从哲学角度探讨人工智能的理解问题和对人生的影响

文/张庆熊

当代分析哲学对人工智能的发展起了积极的推动作用。正因为如此，在目前有关人工智能方法论的教科书中，分析哲学通常占很大比重。然而，我觉得如果要对人工智能的理解问题进行深入反思的话，光了解分析哲学的观点还是不够的。人的理解涉及自我意识、理性反思和思想观念的辩证发展。在这方面，笛卡尔、康德和黑格尔等经典哲学家的观点值得我们重视。他们有关意识、理性和观念的学说直至今天还深刻影响我们对自我意识、认知能力和意义理解的看法。从他们的哲学立场出发，会对人工智能提出怎样的问题呢？他们的论述虽然不像分析哲学家那样在技术层面上与人工智能直接相关，但是会在更广和更高的层次对人工智能提出发人深省的探问。哲学是多元的，不同哲学流派的观点可以为我们提供多视角的切入点，便于我们反思综合。与此同时，我们也需考虑人工智能的应用和在生活世界方面的回应。从哲学角度反思科技与人生的相辅相成和挑战应对的关系，有利于开阔我们的思路。

从笛卡尔的唯理论看人工智能

笛卡尔是近代理性主义（唯理论）哲学的创始人。他在《方法谈》《形而上学的沉思》等著作中建立了欧洲近代有关理性思维方式的主体主义的哲学立场和方法论路线。对照笛卡尔的观点，有如下两个问题值得讨论：

（1）笛卡尔主张理性认知要有一个自明的开端。从笛卡尔的角度看，是否意味着人工智能由于不能确定自我意识的“自明性”（self-evidence），因而难以解决理性认知的开端问题呢？

（2）笛卡尔主张，知识是一个公理系统，以自明的公理为基础，然后借助规则进行推演。人工智能是否能够自己建立和证明这样的公理？人工智能能否自己建立具有概念和公理的框架性革新的理论体系呢？

笛卡尔持第一人称的主体主义的哲学立场。在笛卡尔看来，“我思故我在”的自明性根植于主体的内在体验，而非外部逻辑推导。时至今日，很多人还持类似于笛卡尔的立场，从第一人称的自我意识的自明性出发否定人工智能具有真正像人一样的自我意识。从这一角度看，人工智能的“自我意识”面临根本性挑战：若意识的自明性依赖于主体对自身存在的直接体验，那么人工智能即便能模拟自我指涉的符号系统，其本质仍是人类预设程序的产物。

笛卡尔主张知识应以自明的公理为基础，这些公理是通过主体的理智直观获得的。对照笛卡尔的理智直观的要求，人工智能在进行推理的时候，通常是从人类给定并输入计算机程序的公理或定理出发的，人工智能本身并不能自己建立和确证这样的公理。发展至今人工智能的“创造力”大都还停留在组合优化的阶段，缺乏对公理自明性的认识论奠基和知识体系建构的原创性能力。

从康德的先验论看人工智能

在笛卡尔之后，康德是又一位重量级的近代哲学家。康德反思了笛卡尔等的唯理论和休谟等的经验论各自存在的问题，力图把这两者结合起来，提出自己的先验论的认识论主张。对照康德的先验论，有如下三个问题值得讨论：

（1）康德主张，人的认知基于先天的认知形式和范畴，即感性认知基于时间和空间的直观形式，知性认知基于因果关系等认知范畴。而人工智能的推理基于规则和算法，这是否与康德所说的人的认知基于认知的形式和范畴的思路有相似之处呢？

（2）按照康德的看法，这些先天的感性形式和知性范畴既为人的认知提供了可能性，又为人的认知划定了界限。它们只适用于现象界，并制约了人的经验认知的范围。康德指出，人的认知一旦超越了经验的领域而尝试回答世界的第一原因之类的本体界的问题时就会陷入二律背反。人工智能的算法是否也是这样：既为人工智能提供了可能性，又为人工智能划定了界限呢？

（3）按照康德的看法，人的认知分为感性、知性和理性三个层次。停留在感性和知性层次上的认知活动还属于较为低阶的认知活动，还没有

上升到理性认知的阶段，只有能够对认知的可能性范围及其限度进行反思的认知才达到理性的水平。人工智能能够反思人工智能的可能性范围及其限度吗？人工智能能够发现其中的二律背反吗？人工智能如果不能反思人工智能的可能性范围及其限度，是否像康德所说的那样还没有真正达到理性的层次？为了达到这种理性的反思，人工智能应该如何改进呢？

康德对认知的可能性进行反思的论述，值得我们重视。人制造了工具，但任何一种工具都有其使用的范围和限度。人的理性表现为知道工具的效用及其限度，从而能够用其所长避其所短。同样，人使用知性范畴的推理也是有其效用范围的。迄今的人工智能推理的大模型依然是算法嵌套的产物。即便人工智能的元学习（Meta-Learning）允许优化其自身的某些参数，但这种优化仍受初始算法目标的严格约束。为了防范人工智能万能的“幻觉”，康德的如下警句值得我们铭记：“先验幻相甚至不顾批判的一切警告，把我们引向完全超出范畴的经验性运用之外，并用对纯粹知性的某种扩展的错觉来搪塞我们。”

从黑格尔辩证法的角度看人工智能

黑格尔是自笛卡尔和康德以来欧洲哲学的集大成者。他建立了观念论的辩证法体系，主张绝对观念在一种对立统一的辩证关系中演化，并由此产生出各种形态的观念和事物。人类的认知和理解能力是在精神与物质、主体与客体、普遍与特殊、同一与差异等对立统一的关系中发展起来的。从黑格尔辩证法的角度，我们该如何看待人工智能呢？我想有如下三个方面的问题值得讨论：

（1）黑格尔主张观念自身运动，互相联系和相互转化。因而，观念是活的东西，自己能够生成新的观念。这对人工智能是否能产生新思想的问题有无启发意义呢？黑格尔有关观念自主运动的观点对解释人工智能的“涌现”现象有何助益呢？究竟应该怎样理解和解释人工智能的“涌现”现象呢？

（2）黑格尔主张知识是一个辩证的系统，人的认知是在主体与客体、普遍与特殊、同一与差异的对立统一的关系中发展起来的。这对人工智能的大模型的建构有无启发意义呢？对人机互动的认知架构有无参考意义？

（3）黑格尔主张世界历史的发展蕴含一种历史目的论，人的价值观念的发展是与人类社会的发展相辅相成的，观念运动的辩证逻辑与社会发展的历史逻辑相统一。这对人工智能阐明价值观念有无启发意义呢？人工智能是否能从社会历史发展的线索中梳理出道德和伦常发展的线索，从而发挥伦理教化的作用呢？

参考黑格尔辩证的思想方法论，可以从三个方面反思人工智能的理解问题：

其一，黑格尔有关观念自身运动的观点对人工智能的研究有启发意义。在黑格尔看来，语言是活的东西，文本是逻各斯的展开形态。心灵并非凭空思考，而是借助语言进行思考。语言是思想不可或缺的中介。让我们设想，当一个巨型的电脑储存了千百万册图书，它会不会活起来呢？依然不会，藏书再多，如美国国会图书馆2300万册藏书，依然是一个静态的知识容器。然而，当ChatGPT-3通过45TB语料的递增积累，在1750亿参数的动态关联中就产生了黑格尔所说的从量变到质变的飞跃，涌现出跨领域推理、隐喻理解等非预设能力。这种现象该如何解释呢？让我们设想，图书馆里有一群非常勤奋和智慧的图书卡片整理工作者，他们对图书进行分类和编目，并按照关键词进行梳理，标注其中的相似性和差异性，并且勾画出其中关联性的导图。当读者拿到这样的卡片和导图后，就会感到很有用处。他们会按照这样的卡片和导图寻找图书和阅读图书，并且还会作出评价和写上自己的改进建议。如此循环往复，这样的卡片和导图就会越来越丰富和完善。人工智能所做的工作类似于这群图书卡片整理工作者和读者的共同努力。在此，人工智能所说的“神经网络”在一定意义上可被理解为图书分类、编目和索引的机制。它在形态上好似一层层串联起来的卡片箱，卡片好似神经网络上的神经元，它们之间的串联和信号传输好似神经系统，它们与图书馆中储存的图书相关联，并处于动态调整的过程中。人工智能所进行的“文本整理”不能仅仅被理解为输入语料，因为文本不只是一堆语料，而且包括语料间的脉络。文本中存在逻各斯，文本是逻各斯的展开。只有把握了文本中的逻各斯，才谈得上“理解”。这一观点是黑格尔着力强调的，在此意义上他认为文本是逻各斯的主动展开，“理解”就是把握文本展开的活的脉络。人工智能在输入和整理文本的过程中学习了文本中的逻各斯，从而培育了人工智能的理解能力。

其二，黑格尔有关认识是在主体与客体、心灵与肉体、观念与现实等对立统一的关系中展开的观点对人工智能的研究有启发意义。在现有的

ChatGPT和DeepSeek等大模型上有“对话”一栏。我们知道，“对话”是“辩证法”的起源。“对话”不仅是问答，而且包括论辩。对话者之间为一个论题展开辩护和反驳，并可能以某种方式达成综合性的意见。用黑格尔的话来说，“对话”或“辩证法”由“正题”“反题”“合题”三个环节组成。对话的这三个环节对于论文的写作非常重要。缺乏这三个环节的论文往往显得缺乏内在的逻辑关系，特别是缺乏论辩的逻辑关系。目前人工智能大模型的“对话”在知识问答方面做得比较好，但在论辩的逻辑方面不尽如人意。从这一角度考虑，如何把黑格尔所主张的对立统一的辩证法吸纳到人工智能的思考模式中去，是一个有益的改进方向。

其三，黑格尔有关“逻辑与历史相统一”的观点对人工智能的研究有启发意义。黑格尔主张，历史并非偶然事件的堆积，历史的发展具有合理性，这种合理性表现为符合逻辑性。在此意义上，黑格尔留下名言：“凡是现实的都是合乎理性的，凡是合乎理性的都是现实的。”从黑格尔历史辩证法的角度看，目前人工智能在解读历史文本的脉络或逻辑方面做得很差，还没有找到一条在历史文本中识别历史发展线索的途径。大语言模型在回答问题时，根据对话框给出的主题和上下文语境关联等因素，从概率统计的角度找出最适合的文本组合作为答案。其回答问题的成功，源于其对文本切片的短距离和长距离语言依赖关系的捕捉能力。这里很容易疏忽历史真实的语境与小说虚构的语境之间的差别，也难以区分什么是真实世界中的历史发展的逻辑关联与小说虚构中的想象的情节关联。对历史文本及其可靠程度做出标记，找出历史事件与社会条件和发展规律的联系，对于人工智能的大模型而言仍然有许多可做或可思考的议题。

人工智能的效用和风险及对人生的影响

我们已经从哲学理论的角度反思人工智能的理解问题，现在想从哲学和生活相结合的角度讨论人工智能的效用和风险。

从当代哲学的发展历程看，笛卡尔在意识内部寻求明见性的路线遭到强烈批判，即便沿着笛卡尔路线走的胡塞尔的现象学，在后期也融入生活世界的学说，主张理解是在意向活动与周围世界的交互过程中形成的，明见性发端于日常生活中的经验。从这个角度看，人对知识的学习和理解有如下三个阶段：（1）所学知识进入自己的明见性视域，将抽象的理论与日常生活中的明见性经验相结合；（2）用所学知识来解决实际问题，在生活实践中应用该理论知识；（3）与人交流相关的知识，并能质疑乃至更新相关的知识。

如果我们遵循这一学习和理解的步骤，那么现在问题的关键就在于，是否能巧用人工智能进行学习和提高理解。就“自我意识”这一要求而言，当提问者向AI大模型提出问题的时候，他必须意识到自己提问的意图并需要判断AI大模型的回答是否偏离了问题的主旨。如果AI跑题了，提问者责怪AI是没有用处的，他需要做的是调整自己的提问方式。如果AI的回答很抽象，他可以提示AI用更贴近生活的例子让知识进入明见性视域。理解的最高境界在于反思。反思是对人工智能提出的解答进行质疑，在不断的反驳和追问中划定其界限，进行扩充或做出新的解释。

人工智能在给我们带来巨大效益的时候，也伴随着巨大风险。人们常常嘲笑黑格尔的辩证法是“变戏法”，但是在复杂处境中秉持辩证眼光看问题，是头脑清醒地正视现实的态度。人的认识处于历史的发展过程中，同样人工智能也在继续发展，我们所讨论的是现阶段的人工智能的功用及其局限性。人工智能的产生和应用有其历史的合理性，但如果不正视和正确处理人工智能所产生的问题，人工智能也会失去其存在的合理性。人工智能是把双刃剑，它所产生的正面和负面效应同样令人惊讶。

现阶段的人工智能从总体上说依然是一种工具理性。主流AI大模型的开发遵循“理性决策”范式，即将人的决策行为简化为实现效用最大化的决策工具。现阶段的人工智能带来的另一个弊端是人的异化。人是理性的动物，这句话原本意味着人有自主性，人能通过自己的理性思考自主地选择自己的行为。但由于人工智能的发展，人容易养成一种习惯，遇事必问人工智能，让AI大模型告诉我们该怎么办。就像古代人遇事必求甲骨算卦一样，现代人也会把人工智能当作一种“神明”对待。这样，人与计算机的关系就被颠倒过来。本来，计算机只是一种理性思考和理性决策的工具，但过度依赖智能系统可能导致认知能力和自主决策能力的退化。最后，每事必求智能系统，计算机就从奴隶变成了主人，人反成了受计算机支配的工具。这就是黑格尔所说的“主奴关系”的变化。

〔作者系复旦大学哲学学院教授；摘自《哲学分析》2025年第4期〕