 138 亿年的时间里 ,光线足以到达地球的空间领域

泰格马克关于生命的第一个定义是循环定义 ,他先将生命定义为硬件加软件 ,再用硬件加软件来定义生命 1.0 到 3.0。这种循环定义在逻辑上是错误的。泰格马克关于生命的第二个定义也有明显的问题。什么是“复杂性”? 这是一个模糊的概念 ,用一个模糊的概念来定义生命 ,其本质属性不可能得到揭示。例如 ,一个 DNA、一个 2019-nCoV 病毒都非常复杂 ,它们都能够自我复制 ,但前者不是生命 ,后者虽然是生命 ,但却不是生物。按照这个定义 ,生命 1.0 既不能复制自己的硬件 ,又不能复制自己的软件 ,它们应该不符合生命的定义 ,因而整个的非人类动物都应该被排斥在生命系统之外 ,这岂不荒唐? 泰格马克对生命的定义犯了逻辑上循环定义和定义项不明确的错误 ,是不能自圆其说的。关于生命的定义 ,我们在后文还会做更多的讨论。

2. 关于“被嵌入的生命”(生命 3.0)

泰格马克并不认为生命 3.0 是进化出来的生命 ,而认为它是人工智能发展到一定程度会产生出来的新的生命形态 ,称为“被嵌入的生命”。

“被嵌入的生命”有两重含义 ,一种含义是人工智能嵌入到人体之中 ,改变人体硬件 ,从而进入软硬件都可以更新的“生命 3.0”。目前的芯片植入、脑机接口等技术 ,正向生命 3.0 发展。但是 ,这条路看不出有超越人类智能的普遍智能(General Intelligence) 或通用智能(Universal Intelligence) 出现的可能性。另一种含义是人工智能发展到人类水平的智能(Human-level AI) 以后 ,它逃脱人类的控制 ,依靠自我进化 ,成为超越人类的生命形式 ,即“生命 3.0”。

在《生命 3.0》一书中 ,作者开篇就讲了一个欧米茄团队和普罗米修斯的故事。故事以文学想象和好莱坞手法 ,讲了人工智能体普罗米修斯如何从它的设计者欧米茄团队的控制下成功越狱 ,从赚钱、网络、商业、技术、改进硬件、电影生产、宣传、说服、教育到竞选、政治、获取权力 ,最终控制全人类。这个故事整整讲了一章 ,共 25 页 ,读后不知是科幻作品、好莱坞剧本还是文学创作 ,看来严谨的科学家也难免随性而为 ,信口开河。

泰格马克承认 ,这样的智能体(Agent) 目前尚不存在 ,但他和数字乌托邦主义者都认为 ,这个最终理

想一定会实现,超过人类智能的通用智能一定会出现,生命3.0这个人工智能的圣杯一定会被他们获得。是否如此?本文将在结论部分再加以讨论。

3. 人类是“有限智慧”,那么,存在“超级智能机器”吗

根据哥德尔定理,人类所建构的哪怕是最严格的形式系统,也存在真而不可证的命题;另外,一个这样的形式系统,其自身的一致性自身所不能证明的。这就判定了,人类的理性是“有限理性”,人类的智能是“有限智能”。也就是说,人类的心智和认知能力是有限的。那么,存在无所不能的“超级智能”和“超级智能机器”吗?

在表1中,泰格马克定义了各种智能,按照他的定义和论述,我们可以把这些智能排列成一个等级(图3)。在这个等级图中,无所不能的通用智能(General Intelligence)居于最高等级,在拥有数据和资源的情况下可获得通用智能能力的普遍智能(Universal Intelligence)与之平起平坐,数字乌托邦主义者期望的圣杯生命3.0当仁不让地与通用智能居于最高层级。根据泰格马克的定义,人类水平的人工智能(Human-level AI)、通用人工智能(AGI)和强人工智能(Strong AI)是同一水平的智能。在这两个层级之间,是“远超过人类水平的通用智能”即所谓“超级智能”(Super-intelligence)。显然,数字乌托邦主义者的生命3.0的三级跳,目前恐怕还在第一级HAI/AGI/SAI的水平上。甚至在这个水平上也是问题多多。仅仅是塞尔的中文房间证据(Chinese Room Argument, CRA)对强人工智能的打击,恐怕他们还没有回过神来。^①



图3 生命3.0及人工智能等级图

4. 关于人工智能与人类智能

虽然强人工智能遭到塞尔等心智和语言哲学家的驳斥,但数字乌托邦主义者仍然乐此不疲。

早在1965年,英国密码学家、统计学家、计算机先驱(Computer Pioneers)古德(Irving Good)就提出“智能爆炸”理论,他说:

让我们给“超级智能机器”(Ultra-intelligence)下一个定义,那就是:一台能超越任何人(无论这个人多么聪明)的所有智力活动的机器。由于设计机器也属于这些智力活动中的一种,因此,一台超级智能机器就能出更好的机器;那么,毫无疑问会出现一种“智能爆炸”(Intelligence Explosion),到那时,人类的智能会被远远甩在后面。于是,第一台超级智能机器就会成为人类最后一个发明,只要它足够驯良,并告诉人类如何控制它就行。^②

数字乌托邦主义者认为,数字生命是宇宙进化自然而然、令人期待的下一步,如果我们让数字智能自由地发展,而不是试着阻止或奴役它们,那么,几乎可以肯定地说,结果一定会是好的。谷歌公司创始人兼CEO、数字乌托邦主义最具影响力的支持者佩奇(Larry Page)和贝宝(PayPal)公司联合创始人、SpaceX太空探索技术公司、环保跑车特斯拉公司以及SolarCity三家公司的CEO马斯克(Elon Musk)曾

^① 蔡曙山《哲学家如何理解人工智能》,《自然辩证法研究》2001年第11期;蔡曙山《关于哲学、心理学和认知科学的12个问题与塞尔教授的对话》,《学术界》2007年第3期。

^② M.泰格马克《生命3.0》,第2页。

经展开一场关于人工智能未来的辩论。佩奇说“如果生命会散播到银河系各处甚至河外星系,那么,这应当以数字生命的形式发生。他最担心的是,人们对人工智能的猜疑和妄想会延迟这个数字乌托邦的到来,而且可能会导致邪恶的人工智能发动军事叛乱,接管人类社会,违背谷歌‘不作恶’的座右铭。”马斯克不能接受这种激进的观点,要求他提供细节。而佩奇则抱怨马斯克是“物种歧视”。^①

以上这些大人物,计算机技术、人工智能和IT业界翘楚、自然科学家大都对超级智能信心满满。特别是2016年3月9日谷歌DeepMind公司的AlphaGo击败世界顶级围棋高手李世石,而且在此后的一年内,一个更强大的AlphaGo与全世界最顶尖的20位棋手对弈,没有一次失败。这时,就连通常人都相信,超越人类智能的人工智能已经来了。新闻界甚至惊呼:人类的历史即将结束,计算机和人工智能将统治人类!

为何如此恐慌?这得从围棋的基本原理讲起。围棋是一种围地的智力游戏,对弈双方在一个 19×19 矩阵的棋盘上做争夺地盘的游戏。这个中国人发明的智力游戏规则非常简单:双方一人一手,围地多者为胜。围棋盘上有多少地呢?答案是 $19 \times 19 = 361$,即以围棋盘中的一个交叉点为计算单位,称为子或目。围棋对弈一人一手,如何才能围住更多的地呢?双方就要在死活之中进行拼杀,既要杀死对方的棋子,又要让自己的棋活,同时还要围住更多的地。一盘棋究竟有多少种可能的变化呢?答案是 3^{361} 之多,据说这个数字超过了宇宙中所有原子的总数。^②所以,围棋是一种原理简单、变化复杂的智力游戏,所谓“千古无同局”。所以,对人类而言,围棋不仅是数学和逻辑推导,更是一种艺术。在人的计算能力的范围内,下围棋主要运用的是计算和推导,超出这个范围,人会运用经验和直觉。围棋高手在这两个方面都具备超出常人的能力。计算机程序则不同,它完全依靠精确的计算,AlphaGo还依靠神经网络学习和大数据(以往的棋局)搜索来提高自己的判断决策能力。在AlphaGo与李世石的这场世纪对决的第二局,第37手AlphaGo走出出人意料的一招。现场评论员和中外棋手认为这手棋是业余棋手的水平,场外观棋者调侃说,计算机是否出了问题,芯片是否被人拔了?现场评论断定黑棋已经处于完全的劣势,甚至可以说已经崩盘。现场评论说,如果在这种情况下黑棋还能赢,人类的思维会完全崩溃,可以说人类完全不懂围棋。岂料到第50手以后,左下角战火燃烧,黑棋的发展正好与第37手布下的那颗黑棋取得呼应,正是这手棋让AlphaGo最终赢得比赛,铸就它对李世石的4:1胜局。

AlphaGo制胜的这一招完全超乎人类的理智和想象,为人工智能注射了强心针,使世界陷入恐慌。

对于AlphaGo下出的这一手棋,泰格马克解释说“我将‘直觉’和‘创造力’视为人类的两个核心特征,现在,我要向你解释,AlphaGo展现出了这两种特征。”^③其实,AlphaGo走出的这一步,依靠的并不是如同人类的直觉和创造力,而是比人类更快、更准确的计算能力,以及依靠神经网络学习和大数据搜索而获得的比人类更强的判断决策能力。

结论和更加深入的讨论

总结全文,我们得出几点结论,并做更加深入的讨论。

1. 生命是自然进化的产物,人造生命违背了自然法则

自然是与文化对立的范畴。世界上只有两类事物,一曰自然,二曰文化。^④文化就是人化,是人所创造的一切东西。因此,自然是人类出现以前一切存在的总和,这个存在一直延续到现在,并且还会延续到将来,只要人类不去破坏它。

生命是自然进化的产物。对进化(演化)的解释,迄今有三种理论:达尔文进化论、基因综合进化论

① M.泰格马克《生命3.0》,第41页。

② M.泰格马克《生命3.0》,第117页。

③ M.泰格马克《生命3.0》,第117页。

④ 蔡曙光《自然和文化——认知科学三个层次的自然文化观》。

和心智进化论,详见前文。目前存在的生命形式也有三种:自然进化的生命,克隆的生命和克莱格·文特无中生有的生命。

按照美国生物学家魏泰克(R. H. Whittaker, 1924—1980)提出的生物学分类,自然进化的生命包括原核生物界、原生生物界、植物界、真菌界和动物界五界的生物。这是一个比较完整的纵横统一的分类系统,在纵的方面显示了生物进化的三大阶段,即无细胞生物→原核单细胞生物→真核(单细胞→多细胞)生物。在横的方面显示了生物演化的三大方向,即光合作用的植物、吸收营养的真菌和自己摄食的动物。在五界生物中,处于进化最高端的是动物,人类又处于动物进化的最高端。

非自然进化的生命包括克隆的生命和合成的生命两种形式。克隆的生命是指由单个祖先个体经过无性繁殖而产生的其他个体。20世纪初,韦伯(H. J. Webber)首先提出“克隆”的概念。1938年,德国科学家首次提出了哺乳动物克隆的思想。1963年,英国生物学家和遗传学家霍尔丹(J. B. S. Haldane)在题为“人类种族在未来两万年的生物可能性”的演讲上采用“克隆(Clone)”的术语。霍尔丹说过一句著名的话,他说“现在,我的怀疑是宇宙不但比我们所假想的要奇异,而且比我们能假想的还要奇异。”^①1996年,体细胞克隆羊“多利”出世后,克隆成为世人关注的焦点。今天克隆人已经不是科幻小说里的梦想,而是呼之欲出的现实。由于克隆人可能带来复杂的后果,一些生物技术发达的国家,现在大都对此采取明令禁止或者严加限制的态度。尽管克隆生命已经突破了自然进化的界限,但仍然是从生命到生命的复制。

克莱格·文特的合成生命不仅超越自然进化的界限,而且超越无生物与生物的界限,超越非生命与生命的界限,在他的实验室用化学物质直接合成生命——这是“从无到有”的生命的合成。人类在“超越自然”的路上越走越远。^②似乎已经没有人能够阻挡人类背离自然、自己充当上帝的愿望和行动。

人工智能也要加入生命的行列,他们已经为这个将要出生的生命想好了名字——生命3.0。泰格马克说“计算和信息一样,是独立于物质层面而存在的:计算就像拥有自己的生命一样,与它采取什么样的物质形态无关。”他又说“硬件就是物质,软件就是形态。计算的‘物质层面的独立性’暗示我们,人工智能是可能实现的:智能的出现并不一定需要血肉或碳原子。”^③

就连著名物理学家霍金也大胆预言,人类不但会于未来100年内在其他星球建立新的生存基地,而且还会在此后的一个百年内或更早些,利用基因科技改良现有入种。

这里就提出了一个严肃的问题:人类是自然进化的产物,人类发展到今天,人需不需要再遵循自然?或者人类已经强大到可以凌驾于自然之上并可以为所欲为?

这两个问题涉及一个前提:自然的发展和演化是否有它自身的规律?这个规律人类是否需要遵从?对这个问题的回答,应该向东方的智者寻找答案。荀子说“天行有常,不为尧存,不为桀亡。应之以治则吉,应之以乱则凶。强本而节用,则天不能贫;养备而动时,则天不能病;修道而不贰,则天不能祸。故

① J. B. S. Haldane, “My Own Suspicion is that the Universe is not Only Queerer than We Suppose, but Queerer than We Can Suppose”, https://encyclopedia.thefreedictionary.com/John+Burdon+Sanderson+Haldane#cite_note-66

② 2007年10月6日,克莱格·文特的研究小组用化学物质合成了由381个基因、58万个碱基对组成的人造染色体,并将其植入细菌生殖支原体的外壳中,在这些基因的控制下,新细菌能摄食、代谢和繁殖,堪称人类历史上第一个“人造生命”。2008年初,文特将研究结果发表于《科学》(Science)杂志,宣称他们已成功制造了一种支原体的基因组,完成人造生物的关键一步。他们研究的这种支原体拥有485个基因、58万对碱基,是已知的基因组最小、最简单的生命形态。2010年5月20日,美国科学家宣布世界首例人造生命诞生,命名为“辛西娅”(Synthia)。时代杂志在2000年7月将文特与人类基因组计划代表弗兰西斯·柯林斯同时选为封面人物,又在2007年将他选进世界上最有影响力的人之一。文特的这项工作,开创了化学和生命科学的一个新的重要领域——合成化学(Synthetic Chemistry)。“Synthia”一词源于英文“Synthesis”,意为“综合”,故“Synthia”音译为“辛西娅”,可意译为“综合妹”或“合成妹”。对于文特的工作科学界褒贬不一,赞成者认为应该授予他诺贝尔奖,反对者认为他打开了潘多拉魔盒,突破了科学伦理的底线,是一个“科学坏蛋”(A Bad Man of Science)。——作者注

③ M.泰格马克《生命3.0》,第86—88页。

水旱不能使之饥渴,寒暑不能使之疾,妖怪不能使之凶。”^①孔子曰“君子有三畏:畏天命,畏大人,畏圣人之言。小人不知天命,而不畏也。”^②中国古代思想家和哲学家老子更加明确地说“人法地,地法天,天法道,道法自然。”^③在中国古代先贤的思想中,自然是至高无上的,自然的演化有自己的规律和法则,这些规律和法则是以人的意志为转移的,人只能遵循自然的法则。一切非自然的东西包括人造生命,都是没有存在的理由的。可惜,东方智者的这些思想并未被西方科学家所接受,他们似乎沿着相反的方向,越走越远。

2. 心智和意识是人类生命的唯一标准,不存在其他的标准和定义

为了更深入地讨论人工智能,我认为首先要认识生命的本质和人类的智能,因为人工智能不过是对人类智能的模仿。

按照前述泰格马克的定义,生命是满足以下公式的存在:

生命 = 硬件 + 软件

= (碳基/硅基/其他物质基础) + 计算

在他们看来,生命的物质基础并不重要,生命的基础既可以是碳基,也可以是硅基。^④

我们没有任何理由和证据否定除了碳基之外的其他物质基础的生命存在,例如,著名生物化学家和科幻小说家阿西莫夫(Isaac Asimov)提出了六种生命形态:(1)以氟化硅酮为介质的氟化硅酮生物;(2)以硫为介质的氟化硫生物;(3)以水为介质的核酸/蛋白质(以氧为基础的)生物;(4)以氨为介质的核酸/蛋白质(以氮为基础的)生物;(5)以甲烷为介质的类脂化合物生物;(6)以氢为介质的类脂化合物生物。其中第三项就是我们熟悉的和唯一认识的生命。第一、第二项是一些高温星球上可能存在的生命形式。另外,地球上曾经出现过的那些生活在硫矿里的、厌氧的古细菌就很有可能是以硫作为自己生命的介质;而第四项至第六项,则是一些寒冷星球上可能存在的生物形态。

碳建立了地球上所有已知生物的基本架构。地球生命之所以是碳基的,一个重要原因是一个碳原子最多能够同时和四个其他原子相结合。碳的这一特性使其十分适合形成长分子链——比如蛋白质和DNA。迄今为止,尚未发现硅基生物,因为在地球环境中,硅的活性不如碳,而硅基生物极有可能是厌氧的,所以一般认为充满氧气的环境下无法诞生硅基生命。

泰格马克试图将超级人工智能纳入生命的系列,即生命3.0,只不过这种生命不是碳基的,而是硅基的。但迄今为止我们这个星球上并未出现自然进化系列中的硅基生命。因此,即使我们接受生命3.0为一种生命形式,它也不是自然进化的生命,而是人造的生命。

下面需要讨论的是,生命3.0到底是不是一种生命形式。

很显然,泰格马克的生命定义不能作为区别生命和非生命的标准,因为它不符合我们迄今为止所观察到的生命现象和自然演化的生命的历史。我们即使接受生命3.0作为一种人造的生命,但它与前面提到的“克隆的生命”和“合成的生命”相比,其与“自然生命”的血缘更远,甚至可以说两者无任何血缘关系。泰格马克的“生命3.0”不过就是人的一个新的创造物,人创造了机器,包括火车、汽车和飞机,但这些机器都不是生命。那么,泰格马克的“生命3.0”为什么会是生命呢?生命和非生命的界限究竟在哪里?——我们总得有一个标准来区分生命和非生命,特别地,生命形式中我们还要区分人和非人类的动物。那么,这个标准究竟应该是什么呢?

我们可以从人和动物的区别倒推回去。毕竟,人首先要区分的是人这个物种与其他物种的差异,这当然首先要区分人与最相近的物种——非人类动物的差异。这在人类认识史上有太多的标准和表述,我们在此论证的是人类认知五层级理论所给出的人类心智进化的标准。

① 荀子《荀子·天论》,中华书局2011年版,第362—363页。

② 孔子《论语·季氏第十六》,载《国学十三经》(四),线装书局2006年版,第196页。

③ 老子《道德经》第二十五章,载《国学十三经》(六),线装书局2006年版,第271页。

④ M.泰格马克《生命3.0》,第88页。

根据人类认知五层级理论,人与非人类动物的本质区别是语言,以及在语言基础上产生的思维、语言和思维建构的知识系统,以及知识积淀形成的文化。因此,我们可以根据心智从低级到高级的发展来区分人和非人类动物,同时也用是否具有心智来区分动物与非动物。

从自然进化的方向看,心智从低级到高级的发展形态是:神经层级的心智、心理层级的心智、语言层级的心智、思维层级的心智和文化层级的心智。非人类动物只具有神经和心理两个层级的心智,人类则具有所有五个层级的心智,而抽象的概念语言、思维和文化三个层级的心智仅为人类所特有,称为人类心智(Human Mind)。^①这样,我们就建立了人与非人类动物区分的可靠的科学依据和标准。

从心智对行为的控制方式上看,心智有两种基本的方式:意识和无意识。它们犹如两个舵,掌控着心身的活动。因此,心智和意识(包括它的深层形式无意识)也就成为生命和非生命的标准,此外我们不可能有其他的关于生命的标准和定义。

心智和认知、心智和身体、意识和无意识,它们之间的关系既是历史久远的哲学的基本问题,也是21世纪新兴、前沿的认知科学的基本问题。对意识和心身关系这些非常深刻的问题的讨论,已经大大超出本文的范围,有兴趣的读者可以参阅笔者发表的《认知科学与技术条件下心身问题新解》。^②

3. 迄今为止,人工智能只是人类智能的一种表现形式

要讲清人工智能与人类智能的关系,首先要正确定义“智能”。

泰格马克的定义是:智能(intelligence)是完成复杂目标的能力。^③

此定义中,定义项中“复杂目标”这个概念的内涵是不清楚的,因此,无法揭示被定义项“智能”这个概念的内涵。泰格马克的定义是为了将“智能”赋予非生命的人工智能,从而完成他的“超级智能”(Super intelligence)到“普遍智能”(universal intelligence)、“通用智能”(general intelligence)和“生命3.0”的建构。对此的质疑,详见前文。

对智能的定义,我们还得回到心智和意识这两个最基本的概念。因为人工智能是对人类智能的模仿,所以,我们应该首先定义人类智能:

人类智能就是人类的心智能力,它表现为五个层级的心智和认知能力。

根据这个定义,我们将人工智能与人类智能的对比用表2来表示。

表2 基于人类心智和认知五层级的人工智能与人类智能对照表

水平 AI	I 完全无能	II 低于人类	III 人类水平	IV 超过人类	V 远超人类
心智					
文化层级的心智和认知	艺术、科学、哲学、宗教	格律诗词、书法、绘画			
思维层级的心智和认知	创造性思维、溯因、隐喻	归纳、类比、定理证明	演绎推理	深蓝、AlphaGo	
语言层级的心智和认知	语用加工、隐喻	语义加工、机器翻译	句法加工、机器翻译		
心理层级的心智和认知	直觉、顿悟、经验	情感、情绪、联想	传感器、五感	记忆	
神经层级的心智和认知*	自我、无意识、梦境	高级神经活动	机械行为、机器人	精密行为、自动化生产线	大型设备控制、运输

* 注:神经层级的心智和认知包括身体和行为层级的认知和行为方式。

① 蔡曙山《论人类认知的五个层级》。

② 蔡曙山《认知科学与技术条件下心身问题新解》,《学术前沿》2020年5月(上)。

③ M.泰格马克《生命3.0》第51、67页。

这是一个5×5的矩阵,纵向(Y轴)从下到上表示人类心智的5个层级,横向(X轴)表示人工智能目前和将来可能的水平,25个方格中是目前人工智能已经存在的领域。此表清楚地显示了目前存在的人工智能与人类智能的差距,我们可以看出:

第一,在人类心智的5个层级和5个水平领域内,人工智能完全无能和低于人类水平的领域共有10个,占40%;相当于人类水平的领域4个,占16%;超过和远超人类水平的领域4个,占16%。结论:当前的人工智能远未达到人类心智水平,超过人类心智水平的仅有少数几个领域。

第二,人工智能超过人类心智水平的往往只是单一的领域。例如,战胜人类顶级棋手的“深蓝”(1997)和AlphaGo(2016)仅仅是在思维认知这个层级中大规模数据的确定信息推理这个领域,在其他推理领域如归纳、类比和溯因这些小数据和需要经验参与的不确定推理的领域,人工智能目前未显优势;在定理证明、隐喻和创造性思维等领域,人工智能基本不能发挥作用。

第三,从人类认知的5个层级看,人工智能目前主要分布在低阶认知即非人类动物的认知领域中(图中虚线左下方),其发展水平大多数低于人类和非人类动物的水平。仅在行为层级的某些方面如自动化生产线、大型设备控制、运输等“拼体力”的领域超过人类水平。

第四,从高阶认知即人类认知层级看,除“深蓝”和“AlphaGo”在大运算量确定信息的推理领域独树一帜外,在其他所有领域内人工智能均远逊于人类智能。这就是为什么AlphaGo战胜李世石能够引起恐慌,因为这是人工智能首次在人类认知领域获得胜利。人类认知是以语言为基础,以思维为特征的,“我思,故我在”。现在,在这个自有文明史以来人类独具优势并借以成为“万物之灵”的领域,竟然输给了机器。所以,围棋界引起恐慌,新闻界引起恐慌,并且推波助澜地宣称“人类的历史已经终结”“计算机将统治人类”。但是,正是这个在人类认知单一领域千年难得的一次胜利,恰恰证明人工智能超越人类为时尚早,统治人类则完全谈不上。即使有一天人工智能在以上25个领域中完胜人类,人工智能超越人类、战胜人类、统治人类这样的结论仍然是不可靠的。

第五,在“单兵较量”方面,人工智能与人类智能仍然差距甚远。单一主体的人类智能可以涵盖以上全部25个领域,而单一主体的人工智能迄今为止只作用于单一领域。例如,一位职业棋手除了能够夺得世界冠军,可能他还会弹钢琴、作格律诗、写书法,而AlphaGo除了会下围棋,其他一无所能。所以,人工智能是单一的智能,人类心智则是综合的智能。因此,可以得出结论:单一认知主体的比较,人类综合智能远胜于人工智能。这一结论还有一个重要的推论:就人类发展而言,人与机器最本质的差异是综合,人类的知识是综合性知识,21世纪是综合的世纪,21世纪需要的是知识交叉综合、全面发展的人才。^①

第六,除了文化认知和思维认知的差异,在人类认知的基础语言认知这个层级上,人工智能与人类心智也存在巨大差异。首先是语言系统的差异,人类使用的是在自然进化中产生的语言,称为自然语言;计算机和人工智能使用的语言是人类为它设计的语言,称为人工语言。这种语言只有两个初始符号:0和1。在这个语言系统中,一切的表达式包括词语、语句和元语句(包括计算机指令),都只能用这两个初始符号来表示,否则计算机不能识别。^②我们把这样一个以形式符号为基础,语形和语义都严格形式化的人工语言系统称为形式语言系统。由于自然语言与人工语言差异巨大,因此,在人机之间,有一个语言理解和语言加工的问题。人脑的语言加工分为语形加工(包括词法加工和句法加工)、语义加工和语用加工。^③在语言加工方面,目前的人工智能在句法加工方面差强人意,在语义加工方面困难重重,在语用加工方面则是一筹莫展。例如,汉语是典型的语用语言,汉语的完整意义必须在说者、听者、

^① S. Cai, “The Age of Synthesis: From Cognitive Science to Converging Technologies and Hereafter”, *Chinese Science Bulletin* (《科学通报》英文版), 2011, 56: 465 – 475.

^② 蔡曙山《言语行为和语用逻辑》,中国社会科学出版社1998年版,第347—354页。

^③ 蔡曙山《自然语言形式理论研究》,人民出版社2010年版,第1—3页;蔡曙山《符号学三分法及其对语言哲学和语言逻辑的影响》,《北京大学学报(哲学社会科学版)》2006年第5期。

时间、地点和语境这五大要素的交互作用中来理解,李白的“望庐山瀑布”、杜甫的“壮游”、王维的“山居秋暝”、辛弃疾的“千古江山”、李清照的“常忆溪亭日暮”、王实甫的《西厢记》、汤显祖的《牡丹亭》、曹雪芹的《红楼梦》——这都不是机器能够理解和翻译的。根据塞尔的“中文房间”人工智能模型,计算机的语言加工和翻译,完全没有也不需要语言理解,而仅仅是在两个语言系统之间建立对应关系,并作“能行”加工,即机械的,并在有穷步骤内完成的加工。语言是人类心智和认知的基础,“我言,故我在”。^①人类这种独特的语言存在方式,以及在此基础上的思维存在方式和文化存在方式,恐怕是机器永远也无法实现的。

综上所述,我们可以将通用人工智能定义为在人类认知即语言、思维、文化三个层级全面超过人类的人工智能。很显然,这种理想的人工智能远远没有实现。迄今为止,人工智能只是对人类智能的模仿,是人类智能的一种表现形式。正如谷歌母公司 Alphabet 董事长埃里克·施密特在 AlphaGo 和李世石的世纪大战之前所说“无论最终结果是什么,赢家都是人类。”^②

那么,通用人工智能这个人工智能的“圣杯”是否真的有一天会到来呢?是否有一天人工智能会全面超过人类智能,成为无所不能的最高智慧呢?答案是否定的。我们只需做一个简单的理想实验:

现在就停止发展人工智能又如何?

如果今天就停止人工智能的研究,超过人类智能的人工智能,包括“超级智能”、“普遍智能”、“通用智能”和“生命 3.0”统统都成为不可能。当然,人类也可以选择继续发展人工智能,直至某一天超级智能的出现。即使如此,亦如施密特所说,无论最终结果是什么,赢家都是人类。

所以,一切都在于人(计算机专家和科学家最容易产生的想法)。那么,人是万能的上帝吗?

4. 人永远不能也不可能充当上帝,更不可能是宇宙万物的主宰

在人工智能的讨论中,涉及的不仅仅是技术问题。人工智能问题的正确答案不能仅仅从技术上寻找,而应该从计算机、人和自然的关系中去寻找。

人工智能和人的关系、人工智能和自然的关系我们在前面已经做了深入探讨,下面我们讨论人与自然的关系。

人与自然之间的关系,最终是人的行为与自然的关系,因为自人类出现以后,人类的行为已经深刻地影响了自然。人对自然的影响,可以从文化与自然的关系来分析,因为文化就是人类行为的结果的总和。

文化分为三个层次:科学、哲学和宗教。在科学这个层次上,由于自然科学以自然为研究对象,所以,大多数科学家容易相信自然是可以认识的,是可以控制的,是可以超越的,甚至是可以任意改变的。例如,前面提到的三种非自然进化的生命如霍尔丹的克隆的生命、文特的合成的生命和泰格马克的人工智能 3.0 的生命。在这些生命的形式中,克隆的生命、合成的生命和人工智能的生命都是人类的创造,都属于文化的范畴而非自然的范畴,它们都是人类文化影响自然、超越自然、改变自然的例证。

也有另外一些自然科学家,他们认识到自然的伟大及自然后面力量的伟大,从而产生对自然和上帝的崇拜。一个典型的例子是 17—18 世纪杰出的物理学家、数学家、天文学家、自然哲学家牛顿(Isaac Newton, 1643—1727),他既是“科学史上最有影响力的人”,^③同时也是一名虔诚的基督徒。牛顿认为需要从两个途径来认识上帝,一是对自然的研究,二是对圣经的学习。作为一名科学家,他认为对自然的研究只是认识上帝的一种途径。牛顿临终前说“我的工作和神的伟大创造相比,我只是一个在海边拾取小石和贝壳的小孩子。真理浩瀚如海洋,远非我们所能尽窥。”牛顿以后最伟大的物理学家爱因斯坦是一位自然神论者,他相信斯宾诺莎的上帝,即隐藏在自然规律后面的上帝。他说“在感情上深信,在

① 蔡曙山《论语言在人类认知中的地位和作用》。

② 《AlphaGo 最终局战胜李世石,人机大战总比分 1:4》,新浪科技,2016 年 03 月 15 日, <http://tech.sina.com.cn/it/2016-03-15/doc-ifxqhmv2486669.shtml>

③ 王军礼《谁是科学史上最有影响力的人》, <http://blog.sciencenet.cn/blog-1273577-935464.html>。

无法理解的宇宙中存在一种至高的理性力量,便是我的上帝观念。”他解释道“任何严肃追求科学的人都会越来越确信,有一种精神呈现于宇宙定律之中。这种精神远远超越了人的精神。面对它,力量有限的我们必定会感到谦卑。于是,对科学的追求会导向一种特殊的宗教感情。”他断然否定自己是无神论者,他说“宇宙和谐的秘密无法获得,由此产生的绝对谦卑感将我同与大多数所谓的无神论者区分开来。”他说“科学只能由那些全心全意追求真理和向往理解事物的人来创造。然而这种感情的源泉却来自宗教领域。”^①

哲学家对文化与自然的关系的认识更加深刻,中国古代思想家、哲学家老子在他的哲学中,把“道”这个最高范畴置于“自然”的约束之下。关于老子哲学的自然文化观,可以参看笔者《自然与文化》一文。^②西方哲学家中,斯宾诺莎是自然神论的典型代表,他认为自然和神是统一的,因为上帝就是宇宙,它包括了宇宙的物质和精神世界。他认为人类如果被任何的外界事物所影响,那么他都会成为它的奴隶,并且永远都得不到自由。只有当人类可以和上帝达成共识,不再被这些因素所累时,才会有相对的自由。马克思曾经警告说“人类计划假如不把自然规律作为依据,只会带来更大的灾难。”^③恩格斯则说“我们不要过分陶醉于我们人类对自然的胜利。对于每一次这样的胜利,自然界都对我们进行了报复。”^④

今天人类面对自然,应该如伟大的爱因斯坦所说,谦卑地低下自己的头,而不能继续像过去几百年来那样凌驾在自然之上,无休止地向自然索取,奴役自然,甚至改变自然,否则就可能受到大自然严厉的报复。人永远不能也不可能充当上帝,更不可能是宇宙万物的主宰。

Life Evolution and Artificial Intelligence

—Questions about Life 3.0

CAI Shushan

Abstract: After AlphaGo defeated Lee Sedol in 2016, the myth that artificial intelligence defeated humanity has become a reality. In 2017, Max Tegmark's odd book Life 3.0: Being Human in the Age of Artificial Intelligence gives life to artificial intelligence. In 2018, Stephen Hawking warns of the development of artificial intelligence, and Hawking holds that the thorough development of artificial intelligence may lead to the demise of mankind. These “big questions” should find answers from the evolution of life, the development of the human mind and artificial intelligence, and the relationship between man and nature. Based on Darwin's species evolution theory and modern comprehensive evolution theory (genetic evolution theory), and the five-level theory of human cognition, this paper created the theory of mental evolution, and questioned Life 3.0. Finally, it is concluded that life is a product of natural evolution, and artificial life violates the laws of nature; mind and consciousness are the only standards of life, and there are no other standards and definitions. So far, artificial intelligence is only a manifestation of human intelligence. Human beings can never and cannot act as God, nor can they be the master of everything in the universe.

Key words: artificial intelligence, Life 3.0, evolution of life, human mind, theory of mental evolution

(责任编辑: 苏建军)

① 沃尔特·艾萨克森《爱因斯坦的上帝》,见《爱因斯坦传》,张卜天译,湖南科技出版社2014年版,第338—346页。

② 蔡曙山《自然与文化——认知科学三个层次的自然文化观》。

③ 马克思《资本论》(第3卷),第251页。

④ 恩格斯《自然辩证法》,第284页。