

从记忆与智能的关系看人工智能的发展

杨庆峰

(复旦大学 应用伦理研究中心, 上海 200433)

[摘要]在记忆与认识的关系上,存在着记忆附属论的观点。这一观点主要表现为:记忆是知识的来源、记忆是认识的低级阶段和记忆是知觉的滞留。在“知识驱动”和“数据驱动”的人工智能的代际话语中,记忆附属论依然有效,即记忆附属于智能。知识路径则突出了记忆在知识形成中的作用,记忆则被看作是信息的存储和编码;数据路径则偏重智能体对相关数据的提取,突出的是记忆的提取环节,记忆被看作是数据的存储和提取。从哲学角度看,附属论观点的存在将限制人工智能代际的理解。如果能够摆脱记忆附属论的观点,“经验驱动”作为第三代人工智能的特征呼之欲出。

[关键词]记忆;智能;附属论

[中图分类号]N03 [文献标志码]A [文章编号]1672-934X(2020)02-0001-09

DOI:10.16573/j.cnki.1672-934x.2020.02.001

The Development of Artificial Intelligence from the Perspective of the Relationship between Memory and Intelligence

YANG Qing-feng

(Institute of Applied Ethics, Fudan University, Shanghai 200433, China)

Abstract: There exists the memory attachment theory when explaining the relationship between memory and cognition, which can be mainly interpreted that memory is a kind of knowledge source, the lower stage of recognition, and the retention of perception. In the intergenerational discourse of "knowledge-driven" and "data-driven" artificial intelligence, attachment theory is still effective, which means that memory has attached to intelligence. Knowledge path highlights the role of memory in knowledge formation, and memory is regarded as the storage and encoding of information; data path lays particular stress on the extraction of relevant data by the agent, highlighting the extraction link of memory, which is regarded as the storage and extraction of data. From the philosophical perspective, the attachment theory will limit the intergenerational understanding of artificial intelligence. If the dependency theory of memory can be abandoned, "experience-driven" will emerge as a feature of the third generation of artificial intelligence.

Key words: memory; intelligence; attachment theory

按人工智能代际划分,我们正处于 AI2.0 中。AI1.0 和 AI2.0 的特征已经被清晰地勾勒出来,“基于符号推理的知识驱动、基于深度学习的数据驱动”(张跋),但是对 AI3.0 特征的概

收稿日期:2020-01-06

基金项目:国家社科基金重大项目(17ZDA028)阶段性成果;中科院学部“大数据伦理问题及其治理研究”阶段性成果

作者简介:杨庆峰(1974-),男,陕西白水人,教授,博士生导师,美国达特茅斯学院、斯维本科技大学访问学者,主要从事技术哲学、数据伦理、记忆哲学与人工智能伦理等研究。

括还是比较模糊。要应对这一问题,记忆哲学将会是一个很好的分析途径。^①当我们从记忆哲学视角看,记忆与智能的关系就成为一个基本问题。这一问题的澄清有助于我们理解人工智能的发展趋势。但是由于记忆附属论观念的影响,让这一步变得无比艰难。本文试图从记忆与智能的关系出发,澄清走向真正智能的困境以及勾勒 AI3.0 的哲学特征。

一、“真正智能”实现的困境

如果从智能承载者来讨论智能^②的分类,有三种智能形式:人类智能、动物智能与机器智能。^③人类智能与机器智能的关系也越来越被人工智能学界关注,超越问题就成为人文学界和科学界共同担心的问题。人文学界多讨论机器智能超越人类智能的不良后果;科学界多讨论机器智能是否可能以及如何发展。超越问题的本质是真正的智能何以可能。在这一问题上存在着观念的困境即在人工智能理解中,记忆附属论观念极大地影响了真正智能的实现。“记忆附属论”即记忆隶属于智能,即记忆被看作感知信息的保留、智能生产的低级阶段、智能实体的基本构成部分等形式。在具体阐述之前,我们首先需要梳理一下哲学与 AI 之间的矛盾态度。

在 AI 领域,我们能够感受对形而上学的矛盾态度。一方面是强烈的拒斥态度。这种拒斥主要表现为反对智能的形而上学定义,并且通过技术来实现智能的形式。当我们回到 AI 源头图灵的时候,就会发现他身上表现的非常明显。一般都认为图灵的重要性是提出了“机器是否会思考”的问题,如果是这样,图灵并不拒斥形而上学,而是提出了形而上学的问题。但是这种假象被后来的方法所消除。图灵采取了实验测试的方法来解决这一问题,后来被称为图灵测试。^④这种基于可操作原则的方法得到了这个领域的基本认可。从哲学角度看,这种

基于可操作原则的方法恰恰是其拒斥形而上学的表现。还有一种形式从行为主义角度为智能寻求一种根据,即通过机器与环境的相互关系来阐述智能的实用定义。这种做法导致了智能体(agent)新概念的出现,宝拉·博丁顿(Paula Boddington)将人工智能伦理的与能动性(agency)联系在一起则符合了这一规定性。“AI 与伦理学紧密地与能动性这一基础问题联系在一起。”^[1]这两个过程显示了对于形而上学智能概念的完全反思。另一方面是强烈的接受态度。联结主义者通过神经科学的成果来探讨人工智能的物质机制,他们借助神经科学的新成果不断取得的突破。比如 DeepMind 在走迷宫的问题上,得出了一个惊人的结果:在空间记忆上,人工智能体和人类以及动物表现出相似的神经元结构。^⑤通过这一成果,能够借助为机器智能的可能性做出科学辩护。还有大卫·芒福德(David Mumford)从意识生成的过程探讨意识生成的可能性,比如从情绪与意识的角度来探讨意识生成的可能性,甚至人工智能科学家开始关注情绪的可计算性问题。另外一位就是 S·罗素(Stuart J. Russell)他在讨论智能体的时候,给我们做出了哲学式的规定,在他看来,人工智能是理性主体(rational agent)。“在本书中,我们采取了这样的观点:智能主要与理性行动有关,一个智能能动体在一个情景中采取最好可能的行动。我们研究了这个意义上被构造为智能的智能体问题。”^[2]当这个定义被给出的时候,我们很容易联系到康德。在很多解释中,康德被看作是“理性主体”概念的提出者。在这里我们看到了二者在行动意义上的高度吻合,只是不同也是非常明显的,人工智能的行动是与情境构成一对关联体;而康德的行动则是与主体构成一对关联体,这一传统造就了德国古典哲学。

上述悖论态度显示了关于机器智能讨论的摇摆不定。机器智能是否仅仅是准人类智能?

能否成为真正的智能?强烈的拒斥态度意味着机器智能仅仅是人工智能,与自然智能不同,所有的发展都是在“拟”一“准”的框架中前行;强烈的接受态度意味着为机器智能找寻到坚实的哲学根据。我们把这一问题转化成“真正的智能”,即机器智能成为真正的智能何以可能?

目前,科学家正在通过五种主要方式来实现机器智能。第一种方式即让机器表现出与人类智能一样的行为,也就是“像人一样思考或者行动”^[2]。⑥这条路径导致了让机器能够像人一样感知、决策/判断和预测。第二种方式即让机器智能变得通用起来,能够做到迁移式学习,能够做到举一反三。第三种方式即利用人类与机器神经机制的同构性特征,让机器产生智能行为,比如在空间记忆行为上体现出的同构性特征。第四种即探讨智能行为的神经机制,如在人工神经元的基础上构建深度学习的机制,从而产生相应的行为。第五种方式即探讨智能生成的深层根据,比如情绪是意识的基础条件。

在这五种方式中,第一、二种是表象式的,即表现为与人一样类似的行为或者智能构成;第三、四种是在表象基础上,探讨其依靠的神经机制;第五种是智能生成的条件。在这五种方式中,最后一种应该说是接近了哲学的探讨,探讨智能得以可能的条件。芒福德的工作指出意识生成需要建立在情绪之上,所以这也是给出了一个基本的条件。但是在技术实现上却必须依靠因果关系,也就是让机器具备情绪可以导致机器具备意识。

在上述五条路径上都遭遇到了不同的困境。在表象式路径中,第一种已经遭遇到了塞尔用中文屋实验的驳斥,即机器无法做到理解语言的意义,因此无法说其具备智能。第二种在迁移式学习上遭遇到了灾难性遗忘的困境,这一困境使得机器的持续性学习和迁移性学习变得不可能,为此他们提出了多重路径来加以克服。第三种可能遭遇到的问题是概念的澄

清,比如何以判断不同物种空间记忆的神经元结构是相似的?第四种深度学习路径遭遇到的问题与第二种相同,即灾难性遗忘的问题。第五种路径在哲学上也会碰到困难,比如在什么意义上情绪成为意识的条件?这在哲学上根据并不充分。

上述路径的共同点是都牵涉到记忆因素。在机器智能实现的问题上,记忆附属论是一个基本的假设前提。所以,接下来我们要对记忆附属论这样一个假设前提做出分析。这种分析将表明,记忆附属论观念如何阻碍真正的智能的生成?笔者在《记忆、认知与记忆本体论》一文中专门分析认识论领域内记忆附属论的表现:记忆是知识的来源、记忆是认识的低级阶段和记忆是知觉的滞留^[3]。而记忆附属论在人工智能领域内表现出几乎相似的观念:记忆是信息内容的保留、记忆是智能生产的低级阶段、记忆是智能的组成部分。

二、记忆:信息内容的保留

从记忆的原初含义看,记忆最为基本的规定性是痕迹的保留。亚里士多德的指痕比喻就是这种规定性的最初源头,并影响了整个记忆观念史。直到当代哲学才出现了有效反驳,现象学家海德格尔用石头压在大地上产生的痕迹这一例子批驳了将记忆看作是痕迹的保持的观点;心灵哲学家库肯(Kourken Michealian)提出了“无内容的记忆”观念来反驳了内容的保留的传统观念。但是,哲学概念影响力的体现需要长时间才能表现出来的,所以,我们在人工智能领域内,依然能够看到科学家如何停留在传统的理解中难以抽身。我们仅从两个例子来说明这一情况。

一是在知识生产中。对AI的代际理解,第一代人工智能被看作是“以符号推理模型为基础的知识驱动为方法”(张跋,2019),获取知识被看作是第一代AI的重要目的之一。第二代

人工智能是“以深度学习为基础的数据驱动”为方法。很多学者的解释也是在此基础上进行的。在 S·罗素看来,人工智能能够做四件事情:解决问题、知识—推理—计划、学习、交流—感知—行动。在这个知识生产过程中,存在着三个环节:数据、信息和知识。此外,在对 AI 行为描述中,我们也能够看到知识论因素的存在。OECD 近期发布的《社会中的人工智能》报告指出,“正如 OECD 的 AI 专家组解释道,为了满足人类定义的目标群,一个 AI 系统是能够做出决策、建议或影响真实或者虚拟环境的决定;它使用基于机器和/或人类输入结果来感知现实的和/或者虚拟环境;从这样的感知中抽象出模式(用自动模式,如利用机器学习或者手工的方式);使用模式来形成指向信息或者行动的选择。一个 AI 系统被设计出根据不同的自动化水平来运行。”^[4] 如果从知识获取的角度来看,数据的存储和提取就表现为记忆过程,而这是知识获取过程的低级阶段。数据即感受环节,人工智能感受外界环境的刺激,从行为本身来说是感受过程;信息是有用数据的保持,这就变成了信息内容;在生产过程中,有用的数据成为信息内容保留下来。这个环节被等同于记忆。因为记忆被理解为信息的保留。知识则是有用信息内容被运用产生必要的结果。所以在这个过程中,记忆仅仅被看作知识生产的低级环节。这一观点与哲学认识过程极其相似。我们在传统的认识论模式中的感性、知性和理性就可以看到相似的过程。记忆发生在感性过程,感觉的滞留体现为记忆本身。

二是在机器视觉中。计算机视觉顶会 CVPR 主席德勒克·霍尔姆(Derek Hoiem)曾经讨论了计算机视觉的行为本质。在他看来,“计算机视觉只是记忆,而不是智能”。对于这一观点,笔者曾经去信询问过具体的意思。他做了简单的解释,深度网络是模式识别器。它们把新模式匹配到已知的模式中,转化相关信

息。有时候用方法做一些非常复杂的事情,比如产生图片说明或者估计对象的 3D 形状,好像它们能够“理解”几何或者场景的语言或者结构。在大多数情况下,如果我们没有认真设计,它们仅仅是提取相似的例子,它们的泛化能力就很差。模式识别(和记忆)是智能的重要组成部分,能够基于它自身的模式产生复杂的有效行为。很多看似能够解释图像几何的方法实际上只是在学习过程中记住了图像的几何信息,并通过检索与输入类似的样本来执行预测。预测得到的 3D 模型看似很好,但是这些方法无法泛化到新的形状和场景^[5]。对他这一观点进行哲学分析和阐述有助于超越问题的思考。德勒克的观点指出图像的几何信息保持无法泛化为新的形状和场景。“泛化”可以理解为通用过程,将某种特殊的、个体抽象为一般的、共性的东西。根据记忆理论,“信息的保持”仅仅是感性活动,从认识角度看,信息的保持是指信息内容的保留。而如何从保持飞跃到抽象的确是一个难题。这一问题需要在对深度学习的过程分析基础上才能够有效破解。“依赖硬核知识的系统面对的困难显示 AI 系统需要获取自己知识的能力,这个过程是通过从原始数据提取模式。这种能力就是机器学习。”^{[6](P2)}

这两个过程恰恰是对记忆附属论的一种支持,记忆被看作是信息内容的保持。而在这个问题上,能否产生通用智能就变得有疑问了。我们可以把这一过程看作是智能行为得以可能的前提性探讨了。此外,这一问题也与智能生产本身有着密切的关系。

三、记忆:智能生产的低级阶段和智能结构的基本元素

在人工智能历史中,记忆问题始终伴随着人工智能的研究,成为困扰人工智能研究的一个主要问题。我们从明斯基开始,他在讨论智能的时候提出的一个有趣观念——“智慧从愚

笨中来”,这很容易让我们想到黑格尔的“意识源自恐惧”^[7]的观念。1988年,明斯基在《心智社会》中讨论了智能的生成问题。他多次谈到了记忆问题,如第8章的“记忆理论”、第15章的“意识与记忆”。在讨论记忆理论的时候,他甚至援引了法国小说家马塞尔·普鲁斯特的话语,“确实,此时在我身上品味这种感受的生命,品味的正是这种感受在过去的某一天和现在所具有的共同点,品味着它所拥有的超乎时间之外的东西,一个只有借助于现在和过去的那些相同处之一到达它能够生存的唯一界域、享有那些食物的精华后才显现的生命,也即在与时间无关的时候才显现的生命。”^{[8](P86)}在第15章他引用了休谟的理论,“每个人都乐意承认,一个人感觉到过热而产生的疼痛或者温暖带来的愉悦与他事后回忆这种感觉或者通过想象预期这种感觉相比,思维所产生的知觉是完全不同的。”^{[8](P184)}这两处让我们看到,他已经意识到记忆对于智能产生的重要性。更为重要的是,他在知觉(被烫疼痛或者温暖愉悦)、回忆和想象之间做出了一种简易区别。而这一问题也是困扰诸多哲学家如胡塞尔的头疼难题。可以说,明斯基在记忆与智能关系上已经展示了一种来自人工智能领域的态度:记忆与智能的产生密不可分。当然,除了这些重要的贡献之外,他也强化了一种对于记忆的不利观点:记忆是智能产生的低级阶段。但是,这意味着他并没有真正理解记忆本身。

明斯基对记忆的理解有着非常明显的心理主义的特点,即记忆是一种心理联结过程。他对记忆做出了这样的规定:“记忆是一种程序,它让我们的一些智能体按照以前在不同的时间行动的方式再次行动。”^{[8](P184)}他还对元记忆现象做出过阐述,他的“关于记忆的记忆”就是指向元记忆的问题。他指向的是记忆的产生和存储这一难题,并且提出了他的知识线理论(K线理论),这一理论描述了记忆产生和存储的空间

性。“我们把学到的东西放在离首先学会他的智能体最近的地方以便容易地提取和使用知识;每当我们解决了一个问题或者有一个好主意的时候,它就会与被激活的思维智能体相联结。之后当激活K线的时候,与它相联结的智能体就会被唤醒……”,在他的分析中依然显示了“记忆是联想”的心理学观念。

在智能的讨论中,大卫·芒福德(David Mumford)指出,目前忽视了“情绪”这一重要因素。“没有情绪分析的话,计算机科学家在给机器人编程就会出错,无法使之能在与人类互动时正确模仿并回应情绪,我们把这种至关重要的能力叫做人工共情(artificial empathy)。我甚至承认,如果我们希望AI真正拥有意识,我相信它必须在某种意义上拥有自己的情绪。”^[9]他在这一问题上采取了共情推演的方法,如从人有意识推演到动物是否有意识这个问题,“意识并非非黑即白,不是要么有要么没有。它应该以程度来衡量。”“对时间流动的感知才是意识真正的内核。……我们每个人都拥有对讯息万变的当下的连贯体验……这种体验与知觉在本质上截然不同,而且比它更基本,这就是让我们拥有意识的东西。”“机器人拥有某种真正的情绪。”

如果说,明斯基将记忆看作是智能生成的低级阶段。那么另外一位学者安东宁·图因曼,我们会看到他吧记忆看作是智能算法的基本组成部分。“我讨论了认知、识别、记忆、抽象、分析、理解和信息检索,作为智能算法的基本部分。”^[10]纽约大学教授杨立昆(Yann LeCun)认为,人工智能变革的点在于“无监督学习”,关于智能与常识部分的模型则是“感知+预测模型+记忆+推理规划”。在他的模型中,我们很容易看到,记忆只是智能模型的一个有机部分,除此以外,还有感知、预测和推理等多种因素。

反观国内人工智能领域,也存在着相似的

看法。在知识路径的框架中,记忆被看成与信息相关的过程。在知识生成的问题上,公认的被接受的模式是“数据—信息—知识”,而在这个模式里,记忆是处于第二个层次,即与信息相关的层次,表现为信息的编码、存储和提取。此外,在陈霖、张跋等院士看来,记忆被包含在“认知层次”“认知基本单元”中,认知层次由知觉、注意、学习、记忆、情绪和意识等构成。在张跋看来,记忆被包含在决策和行动中,如记忆、遗忘在决策和行动中的作用。他们的这些观点都支持了记忆附属论观点的有效性。

从人工智能发展史我们可以看出,大多数学者还是接受了记忆附属论的观点,即记忆是智能的附属,或者是知识产生的低级阶段,或者是智能算法的组成部分,更或者是认知基本单元的组成部分。但是,在人工智能领域内,对记忆附属论观点的反思也有所体现。我们可以从深度学习领域看到这一点。这一领域为我们提供了新的视角:记忆是智能能力展现的条件。

四、记忆:智能得以可能的条件

作为第二代人工智能的深度学习路径无疑走出了一条新路,它是建立在经验之上的学习方式。“……允许计算机从经验中学习,通过概念的等级来理解,每一个概念都用与之更简单地概念加以界定。……概念等级允许计算机通过更为简单的概念来学习复杂概念。如果我们画一个图表来显示这些概念如何建立起来,这个图就是带有许多层的深度图,因此,我们把它称为‘深度学习’。”^{[6](P2)}在这个界定中,触及到经验与记忆、智能与记忆的关系问题。在这种路径中,科学家不再仅仅局限在记忆是智能的附属部分的结论上,他们往前迈了一步,将记忆理论看作是破解人工智能行为的关键性因素。

在深度学习中,记忆被看作是通达智能的必要条件。这意味着深度学习的提出并不仅仅是人工智能技术的进展,其意义远未被估计出

来。在我们看来,深度学习是人工智能发展的飞跃。在传统的认知范式中,记忆仅仅是整体行为的一部分,甚至是低级阶段,在追求高级能力的过程中,这种低级的能力完全可以被忽视。但是,深度学习改变的是记忆对于人工智能中的地位。在人工智能的决策和行动中,记忆和遗忘远不是低级阶段,而是条件。在哲学看来,这远远不够。记忆与智能的真正关系远没有被揭示出来。

从过程来看,深度学习过程是一个基于在先经验的提取特性的过程。“深度学习通过把预期复杂的测绘分解成巢状的简单的测绘序列来解决这一问题,每一个被模式不同的层次来描述。在可见层上呈现输入信息,之所以这样命名是因为它包含了我们能够观察的变量。然后一系列隐藏层逐渐从图像中提取抽象的特性。”^{[6](P6)}在大多数借助历史数据和图像进行学习的深度学习中,算法是从对象之表征物中提出特性。这个过程有点类似于卡尔纳普的世界之逻辑构造,把更加抽象的表征建立在较少抽象的表征基础上。“深度学习是一种特殊的机器学习种类,通过学习把世界表征为概念的网状系统来获得更大的能力和弹性。这个过程是通过把每一个概念建立在在相对简单的概念之上,把更加抽象的表征建立在较少抽象的表征的基础上来实现。”如果从一个正方形的学习来看,这个过程经历了从“边”到“轮廓”,从“轮廓”再到“对象”的过程。“这儿的图像是被每一个隐藏单元表征的特性类型的可视化。考虑到像素,通过比较邻近像素的明亮度,第一个层次可以来轻易分辨‘边’,考虑到第一个隐藏层的对‘边’的描述,第二个隐藏层能够轻松寻找角和扩展的轮廓,这些被识别为‘边的集合’。考虑到第二个隐藏层用角和轮廓对于图像的描述,通过找到特定角和轮廓的集合,第三个层次能够察觉特定对象的整个部分。最后,用它包含的对象部分的名义,图像的描述能够用来识

别图像中表现的对象。”^{[6](P6)}这里的对象都可以被称之为“可描述性的特性”。深度学习是“阅片无数”,它需要学习无数的照片和数据,从而形成自身的经验。“许多人工智能的任务解决的方式是:设计出合适的满足那些任务而提取的特征集,然后把那些特征提供给简单的机器学习算法。例如,一种从声音中识别说话者身份的有用特征是说话者声道大小的估计值。因此它会给出强的线索来显示说话者是男人、女人还是小孩。”^{[6](P3)}然而,还有一些任务是无法提取特性集的,这被称之为“不可描述性特性”。作者举了一个例子,识别图像中的汽车的例子。“汽车有轮子”,但是这样一个特性很难描述。“一个轮子有简单的集合形状,但是它的图像可能被落在轮子上的阴影轮子、金属部分发出眩目的阳光弄复杂了,或者前景中模糊轮子部分的对象也会产生这样的结果。”^{[6](P3)}这种情况人也会遇到。比如在黑夜无法识别一个人或者一个东西。这一点在现象学中也被作为分析的对象。

这种持续性学习获得的经验成为人工智能机器做出决策和判断的重要根据。持续性学习过程要面对的问题是记忆和遗忘。记忆是确保机器学习经验的积累,克服遗忘的灾难性后果是为了确保后期学习的经验不会影响到先前学习的经验。

但是,机器和人类的差别还是非常明显的。我们设计出这样一个思想实验,让机器阅读一封遮蔽作者的信件,然后判断出作者。这个实验曾经被胡塞尔用来描述人类如何识别出作者的过程。“例如,我们手拿一封旧信,它以不确定的一般方式指示着某人,但我们不知道这个人是谁。信的笔迹看起来是我们所熟悉的,而且这时浮现出对好些人的回忆,我们拿不准是谁。当阅读第一行时,对收信情境的确定的、但却绝非直观的回忆浮现出来,而且这个人随机被确定了,当继续阅读这封信时,这个裁定得到

确认。”^[11]面对这样一个过程,机器如何识别?按照一般的理解,机器会根据字迹的比较来判断出作者。仅此而已。但是对于人类而言,字迹是熟悉的字迹,相伴随的还有对人与事的回忆,更有对叙事情景的回忆。所以在这个过程中,除了认知之外,还有回忆的作用。这样以来,我们从胡塞尔这里看到了一个隐藏的问题:记忆与认知的并列因素远远被忽略了。而我们需要做的是重新呈现出这一对关系,并让这一对关系的思考能够为人工智能的理解提供哲学根据。

五、与智能并列的记忆及其人工智能发展

以上所展现的是人工智能领域对记忆附属论的不同看法,这些看法严格意义上来说并不构成反思和批判。真正的反思和批判还是源自哲学自身的。所以,我们需要回到哲学来看一下记忆与智能的一种原初独特关系的样式。

中世纪时,奥古斯丁与安瑟伦特别指出记忆与智能是灵魂的并列的两大力量之一。“奥古斯丁还遗留给中世纪基督教义一个‘三位一体’观,即灵魂具有三重力量:‘记忆、智能和天赋’(见西塞罗的《论发明》第2章,第53篇,第160节)。在他的《论三位一体》的专述中,这三位便是‘记忆力、智力和天赋性’,三位一体即三位在人身上的映射。”^{[12](P80)}“秉承圣奥古斯丁的思想,圣安瑟伦又提出勒‘智能、意志、记忆’这样的三位一体观。安瑟伦将其堪称是灵魂的三重‘高贵性’。”^{[12](P88)}这些讨论都是对记忆在灵魂功能地位的描述。“认识主体具备四种相对独立的经验直观能力:知觉、记忆、归纳和证言,它们对于现象的经验认识的四个来源。”^[13]笔者曾经撰文指出,在灵魂构成中,记忆被看成是灵魂的三大部分之一,与智能并列^[3],这一观点在灵魂功能观念中又得到了加强。

整个哲学传统所描述出的这种关系让我们

必须去思考人工智能所建立的智能与记忆的关系,对于他们而言,作为智能体的人工智能,其功能实现过程中,记忆很显然也必须是诸多功能之一,失去了这一维度,人工智能是不完整的。但是,在实现过程中,把记忆放置在认知之下的做法很显然存在问题。因为在这种做法中,认知被看成是智能体的主要功能之一。但事实上却完全行不通。如此,在智能追求的道路上,无论是模仿人类智能制造出机器智能,还是机器自身生长出智能,更为重要的是回顾更加本源的概念。来自谷歌的顶级人工智能研究团队 DeepMind 更是致力于要“给人工智能加点记忆”。加入的记忆远不是点缀元素,而是体现了深层次的追求,或许这些科学家已经意识到智能与记忆是灵魂真实存在的两种力量。仅仅获得智能远远不够,还需要获得回忆的力量。这才能够导致完美的灵魂的诞生。

从此出发,或许我们能从记忆哲学而不是技术上勾勒出 AI3.0 的基本特征。AI1.0 基于符号推理的知识驱动,而 AI2.0 是基于深度学习的数据驱动。AI3.0 需要克服的是智能的可解释性问题,是真正智能的问题。所以, AI3.0 可以看作是深度学习的“经验驱动”特征。“经验驱动”中的经验并非传统知识论框架之中的作为知识来源的经验,而是作为过去记忆之保留的经验。之所以如此至少有三个方面的理由:一是来自哲学自身的理由。正如我们前面分析的,缺乏记忆维度的智能并非真正的智能,真正的智能必须与记忆维度共在。二是来自上述特征概括的根据。符号推理与深度学习都是从机制上来说的,只是在人工智能的研究中,物质性的神经机制是最为看重的因素,而多少忽略了智能自身的精神性因素。而我们在芒福德那里看到的恰恰是一种很好的观念,他从意识自身的生成来讨论智能的核心特征。他的分析最终指向了两个方向:情绪和时间性。这种分析颇具哲学意味。如果从此出发,我们所

说明的维度中记忆也是意识得以成为对象的条件。如果是这样,记忆作为智能产生的条件就被确立了下来。而这样一来,“经验驱动”就成为可以理解的特征表述了。而且这一表述比数据和知识还要更加稳固,因为后者更多是对象式的结果,而记忆始终是作为是的对象成为可能条件的规定性上。三是经验驱动相比知识驱动和数据驱动更加具有说服力。人工智能专家强调深度学习是从经验之中进行的学习。如果从此观点出发,经验驱动要比数据驱动更具优先性。数据驱动主要是指深度学习的素材而言,即大数据成为人工智能的驱动力。而数据只是原始的素材,要把数据变成相应的信息还需要提取模式与抽象模式。而这是对深度学习深入挖掘的必然结果。而当触及到这个维度时,经验就凸现出来。如此, AI3.0 的特征表现为经验驱动就变得具有一定的根据了。当然对于人工智能学家来说是否具有说服力还需要相应的检验。

[注释]

- ① 笔者曾经提出记忆哲学是理解人工智能的钥匙(《记忆哲学:解码人工智能及其发展的钥匙》,《探索与争鸣》,2018 年,第 11 期),在这一观点的引导下展开思考,智能与记忆作为记忆哲学的基本问题需要解决和面对。
- ② 从范畴来看,“智能”(intelligence)概念并不属于严格的哲学范畴,我们很难从传统哲学家那里找到比较系统的智能论述。从亚里士多德到胡塞尔,我们看到的是从灵魂到意识的相关论述和分析,却没有智能;即便对于最有可能的黑格尔来说,也只是精神概念成为最根本的概念。这个概念只是随着近代实验心理学的出现,才有了地位。这个概念也无法避免被数量化的命运,所以智能与智力测试密切联系在一起。心理学的做法为智能确立了一个科学的标准,即可以被测量的指数,也就是后来 IQ 合法性的确立。所以,从智能本身来看,至少存在着四个事实需要注意:(1)从质上说,智能是通过力量和能力表现出来;(2)从量上看,智能完全可以被测量,并通过某种方式加以表达;(3)从解释来说,可以通过意识、灵魂、心理等来解释和理解智能概念的相关问题;(4)就智能本身来说,还需要关注到智能的承载者。

- ③ 从智能承载者来看,人类智能与机器智能为两个被对立起来的端点。如果我们接受(3),如用灵魂来解释智能的话,可以从哲学史上找到人类智能和动物智能的合法性源头。亚里士多德在《论灵魂及其他》一文中指出了植物灵魂、动物灵魂和人类灵魂的三分法。他把植物灵魂归于感觉机能、动物灵魂归于运动机能、人类灵魂归于理性机能。如此解读的结论是:人类智能与人类灵魂相等同,而动物智能与动物灵魂相等同。但是,这种独立存在的问题是,忽略了动物智能在整个智能类型中的地位。此外,在人工智能灵魂讨论问题的时候,我们会看到有一些混淆,如人工智能与机器智能的关系。一般情况下,人工智能与机器智能被划上等号。但是,二者存在最大的区别是,人工智能主要是强调的是智能的实现问题,其背后的根据是智能的可测量性和技术可实现性。而机器智能则突出了智能承载者,与机器并列的人类和动物会成为主要的对象。而如果是这样,人工智能理应与自然智能对应,而自然智能包括人类智能、动物智能等形式。
- ④ 图灵测试的经典例子来自以下这篇文章,Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433—460.
- ⑤ 这篇文章主要讨论了人工智能体具备先天的类人空间表征结构,见 Andrea Banino et al, 2018, Vector — based navigation using grid-like representations in artificial agents, *Nature* volume 557, pages 429—433。能够为人、动物和智能体的同构分析提供科学根据。
- ⑥ 根据罗素的分析,像人一样思考来自 Richard Bellman (1978)、John Haugeland (1985) 的观点;像人一样行为来自 Ray Kurzweil (1990)、Richard Karp 和 Kevin Knight (1991) 的观点。

〔参考文献〕

- [1] Paula Boddington, Towards a Code of Ethics for Artificial Intelligence[M]. Springer, 2017: 36.
- [2] Stuart J. Russell. Artificial Intelligence——A Modern Approach[M]. Prentice Hall, 2010: 30.
- [3] 杨庆峰. 记忆、认知与记忆本体论[J]. 南京社会科学, 2018(7): 32—40.
- [4] OECD Multilingual Summaries Artificial Intelligence in Society Summary in English[EB/OL]. <https://www.oecd-ilibrary.org/docserver/9f3159b8-en.pdf?expires=1562289689&id=id&accname=guest&checksum=7EB8CC4E5DCB12BD8B654ED5B5BDF175> [2019/7/5].
- [5] Daeyun Shin, Charless C. Fowlkes, Derek Hoiem, Pixels, Voxels, and Views. A Study of Shape Representations for Single View 3D Object Shape Prediction[EB/OL]. <https://arxiv.org/pdf/1804.06032.pdf> (2018—06—12) [2019—07—05].
- [6] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning[M]. The MIT Press, 2016.
- [7] [德]黑格尔. 精神现象学[M]. 贺麟, 译. 北京: 商务印书馆, 1979: 128.
- [8] [美]马文·明斯基. 心灵社会: 从细胞到人工智能, 人类思维的优雅解读[M]. 任楠, 译. 北京: 机械工业出版社, 2018.
- [9] David Mumford. Can an Artificial Intelligence Machine be Conscious[EB/OL]. <http://www.dam.brown.edu/people/mumford/blog/2019/conscious.html>, 2019—04—11.
- [10] [荷]安东宁·图因曼. 智能就是算法吗? [M]. 答凯艳, 等, 译. 北京: 机械工业出版社, 2019: 23.
- [11] [奥匈]胡塞尔. 被动综合分析[M]. 李云飞, 译. 北京: 商务印书馆, 2017: 124.
- [12] [法]勒高夫. 历史与记忆[M]. 方仁杰, 倪复生, 译. 北京: 中国人民大学出版社, 2010.
- [13] 周昌忠. 科学的哲学基础[M]. 北京: 科学出版社, 2013: 98.