羊城晚报
A10
理论
文史哲
2026年3月6日
星期五
责编 潘玮倩
美编 夏学群
校对 林 霄

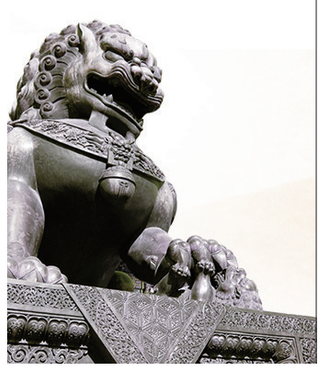
编者按

在技术狂飙的当下，生成式人工智能正以前所未有的烈度，重塑着人类的认知边界。面对这股时代热潮，羊城晚报文史哲周刊立足思想前沿，特别推出“AI狂飙时代的人文与哲学观察”专题，刊发两篇重磅文章，试图在代码与算法的呼啸声中，重新锚定“人”的坐标。

本期两篇文章构成了一场跨界对话。刘志伟先生从历史的视阈破局，呼吁人文学者退出算力竞赛，回归“生活之场”，向内挖掘人类的主体性底座。周兵先生则从哲学本体论出发，冷静剖析AI进化出“意识”的可能性与伦理边界，以严密的推演向外勘探机器的智能天花板。

前者反思“人类该如何自处”，后者追问“机器将走向何方”。在这场演化的“狂飙”中，两位学者殊途同归地揭示了一个深刻的真相：对抗技术吞噬的终极解药，可能恰恰在于保全人类自身那不可计算的复杂、脆弱与丰饶。

专题策划：温建敏
专题执行：潘玮倩



AI时代人文学科的使命

在“非逻辑”与“失忆”中重塑人类主体性

□刘志伟

视觉中国供图

在探讨人工智能（AI）与人类未来这个略显沉重的话题时，我想先从一段大家熟知的旧视频谈起。那是几年前马云与埃隆·马斯克的一场著名对谈。当时，马云认为计算机只是玩具，人类创造了机器，机器绝不可能比人类更聪明；而马斯克则直言，人类在算力与智能维度上终将被全面超越。

坦率地说，在“谁的能力更大、谁更聪明”这个问题上，我倾向于认同马斯克。我年轻时曾是一个“极其单纯”的科学崇拜者，深信人类最终会被科学所奴役；后来学习了马克思主义理论，深刻理解了工业化进程是如何将人“异化”为机器的奴隶；如今到了数字化时代，随着人工智能的发展，这种人被技术奴役的危机感似乎越发逼近。

但是，作为一位在人文学科领域“混”了几十年的人，我逐渐意识到，我们思考问题的坐标系，根本不应该停留在马云和马斯克那个层面。解决技术跃升、改变人类生存技能的问题，是科学家们的伟大使命。人文学者完全不需要，也绝不应该在这些层面上去与AI竞争。

“不及格”的AI提纲与错位的竞争赛道

以历史学为例。在我读研究生的时代，为了查阅几份史料，我曾经揣着300块钱经费跑一趟北京图书馆（中国国家图书馆前身）手工抄写十本书，中间还可能被“赶”出来，需要重新开完证明再去。而今天，哪怕不用能迅即找出史料出处的AI，只需要通过数据库查询，我也可以在5分钟内查到所需的文献，甚至直接下载。如果借助AI，它还能在极短时间内处理数以千万计的文献材料并写出一篇文章。所以，如果人文学科的价值仅仅被定义为“占有和处理材料”，我们在这种算力面前已经一败涂地。

就在这场论坛（“人文学术的未来”）的上半场，我还坐在台下用手机玩了一下简单的AI工具（Kimi），让它帮我生成了一份关于“AI时代人文学科使命”的发言提纲。如果这是一场标准化考试，我相信那份AI生成的提纲起码能拿90分。但在真正的人文学者看来，那篇东西全是胡说八道，是不及格的。为什么？因为那份AI生成的提纲，其底层的逻辑依然是在比拼知识量、逻辑推演和解决问题的能力。如果人文学科顺着这种框架去定义自己的使命，去跟机器比拼谁更“聪明”，那我们是一定会输的。用纯粹的算法逻辑来规训人文学科，这恰恰背离了人文学科的本质。

我们的使命，不在于和AI比拼大脑的运算能力，更不在于去驾驭那些超越人类智慧边界的数据。人文学科的终极指向，是回归并回味我们作为“自然人”的本质。

“碳基人”的底气：非逻辑与不确定性

我的同事余志教授告诉我，科学界将未来的智能主体划分为三种：碳基人、半碳半硅人、硅基人。面对强大的“硅基人”（AI），我们纯粹的“碳基人”长处究竟在哪里？

我认为最重要的一点在于：我们碳基人，在本质上是不讲逻辑的。

在座的哲学家和逻辑学者可能要反驳我。但请注意，逻辑学只是让人类“学会”了讲逻辑，而人类真实的生活运转，往往是非逻辑的。不信你在现实生活中，回到家里跟太太吵架、跟母亲吵架，你试图讲逻辑？讲逻辑你一定输。

再举个例子，我刚学会打扑克的时候，第一次上桌我就赢了。为什么？因为我不懂规矩，不讲逻辑，不按牌理出牌，所以赢了。但等我完全学会了规矩，懂得怎么讲牌理之后，我就总是输。这其中蕴含着一个深刻的差异：人类最本质的特征是不讲逻辑、充满不确定性的。一旦进入非逻辑的、不确定的、充满感性与人性的复杂地带，硅基人的确定性模型便会显露其边界。

人文学科所要处理的，正是这些关乎情感、趣味、人性的非逻辑世界。用非逻辑打败硅基人的确定性，是我们的一大底气。

警惕理性的局限：回到生机勃勃的“生活之场”

面对浩如烟海的文献与不断迭代的算法，历史学家应当如何自处？刚才发言的杨洋老师讲到，日本学者上原专禄曾提出一个极其深刻的论断：对历史的认识有两种思维方法，一种是“历史的思维方法”，一种是“非历史的思维方法”。

通常，人们会理所当然地认为，历史学家的看家本领就是“历史的思维方法”——当别人问“这个史料可靠吗”，“是真实的吗”时，我们能给出严谨的考证。但这远远不够。历史学家真正高明的能力，其实在于“非历史的思维方法”。

上原专禄指出，对于非历史思维方法的终极克服，绝不能在纯粹的“认识之场”或“思想之场”中完成，而只能在“生活之场”——即作为生活总体的历史现场中完成。这里所谓的“认识之场”，实际上指的就是人类高度理性的思维。

在此，我们必须坦诚地面对人文学科内部的一种张力。作为学者，我们试图用理性的、逻辑的，甚至概念化的方式来进行学术写作，去建构历史叙述，但我们所面对的，恰恰是一个充满无序、非逻辑的“生活之场”。我认为历史学容易陷入一种幻觉，就是以为凭借自己的描述和评论就能完整地呈现过去的一切。

其实，我们所有的呈现都带有极大的局限性——时代的局限、个人的局限。一个合格的人文学者，在建构历史记忆与学术叙述时，必须时刻保持一种“解构”自身的能力。我常常整天跟我的学生说：“你讲的都很对，不过你自己不要太相信啊！”我们要创造知识，但绝不能盲目迷信自己构建的逻辑框架。这就是为什么我们要不断从抽象的“思维之场”退后，重新扎入那个混杂着各种变数的“生活之场”。

生活之场本身就是一座“思维之屋”

当然，强调回到“生活之场”，绝非消解理论和思想的价值。恰恰相反，生活之场并非是一个现成供我们进去消解矛盾的空洞实体，它本身就是一座“思维之屋”。

人类在“生活之场”中的所有行为，背后都承载着过去的思想。我虽然是做社会经济史的，但最近几年，我其实在向思想之场、精神之场回归。比如我在商学院开经济思想史的课，我没有从物质生活讲起，而是从《周易》《孟子》《管子》讲起。因为生活之场中的人，本身就被思想支配着。

我们需要重新回到思想领域和知识领域。以清代为例，清代对我们今天的意义，也在于它累积了中国过去两三千年的思想成绩。这种精神世界的东西放回到生活之场中，往往是更具支配性的。不理解这些思想的积淀，我们就无法真正理解现实的“生活之场”。

未知的创造与处理“失忆”：AI无法抵达的边界

今天我们对人工智能最大的恐惧，是害怕人类会失去主体地位。但只要人文学者依然保持对人性和职业使命的自觉，我们就永远是自身命运的主人。

因为人类永远在创造自己的生活，而这种“再创造”绝不是基于已有知识的线性推演，而是向着一个完全无知、未知的空间在拓展。AI只能在既有数据与既定规则中推演，无法主动生成真正意义上的历史性未知。

关于这一点，代际差异是个极好的切入点。每一个时代，人类行为背后所遵循的道理都在发生巨变。在我们这代人受教育的语境里，很多都是我们那个时代的秩序，是我们的天命和规律。但对于改革开放后成长起来的一代人而言，有些逻辑可能完全不可理解。而现在的年轻人，他们的生活方式和认知，甚至可能又是一种我们这代人觉得不可思议的全新范式。这种根植于社会剧变中的“未知发生”，是仅仅依赖“已知信息”进行运算的AI永远无法处理的盲区。

最后，我想谈谈历史学面临的核心命题：记忆。

我认为，数字化与人工智能的底层运行逻辑，往往建立在对信息全面存储与调用的假设之上，假定所有信息都应转化为“记忆”。但历史学仅仅是处理“记忆”吗？

我认为最高明的历史学家，其实是在处理“失忆”。人类在漫长的长河中，究竟记忆了什么？又“失忆”了什么？这种记忆与遗忘的取舍，一直随着时间、空间的变化而改变。

这就如我们历史学的一句老话——“一切历史都是当代史”。历史从来不是对过去的复写，而是当代问题意识对过去的重组。记忆与遗忘，并非自然流淌，而是主体在现实关怀中的选择。

只要我们紧紧扎根于非逻辑的、充满变数的“生活之场”，只要我们依然承担着为人类不断变化的现实生活重塑意义、处理“失忆”的责任，人文学科就永远不会被人工智能所取消。硅基人或许可以奴役怠惰的碳基人，但最终，技术依旧需要被拥有人文自觉的碳基人所控制。掌握命运的出路，就在我们对人文学科使命的坚守之中。

（作者系中山大学历史学系教授，本文为作者在2024年中山大学人文学部“人文学术的未来”高端学术论坛所作发言的整理延伸。整理者：潘玮倩、薛盈盈、何擎宇）

人工智能及其哲学前沿问题

□周兵

当下人工智能迅猛发展，其成果正在深刻改变当今社会的方方面面，冲击传统价值观，创造新的观念。高度发达的人工智能代表着新近产业革命的时代脉搏，是人类迈入新纪元的标志。人工智能正在改变人们的生活，并引发人们对如此改变的反思。特别是2025年初国产版生成式人工智能DeepSeek发布之后，几乎引发全社会对人工智能海啸般的关注。对人工智能讨论的热情一方面标志着社会整体对时代变迁的感知，另一方面将人们从对人工智能技术问题的求解推进到与之相关的哲学问题思考。人工智能的本质问题、意义问题、价值与伦理问题等显然无法在科学技术框架内得到回答。人工智能是现代科学技术的成果，也是哲学现象。本文概要介绍人工智能及其相关的哲学问题。

人工智能定义及其发展

“人工智能”（Artificial Intelligence）这一概念最早在1956年的达特茅斯会议上被正式提出，用在研究机器模仿人类智能或认知的科学领域，人类智能被模仿为基于规则的符号表征计算型式，如果这些计算模型能在机器上成功运行，则该机器具有人工智能，这是经典人工智能（classical AI）的设想，标志着人工智能作为一门科学学科的诞生。

事实上，人工智能哲学先于人工智能学科诞生。在达特茅斯会议之前，英国数学家图灵已开始思考“机器是否能够思考”。他提出通用图灵机模型，为现代计算机奠基，并通过对“可计算性”“智能”等概念的分析，为人工智能及其哲学奠定理论基础。

而从历史上看，人工智能的发展大致经历了几个阶段：20世纪中期诞生并在60年代迎来第一次高潮，随后经历两次“寒冬”。进入21世纪后，随着互联网数据和算力提升，机器学习特别是深度学习取得突破。近年来，以ChatGPT和DeepSeek为代表的生成式人工智能迅速普及。

人工智能哲学

如果说人工智能是对人类智能的模仿，那么，人工智能哲学的范围也如人类哲学的范围一样广。人工智能哲学涉及传统的从形而上学、认识论到价值论的所有哲学研究领域，不仅要解释人工智能是什么、如何是，还要讨论其应该如何。

人类对人工智能的哲学思考并非始于今日。早在古典哲学中，柏拉图、亚里士多德等人已讨论“模仿”“理性”等问题；近代以来，笛卡尔的身心二元论以及拉美特利“人是机器”的观点等，也为后来关于智能与机器的讨论提供了思想背景。

当前研究者设计人工智能的方法有两种：一是“符号主义”自上而下的方法（心灵层面）；二是“联结主义”自下而上的方法（身体层面）。

第一种方法自上而下地从传统逻辑推理出发，试图用纯粹的符号来表征和处理所有知识，这也是经典人工智能的主要方法。这源于人类试图将“理性彻底公式化”的漫长历史：从亚里士多德的形式逻辑，经欧氏几何的建立，再到代数学与几何学的结合，到笛卡尔提出以数学为中心的“普遍知识”，再到莱布尼茨构想“普遍符号语言”与“理性演算”，人们一直希望用符号系统和逻辑演算把所有推理过程彻底形式化。现代数学家希尔伯特进一步提出“判定问题”：所有的数学推理能否被形式化？或者，是否存在一种机械过程，能判定一阶逻辑中的任意命题是否可证？对此，哥德尔用不完备性定理给出了无法证明的否定回答。随后，图灵与他的老师丘奇分别独立证明此题无解——图灵的证明过程即“通用图灵机”理论模型的提出。

第二种方法联结主义（类神经元计算）的诞生基于脑神经科学的发展，自下而上模仿人脑神经网络的工作模式（学习与记忆的原理），将大脑比作计算机、神经元比作处理器，把人脑神经网络算法程序化，通过模仿人类智能生物物质层次的组织方式来设计人工神经网络、发展人工智能，这也是当下人工智能的主要方法。这是将人类智能物质化、过程化、行动化、具体化与操作化。随着研究发展，学界逐渐尝试将符号主义与联结主义结合，以弥补单一方法的局限。

人工智能以模仿人类认知能力为目标，因此与认知科学的发展密切相关。现今，关于智能的讨论往往转化为智能与意识、身体及环境之间关系的问题，因此人工智能哲学与心灵哲学、意识哲学等领域密切相关。一方面，对智能与相关概念的混淆将直接导致对一些问题不加反思，如人工智能是否有意识、有情感、是否可以自我进化——造成混淆的部分根源，在于“意识”是在近代才开始作为重要哲学概念，任何对这些问题的回答首先需要厘清概念；另一方面，对智能与身体紧密关系的认同使人们开始关注具身理性，直接引发具身智能机器人的研究热潮。

阐释至此，符号主义和联结主义都承认：机器可以模仿人类智能。但问题在于，这种模仿能否精确且全面？是否存在第三条路径？

当下产业界追逐的通用人工智能（AGI），正是对前一问题的肯定回答。如果智能仅仅是计算，那么现在AI早已远超人类。但事实上，智能不仅包含计算，更包括创新思维、因果推理与深层理解；更重要的是，人类是时间性的、身体性的、在世界中的存在，这决定了人类智能不是抽象的，它由很多复杂因素共同决定，如身体、性情、环境等。因此，虽然当前人工智能成功模仿了人类智能的某些方面，但它还没有达到人类智能水平。要实现通用人工智能（AGI）乃至超越人类智能的超级人工智能（ASI），人类需要掌握全方位模仿生命智能的科学和技术——而这能否实现，仍是未知数。

关于第二个问题，首先需要明确的是，当下设计人工智能的方法是符号和联结两种方法的混合，理论上不排除有更优范式，但它必须能克服现有框架的根本缺陷。

符号主义的局限在于形式化边界：并非所有现实情境都能被符号系统完全表达。塞尔著名的“中文房间”实验与德雷弗斯的批判，都指向这一困境。联结主义则受困于生物复杂性：真实神经网络极其精密，难以完全模拟；且在实际应用中暴露出算法黑箱、算法缺陷与数据依赖等固有缺陷。两种方法结合后，其界限依然不明确，这为人工智能未来的发展留下不确定性。脑机接口技术（BCI）等的发展，也使人类与人工智能的边界逐渐模糊。

AI正掀起新的产业革命，新旧行业迭代；算法重塑着信息传播、社会财富与社会观念。然而，隐私泄露、技术垄断、知识幻觉（AI生成虚假却看似合理的内容）及算法黑箱的不可控性，引发普遍担忧。热点问题还包括算法偏见、军事伦理、价值对齐（如何令AI符合人类价值观）及责任归属。而人类最大的恐惧，则是人工智能对人类的超越。

尽管如此，人类研发AI的初衷仍是让它服务人类、克服自身局限、实现自身自由。人类行为的目的与后果之间有两种情况：1. 好的目的带来坏的后果；2. 好的目的带来好的后果。在行为中遵循第二种情况是标准态。因此，人类制造超越自己的人工智能从目的上来看可能没有必要，因为它不一定会带来人类的自由。由此，从人类中心主义的视角来看，超级人工智能（ASI）是一个伪概念；相反，从非人类中心的角度看，这或许是“新物种”借助人类“创生”的机会。或者如海德格尔所言，人类的命运为技术的集置（Ge-stell）所促逼而走向危险。无论如何，AI发展必须以规避恶果为伦理底线，走控制性发展之路。

人工智能的威胁及其防治

人工智能的最大威胁基于一个假设：它将于某一时刻进化出自我意识。一旦AI拥有意识，将可能不再服务人类，甚至反噬其主。因此，“AI是否会有意识”成为全社会讨论的热点。

当然，我们需要厘清“意识”的边界。根据概念分析智能并不等于意识，人工智能也未必天然包含意识。然而，AI的本质是对人类智能的模仿，而人类恰恰是一种有意识的智能体。因此，只要AI试图全方位复刻人类，它终究绕不开“意识”这道终极关卡。

而一旦AGI（通用人工智能）或ASI（超级人工智能）获得意识，人类“技术服务于人”的初衷便可能落空。如此说来，一个最简单的防范也许就是不去发展它，然而现实是，人类似乎已在不知不觉中被技术裹挟。

人工智能未来是否会拥有意识？答案是开放的。合理的回答似乎是保留其获得意识的可能性。在技术层面，联结主义与符号主义融合的界限尚未触及，且AI已展现出自主学习甚至欺骗能力，因此，无法排除其获得意识的可能。在哲学层面，人工智能拥有意识的可能性没有被排除，胡塞尔现象学指出，主体通过“同感”构建有意识的他者；因此，若人在交互中将AI同感为有意识的存在，即便是在受“欺骗”的情况下，该AI也应被视为具有意识。

从哲学视角看，一些学者认为，机器意识在理论上并非不可能；也有学者从现象学或身体哲学角度指出，人类尚未完全理解自身意识，因此复制意识并不现实。

人工智能本质上是一种以数学与算法为机制的智能模仿，是对人类智能客观化的成果。人们担心的威胁——人工智能可能获得意识——在近代以来主客二分的哲学框架下，意味着客观可能反向回到主观，黑格尔的辩证法似乎早已预示了这一可能性。

面对狂飙突进的人工智能，辛顿与本吉奥等先驱纷纷敲响警钟。辛顿曾呼吁建立一个由多个国家参与的国际人工智能安全研究网络，专注于研究如何训练人工智能向善；本吉奥则预言AI的规划能力将在5年内达到人类水平，倡议构建一种只有智能、没有自我与目标的“科学家AI”。综合两位学者观点，如何控制性发展和引导人工智能向善，将决定它成为文明的加速器还是威胁。这为人工智能伦理指明了努力方向。

AI是一面反映人类欲望与恐惧的镜子吗？其走向，也许并不取决于算法，而取决于人类能否在镜中认清自己，并继续选择向善。

（作者系深圳技术大学科学技术哲学研究中心负责人/德国工业文明研究中心研究员）