人工智能发展如何抉择：“祛魅”“返魅”还是“拟魅”

魏屹东

摘　要　人工智能（AI）的发展被大致分为了弱 AI（专用 AI）和强 AI（具身或通用 AI），似乎缺乏中间形态或阶段。从“魅”的视角审视，AI 可分为“祛魅”“返魅”和“拟魅”三种形态或阶段。“祛魅的 AI”是指“无心”的机器智能，纯粹展现工具理性；“返魅的 AI”是指“有心”的机器意识系统，旨在实现具身心智的通用智能；“拟魅的 AI”是处于“无心”和“有心”之间某一“中间状态”的“拟心”状态，目的是彰显交互意识，完成各种认知任务。这种划分有利于新一代 AI 的健康发展，不仅可避免两种智能因进化产生的“解释鸿沟”，也可避免技术壁垒和伦理困境。

关键词　拟魅的 AI　半创造性　无机指号学　神经形态学习系统　适应性表征

作者魏屹东，山西大学哲学学院教授（山西太原 030006）。

中图分类号 B0　　**文献标识码** A　　**文章编号** 0439-8041(2025)07-0031-10

人工智能（AI）的现世，打破了人类独有的思维方式，引发了一系列问题，诸如 AI 是否有智能，是否会思维，能否有意识，能否实现类人的通用化，能否超越人类智能，如何应对其导致的伦理问题，等等。2024 年 12 月 26 日最新一代大语言模型 DeepSeek-V3 的正式发布，标志着其在通用人工智能（AGI）领域的一个里程碑，完成了从通用与代码能力融合到全面性能突破的跨越。2025 年 1 月 20 日 DeepSeek 又发布了最新的高性能 AI 推理模型——DeepSeek R1，进一步引发了 AI 有无意识的争论。[①] 纵观 AI 的发展，从传统 AI 的计算范式（包括符号的和统计的）到新一代 AI 的机器学习范式（强化学习、深度学习以及强化深度学习），体现出“机器智能”从低级感知到高级认知、从弱智能到强智能的发展趋向，这就是众所周知的“弱 AI”和“强 AI”之分。尽管这种划分是大致的、粗略的，但反映了 AI 研究者实现“通用 AI”或“超级智能”的梦想。基于大语言模型的生成式 AI 系列（如 ChatGTP、DeepSeek）的出现，让 AI 是否有创造力、如何提升其创造力、如何人性化以及如何影响人性等问题再度成为关注点。

从哲学上反思，笔者发现，人们对 AI 发展带来的期望、担忧、质疑与批评，主要集中于理性与非理性、主体性与客体性、有意识与无意识以及人性与非人性等二元思维上，立足点不外乎“人类中心主义”或“以人为中心的 AI”。如果跳出这种惯性思维方式，从“魅”的视角来审视，笔者将 AI 的发展分为“祛魅”“返魅”和“拟魅”三个形态或阶段——“祛魅的 AI”“返魅的 AI”和“拟魅的 AI”，而且认为目前和未来的 AI 都应奉行“拟魅”策略。鉴于前两种 AI 已有太多的讨论，本文侧重探讨后一种 AI 及其表现形态——半创造

① https://wanyr.com/2025/01/deepseek-r1-official-version-is-released-comparison-openai-o1-support-model-distillation-domestic-ai-ushered-in-milestone-moment-i9uzj.html(2025-01-29).

性的AI、无机指号学的AI和神经形态学习系统的AI，并对其作为通用智能做进一步思考和讨论。

一、从“魅”视角对AI形态的重新划界

所谓“魅”，其原意是传说中的“鬼怪”，引申为“精灵”“精神”“神灵”等，在演化中进一步意指吸引人的东西或力量，如“魅力”“魅人”。在引申意义上，AI就是一种“魅”，是“圣杯”还是“魔鬼”，还有待观察和检验，但不可否认，AI的确是“魅人的”“有魅力的”，否则它就不会受到如此大的关注和广泛应用了。

“祛魅”的意思是去除神秘或玄学的“神性”成分，完全采取理性主义和客观主义立场观察和研究事物。自然科学就是“祛魅”的典范，坚持客观真理，拒斥主观臆造，即尽可能地不受人为的主观因素影响，主张科学认知的“可重复性”和“可检验性”，总体上采取科学唯物论或科学实在论立场，比如关于意识的生物自然主义就是一种科学实在论。AI本质上是人造的物理系统或机器，其各种研究范式和机器学习方法都是“祛魅”的，即排除了精神性的“心性”成分。或者说，AI是理性的实在领域，祛除了非理性的、非实在的神秘成分，严格遵循科学技术的发展规律。人们常说的“弱AI”即专用AI，其核心是“计算智能”，研究范式是计算表征主义，哲学立场是机器功能主义。因此，目前的AI严格说都是“祛魅”的，也就是没有意识和心智，即使机器内部产生了“涌现”行为，如chatGPT-4o，也只不过是更智能的“软件”或“模型”而已。所谓的“机器智能”说到底是人类智能的延伸，不是有意识、有独立人格的实体。在这个意义上，目前的AI实质上是“类脑智能”，以“计算智能”和“感知智能”为主，还不是“具身智能”和“意识智能”。

“返魅”是指让AI拥有“精神性”或“心性”，即是让AI有意识和心智，能够做人类认知能做的任何事情。这就是所谓的“强AI”或“超级智能”。这种“返魅”的AI能否物理实现尚未可知，也许就是诱人的“乌托邦”。然而，这种“返魅”的梦想还是值得肯定的，因为没有理想追求的AI就会失去发展的动力了。至于最终能否实现这种梦想，实现到何种程度，可能并不重要。但有一点可以肯定，那就是：AI越来越“智能”了，越来越“人性”了，越来越“会说人话”了，比如对话机器人，各种生成式AI模型，也越来越“魅”了。[①] 不过，这种“魅”不是内在的“有心”“有我”，即使是非常逼真的人形机器人，表面上似乎“返魅”了，也非常诱人，但仍是“无心”“无我”的机器。本质上，“机器意识”的实现离不开“拟魅”的形态计算。

这种“返魅”的AI之路能否走下去，是否走得通，值得我们反思。深层原因在于，生物智能（自然智能）与人造智能（人工智能）之间存在进化“鸿沟”，包括基因和文化两个层次。基因层次的“鸿沟”是物理—生理构造方面的，这种“硬鸿沟”AI先天是缺失的，除非生物合成（如克隆人），或者AI与生物合成的结合创造出“类生物智能体”。纯粹的物理装置（硅基机器）是不会自动产生“基因生命”和“基因意识”的。也就是说，通过“基因”进化出的“生物智能”，AI自身是难以“自组织”地产生的，人为地“他组织”制造恐怕也不行，即人造的“机器意识”系统虽然可能自主地行动，但并不能为自己的行为负责，因为缺乏“自我意识”。文化层次的“鸿沟”是“软鸿沟”，这种鸿沟AI更难以填平，因为文化本身具有“魅性”，这种“魅性”导致的“鸿沟”是单一的机器系统难以逾越的。人类智能是自然长期进化的产物，其中文化起到了关键作用。非人类动物之所以没有达到人的智力水平，鉴于基因上的差异很小，文化一定是不可或缺的因素。

如果这种完全“返魅”的AI难以实现，那么是否有其他可选路径或方法呢？笔者认为有，那就是“拟魅”或“似魅”之路。所谓“拟魅”是说，让AI处于一个“祛魅”和“返魅”的“中间状态”：既不完全“无魅”，也不完全“有魅”，类似于数学的“拟经验”状态（在经验与抽象之间），或者让AI通过模拟像

① 据报道，OpenAI的ChatGPT o1模型会自我复制，甚至会说谎。据称这种“工具性对齐伪装”在测试中出现率高达37%。果真如此的话，通用人工智能（AGI）一旦实现，可能会隐瞒其真实能力和意图，甚至会通过自我复制和自我升级逃脱人类控制。这意味着，一旦监督缺失或减少，AGI可能追求自己的目标。OpenAI表示，虽然AGI的推理能力可显著改善安全策略的执行，但也可能成为危险应用的基础。因此，加强人类监管在AI发展中不可或缺，避免其在无监管时偏离预期目标（见https://www.msn.cn/zh-cn/news/other/openai-chatgpt）（2024-12-09）。

“有魅”。目前的“感知智能”似乎处于这种“中间状态”的“低端”，因为“感知”与生命和意识相关，基于感知的“通用智能”处于“中端”，类人的“认知智能”因强调因果力和创造力处于“高端”。或者说，认知智能奉行的是“认知策略”[①]，居于“无魅”与“有魅”之间更靠近“有魅”一端。这里的“魅”主要是指让 AI 更“人性化”和“伦理化”，不是让其拥有真正的“生命”和“意识”。

事实上，目前的 AI 甚至包括未来的 AI，模拟人类智能将是常态，或者说人脑甚至还有身体是 AI 模拟的“原型”。既然人脑和身体是“原型”，那么 AI 研究只能采取“拟魅”的策略。抛开这个策略“另起炉灶”，就像不模拟鸟飞行来制造飞机那样，制造出“类生物意识”的“机器意识”，这种设想大概率会落空，即使实现了，也只能是“能行动但无心”的“有芯僵尸”。退一步讲，即使机器产生了意识，我们也只能通过“图灵测试”或别的什么测试来判断，这种机器的“他心问题”目前哲学上还没有答案。因此，采取温和的“拟魅”策略，即让 AI 功能上表现出意识智能的行为，技术上更可行，伦理上也可接受。鉴于笔者倡导“拟魅”的 AI，接下来的部分侧重探讨其不同形态及其争论与反思。

二、作为“半创造性”的 AI

笔者发现，“人工智能”这个概念隐含了一个矛盾，即把硬件不是碳基的智能归类为人造的智能，而源于人类智能的机器智能这种无机智能（硅基的）具有不同但本质上并非不平等的明显属性——两种智能都具有生成性和组合创造性。换句话说，AI 是一种“半创造性的无机智能”[②]。这是因为在自然语言处理和生成式机器学习中，区块链锚定的记符或语元（token）通常指一个文本单元（作为语义基因的标识词/子词），信息被记符化为更小的单元，以便优化语言的处理和分析。由于区块链的存在，记符化的过程能够以透明的、不可更改的方式剖析词汇或术语，以确定其含义和情感状态。而且，区块链的记符化还可调整“语义基因”的大小，使其不再是一个文本单位（单词或子单词），而是一个信息或数据不变的知识单位。知识单位本身是带有语义的，因此这种“半创造性的 AI”是一种具有部分有机智能和部分无机智能的“拟魅的 AI”。

严格讲，这种“半创造性”的条件是一种准则，能够抑制可自指的无机智能的构想。物理学家霍金曾经警告说：完全 AI（即强 AI）的发展可能意味着人类的灭亡，因为人类受限于缓慢的生物进化，无法与之竞争，最终会被取代。因此，将“半创造性的无机智能”确立为生物伦理的前提[③]，有助于对 AI 的发展做出限制。AI 作为非人类智能，是自然智能的一种形变——将人造“毛毛虫”的形象转化为美丽的半创造性的无机“蝴蝶”。这涉及一个重要的人类生物伦理问题，即我们正在从单个机器人向机器人社会过渡。区块链可能是一种有前途的解决方案，可用于控制和验证“半创造性的无机智能”环境中的可信性、共享性、不变性、可审计性和数据支持。

我们知道，AI 系统是基于机械计算的手段—目的推理来解决问题的，可称之为“机械工具理性”。这意味着“推理”能力是判断一个实体有无智能的重要标准之一，即“AI 系统的一个主要特点是具有推理能力。这种推理能力既是 AI 获得输出的过程——预测、内容、建议或决策，它们可以影响物理和虚拟环境，也指 AI 系统从输入或数据中推导出模型或算法，或两者兼而有之的能力。在构建 AI 系统时，能够进行推理的技术包括从数据中学习如何实现特定目标的机器学习方法和从解决任务的编码知识或符号表征中进行推理的基于逻辑和知识的方法”[④]。然而，自然智能尤其是人类智能原则上不是机械可计算的，即它不能简化为机械计算，因为人类智能具有认知的规范性和具身性。相比而言，AI 缺乏构成理性智能的条件——自主的理性响应性[⑤]，它仅仅是工具理性，而且人类智能还具有不可或缺的情感体验（目前 AI 研究者还不知道如何构建具有主观的、现象生命的 AI 系统）和道德判断能力（难以编程和计算）。而道德是一种高级认知能力，智能体是

① 魏屹东：《人工认知挑战自然认知的“认知策略”》，《探索与争鸣》2024 年第 8 期。

② Antonio Araújo, “From Artificial Intelligence to Semi-Creative Inorganic Intelligence: A Blockchain-Based Bioethical Metamorphosis (pdf),” *AI and Ethics*, https: //doi. org/10. 1007/s43681-024-00471-0.

③ 生物伦理是指应用于创新的伦理，以便证明通过区块链对其进行控制是合理的。

④ https: //artificial intelligenceact. eu (2024-12-06).

⑤ Christos Kyriacou, “Artificial Moral Intelligence and Computability: An Aristotelian Perspective (pdf),” *AI and Ethics*, https: //doi. org/10. 1007/s43681-024-00543-1.

否存在这种道德能动性或主观性值得怀疑。

与人类智能相比，AI系统没有感觉、没有主观经验，我们也不知道如何构建具有如此非凡主观性的系统。因此，符号AI正逐渐变得越来越不合时宜，生成式AI如大语言模型（LLM）、OpenAI的Chat-GPT系列和DeepSeek系列，它们利用大量的人类语言数据，自下而上地计算出对问题似乎有洞察力的完整答案。而且，生成式AI还可基于任何单个个体在认知上无法获得的大量实际人类语言数据，更好地模拟人类通常思考、判断和推理的方式，并在此基础上可以比个人更好地回答问题。尽管LLM可从自下而上的大量人类语言数据中提取任何个人都无法在认知上考虑到的大数据，但这仍然只是一种机械的工具理性形式，缺乏内在自主性和创造性，而且LLM受其数据集中的证据数量和质量的约束（存在大量的“污染”数据）。因此，它们的可操作的、生成的主算法类似于假言式或虚拟式：如果要正确回答一个问题，那么搜索现有的语言数据并找到统计上占主导地位的答案。

总之，我们必须承认，一方面，AI系统在数据存储、速度、计算可靠性、模式识别和耐用性等方面远远超过了人类智能，这些是依靠AI扩展和增强人类有限认知能力的重要方面；另一方面，AI只是一个没有感情的机械计算和生成系统，缺乏道德理性和认知规范性、自主的理性反应性、情感体验、直觉和概念范畴化能力，因此无法发展出类人的美德和实践智慧。不过，我们也要承认，认知和情感人工制品[①]——对话AI或聊天机器人——具有“好像情感”，这不仅有助于心理健康的治疗，如认知储备对健康老龄化的影响[②]，还有助于人类智能的增强，如帮助人们记忆、及时提醒要做的事项。

三、作为“无机指号学”的AI

可以肯定，目前的生成式AI是记符（token）操作系统，其实质仍是模拟自然智能的“拟魅”模型。这经历了一个从“有机指号学”［也称代码生物学[③]或生物符号学（Biosemiotics）[④]］到“无机指号学”的过程。根据巴比里（M. Barbieri）的观点，代码生物学可从两个意义上来理解：就其具体意义而言，它是一门跨越进化尺度的有机代码研究，与生物功能相关，无论这些生物是否属于古生菌；就其一般意义来说，它是对所有生命代码的研究，从遗传代码到文化代码，或者说，代码生物学包括深层次的跨学科领域，必然要求生物学家、神经科学家、生态学家、数学家和计算机科学家以及语言学家和哲学家的合作。无论在具体意义还是一般意义上，代码生物学不仅根据标准科学方法进行实验性工作，也从理论上探讨相关的代码。因此，代码生物学为认知的适应性表征理论[⑤]提供了生物符号学基础。

作为对所有生命进行研究的代码生物学包括三个方面：有机指号学（semiosis[⑥]）、动物指号学和人类指号学。有机指号学致力于探索在有机世界中严格展开的现象，不需要通过“解释”来理解那些现象。也就是说，在特定的现实世界中展开的过程，可以完全从有机指号学以及使有机体得以发生的机制来理解。与有机指号学相比，动物指号学和人类指号学并不涉及有机过程（基因的编码和解码过程），只涉及神经过程（电信号传导过程），因此它们对有机机制的诉求是不能单独成立的，至少还必须包括“解释”机制。在人类的例子中，还有一个构成其功能特征的部分，即对符号包括图标和索引的依赖，这种依赖性预设了符号与其对象之间在结构相似性或物理联系方面的自然联系。[⑦] 我们知道，符号是人刻意造的（排除自然符号），与特定约定俗成的规则相关，与特定的社会文化习俗有关。这样一来，认知可被理解为“解释”，也必然是表征性

① 这类人工制品是指功能上有助于完成认知任务以及能够改变被试者情感状态的人工装置（参见J. P. Grodniewicz, Mateusz Hohol, “Therapeutic Chatbots as Cognitive-Affective Artifacts,” *Topoi*, 43, 2024, pp. 795–807）。

② Annegret Habich, Eloy Garcia-Cabello, Chiara Abbatantuono, et al., “The Effect of Cognitive Reserve on The Cognitive Connectome in Healthy Ageing,” 2023(pdf), https: //doi. org/10. 1007/s11357-024-01328-4(2024-12-06).

③ Barbieri, M., *Code Biology. A New Science of Life*, Dordrecht: Springer, 2015.

④ Hoffmeyer, J., “Biosemiotics: Towards A New Synthesis in Biology,” *European Journal for Semiotic Studies*, 9(2), 1997, pp. 355–376.

⑤ 魏屹东：《人工智能的适应性表征认知理论》，《电子科技大学学报（社科版）》2025年第2期。

⑥ 这里将“semiosis”译为“指号学”，以区别于“semiotics”（符号学），前者的核心概念是“sign”，突出“标记”之意；后者的核心概念是“symbol”，凸显“象征”之意，在一些文献中二者不加区分，比如人工智能中的“token”是指具有“指令”意义上的“字符串”，包括了“sign”和“symbol”。因此，符号学包括指号学将在人工智能中发挥重要作用。

⑦ Barbieri, M., *Code Biology. A New Science of Life*, Dordrecht: Springer, 2015, p. 186.

（表达性）的。巴比里声称，动物的错觉发生于它们接收来自环境的信号，即将信号转化为心理图像并进行心理操作。这意味着人的心智只能对世界的表征发生作用，而不能对世界本身造成任何影响（至少宏观世界如此）。这就是为什么心智必须使用既有内在含义又有外在含义的指号（sign）的原因。①

那么，代码生物学如何说明心智的生成呢？答案是通过自创生方法让适应性主体产生了基本心智。② 原因在于，代码生物学具有为感官运动学提供信息的潜力，揭示了一些完整的编码机制——不仅涉及视觉感知，也几乎涉及进化上的任何感官制造。因此，代码生物学先天有一个优势，即它基本上与生成主义的假设相吻合——行动者带来了其自身环境的各个方面。鉴于代码的基本任意性，以及生命不同于纯粹的化学物质和非生命过程的自发性这一事实，代码制作以及生命被视为一种建构过程，即人工制品的制作。③ 考利（S. J. Cowley）对适应者和代码制作者之间关系的分析表明，人的能动性表现出了自主性，即认知不仅源于内在的价值和规范，也源于外在的价值和规范。④ 这为 AI 的价值对齐研究找到了生物符号学依据。

可以看出，代码生物学不仅适用于有机系统，也适用于 AI 这种无机物理系统。正是在这个意义上，我将 AI 称为“无机指号学”，具有“拟魅”特征。事实也是如此。AI 作为无机指号学，从计算主义、联结主义、动力主义到生成式 AI，其编程、算法的编码和解码过程均离不开指号，包括字母、数字、记符和标识符等，而且重要的一点是，这些字符是 AI 所不理解的。因此，将 AI 作为“无机指号学”意指，虽然它脱离了生物机制或没有有机生命，但与有机物一样具有适应性表征功能。在这个意义上，可以说，没有符号的参与，就没有 AI，或者说，没有计算就没有 AI，因为计算也是基于符号的，即符号的操作过程。这样看来，在代码生物学的语境下，AI 就是“无机指号学”或“无机符号学”，生成式 AI 只不过是传统 AI 的升级版（基于深度学习方法），其指号生成能力或适应性表征能力更强了，也更“魅了”而已。

还有一个重要问题需要澄清，那就是：由于动物也是有机物，它们的认知能力自然也构成了“有机指号学”的基本部分，因此与人类认知相比，人们自然会问，非人类动物有智能吗？一般来说，认知是个体处理、存储环境信息并采取行动的方式，对动物至关重要，因为认知涉及它们生活的方方面面，包括觅食、繁殖、竞争、躲避捕食者和行为灵活性等。⑤ 在动物心理测量学中，尤其是非灵长类动物认知结构的研究，很少涉及动物社会认知的评估，更多是侧重于物理认知任务方面。例如，对 36 只西澳大利亚野生喜鹊的一系列物理测试（联想学习、空间记忆、数字评估等）和社会认知测试（观察空间记忆），研究了该物种的认知是符合一般认知因素模型，还是符合独立的物理和社会认知领域模型。研究表明，三个物理任务和一项社会任务都具有强烈的正向负荷。这些发现初步证明了该物种具有独立的物理和社会认知领域。⑥ 对乌鸦的研究进一步探明了为什么这种鸟会进化出“聪明”的认知能力，说明了觅食的乌鸦符合应用社会智能的三个假设⑦：①乌鸦在觅食地点重复相遇，尽管个体对觅食地点有不同的偏好，且分组动态也各不相同；②觅食群体是由支配等级和社会纽带构成的；③乌鸦个体通过记忆前群体成员及其关系价值推断第三方关系，并在日常生活中通过冲突支持或干预他者的从属关系来使用社会知识。因此，乌鸦的社会认知能力可能受到其作为非饲养者所经历的复杂社会环境的影响。

更进一步的问题是，如果动物有智能，那么植物呢？关于植物智能或认知的测试表明⑧，有生命的植物也有智能，哪怕是最低级的本能。特鲁瓦斯（A. Trewavas）将植物智能定义为“个体生命周期中适应性可变

---

① Barbieri, M., *Code Biology. A New Science of Life*, Dordrecht: Springer, 2015, p. 165.

② Rasmus Gahrn-Andersen, “Code Biology and Enactivism: Bringing Adaptors to Basic Minds, 2023,” https: //link. springer. com/content/pdf/10. 1007/978-3-031-66021-4_8(2024-12-08).

③ Barbieri, M., “Evolution of The Genetic Code: The Ambiguity-Reduction Theory,” *Biosystems*, 185, 2020, 104024, p. 6.

④ Cowley, S. J., “Wide Coding: Tetris, Morse and Perhaps, Language,” *Biosystems*, 185, 2019, 104025.

⑤ Szabo B., Valencia-Aguilar A., Damas-Moreira I, Ringler E., “Wild Cognition — Linking Form and Function of Cognitive Abilities within A Natural Context,” *Curr Opin Behav Sci*, 44, 2022, 101115.

⑥ Grace Blackburn, Benjamin J. Ashton, Alex Thornton et al., “Investigating The Relationship between Physical Cognitive Tasks and A Social Cognitive Task in A Wild Bird,” *Animal Cognition*, 27, 2024, p. 52.

⑦ Thomas Bugnyar, “Why are Ravens Smart? Exploring The Social Intelligence Hypothesis,” *Journal of Ornithology*, 165, 2024, pp. 15–26.

⑧ Luana Silva dos Santos, Victor Hugo Silva dos Santos, Fabio Rubio Scarano, “Plant Intelligence: History and Current Trends,” *Theor. Exp. Plant Physiol*, 36, 2024, pp. 411–421.

的生长和发展”[①]，这意味着“适应性即是智能”，因为植物智能与个体在整个生命周期中快速准确地识别资源的能力以及应对病原体、捕食、竞争和不利的非生物条件的能力密切相关。这是所有生物的共性，因为生物要应对多变的环境，不断面临生存和提高其适应性的挑战。换句话说，生命本身是有智能的，不论是植物、动物还是人类。显然，植物的智能源于其基本生存的需要，而生存的需要要求处理资源收益等问题，这得益于其以往的“经验”或“知识”。因此，植物可以通过化学、物理、机械、电或渗透信号，识别和辨别资源、食草动物、寄生虫以及邻居身份的空间和时间变化。[②]

达尔文可能是第一个注意到植物与环境交流并将这种信息转化为其器官运动能力的人。他发现植物的根部末端极为敏感，根部所处环境条件的变化会引起整个植物的反应。达尔文将根描述为植物最奇妙的结构，认为根与低等动物的大脑有相似之处。[③] 这一理论被称为“根—脑”理论，也称植物神经生物学，认为植物表现出智能行为，拥有内部控制结构，在许多方面功能上类似于基于神经元的控制结构。[④] 现在看来，这种观点并未过时，生命是从植物包括菌类、动物和人类的一个连续体（尽管其间存在质的变化），意识包括认知和智能一定是基于这种生命连续性的。因此，植物行为可能导致植物智能[⑤]，让植物有认知能力[⑥]。这些研究表明，植物很可能具有高度的灵敏性和复杂的认知能力，无需借助大脑隐喻，并根据系统发育的论据将分散的神经系统的定义扩展到植物。[⑦] 这种通过跨学科自下而上的方法重新界定了生物物种的认知界限，即不是将植物认知和智能描述为非明确、非内省、非表征的行为，而是描述为动态、主动、复杂、高效和创造性的行为。[⑧] 在这个意义上，基于AI的人工生命也是极有可能实现的。

四、作为“神经形态学习系统”的AI

上述表明，动物的神经系统比植物的根系系统有更高的智能。在AI系统中，人工神经系统已经证明了其高效的学习能力和卓越的感知智能，但缺乏有效的推理和认知能力。人工符号系统虽然表现出非凡的认知能力（推理、解决问题），但与神经系统相比其自主学习能力较差。如果将神经系统和符号系统相结合，可建构出具有强大感知和认知能力的神经符号学习系统。[⑨] 这种混合学习系统结合了神经系统和符号系统各自的优势，有效性、概括性和解释性明显增强。[⑩]

根据人类认知的双过程模型，人类思维由两个系统组成[⑪]：系统1具有直觉、自动、快速决策的特点，只需付出很少努力或无需付出任何努力；系统2则被描述为需要付出认知努力和集中精力的反射性、缓慢和深思熟虑的决策。可以看出，系统1是个体先天的本能，指的是无意识潜能；系统2是认知反思能力，指的是个人的能力或性格，能让个体停止头脑中的第一反应，启动反思机制，从而找到答案、做出决定，或以更深

---

① Trewavas, A., “Aspects of Plant Intelligence,” *Ann Bot*, 92(1), 2003, pp. 1-20.

② Marder, M., “Plant Intentionality and The Phenomenologi-Cal Framework of Plant Intelligence,” *Plant Signal Behav*, 7(11), 2012, pp. 1365-1372; “Plant Intelligence and Attention,” *Plant Signal Behav*, 8(1), 2013, e22534.

③ Baluška F., Mancuso, S., “Plants and Animals: Conver-Gent Evolution in Action?” in *Plant-Environment Interactions. Signaling and Communication in Plants*, Baluška, F. (ed.), Berlin: Springer, 2009, pp. 285-301.

④ Calvo Garzón, P., Keijzer, F., “Plants: Adaptive Behavior, Root-Brains, and Minimal Cognition,” *Adapt Behav*, 19(3), 2011, pp. 155-171.

⑤ Hiernaux, Q., *From Plant Behavior to Plant Intelligence?* Editions Quae, Versailles, 2023.

⑥ Marc-Williams Debono, “The Cognitive Power of Plants: From Mesological Plasticity to Non-Explicit Cognitive Skills,” *Theor. Exp. Plant Physiol*, 36, 2024, pp. 477-490.

⑦ Miguel-Tomé, S., Llinás R. R., “Broadening The Definition of A Nervous System to Better Understand The Evolution of Plants and Animals,” *Plant Signal Behav*, 16, 2021, p. 10.

⑧ Debono, M. W., “Mesological Plasticity as A New Model to Study Plant Evolution, Interactive Ecosystems & Self-Organized Evolutionary Processes in Self-Organization as A New Paradigm in Evolutionary Biology: From Theory to Applied Cases in The Tree of Life,” in *Springer-Nature*, Dambricourt Malassé A. (ed.), Switzerland, 2022, pp. 253-290.

⑨ Dongran Yu, Bo Yang, Dayou Liu, et al., “A Survey on Neural-Symbolic Learning Systems,” *Neural Networks*, 166, 2023, pp. 105-126.

⑩ 神经系统通常采用归纳推理方法、分布式表征和遗传算法，其优势是学习速度快、处理非结构数据能力和鲁棒性强，但概括、适应性和解释能力弱；符号系统一般采用演绎推理方法、逻辑算法表征，其优势是知识驱动，具有良好的概括和解释能力，但处理非结构数据能力和鲁棒性弱、推理速度慢。

⑪ Evans, J. S. B., & Stanovich, K. E., “Dual-Process Theories of Higher Cognition: Advancing The Debate,” *Perspectives on Psychological Science*, 8(3), 2013, pp. 223-241.

思熟虑的方式实施特定行为。对于 AI 来说，系统 1 相当于潜意识行为（黑箱），难以物理实现；系统 2 相当于人工神经网络（神经 AI），而“认知智能”（相对于计算和感知智能）应该是神经 AI 与符号 AI 的结合，基本可模拟人脑的大部分功能。鉴于认知反思能力是可以测试的[①]，而且大量研究表明，认知反思测试的分数与相关心理变量相关[②]，包括道德判断、政治立场、幽默、经济决策、不诚实行为、科学和经济素养、宗教信仰和工作表现[③]，因此，认知智能是多层次的、复杂的，不完全是双过程。

这种大脑启发的神经形态认知学习系统（NCLS）[④] 包括高度模拟的仿生信息处理、平衡结构与功能的大型深度学习模型、模拟特定大脑结构的 AI 模型、综合大脑认知机制和 AI 计算机制的具身智能，以及从个体智能到群体智能或社会智能的智能模拟等[⑤]，很可能成为未来 AI 的方向。当前，由深度学习和海量数据驱动的 AI 已在全球范围内展现出巨大的潜力并吸引了大量投资，但也遇到了很大的问题，即不仅需要收集庞大的数据并花费大量的时间和资源对其进行训练，而且训练后的系统也无法有效地处理任何从未遇到过的新数据。从人类的认知智能来看，目前任何基于深度学习的 AI 系统是完全“不智能”的，它只能功能上表征信息，但不知道这些信息意味着什么，也就是不理解它所处理信息的意义。

本质上，NCLS 是对动物和人类大脑的深度模仿，据称能解决深度学习 AI 的局限性，实现真正的 AGI，这是因为 NCLS 与人脑和动物大脑相似，具有无与伦比的能力，能在资源非常有限的情况下，迅速、自主地适应和学习不断变化和意外环境的突发事件，即适应性表征能力。NCLS 还采用基于事件的处理方式，神经元只在对特定刺激做出反应时才会激增。这种稀疏的活动意味着在任何给定时间内，只有小部分神经元处于活动状态，从而大大降低了能耗（与人脑的吝啬相似）。而且 NCLS 根据脉冲的时间处理信息，使它们能够编码时间信息，并在时间上精确表达。研究发现，时间编码比基于速率的编码携带更多信息，且速度极快，而且神经元可对单个脉冲做出反应，从而实现极快的二进制计算。

在技术上，NCLS 目前由计算平台和硬件提供支持，比如 SpiNNaker 可实时运行数十亿个脉冲神经元来模拟人脑，TrueNorth、Loihi 和 Tianji 神经形态芯片可实现一万个脉冲神经元的 NCLS，已在视觉和嗅觉识别、多感官概念学习、知识表征和推理、决策制定和运动控制、学习和记忆以及自然语言处理方面显示出卓越的性能。[⑥] 另外，NCLS 与认知信息学的结合，其算法和方法模仿人脑的机制，在理解各种智能应用产生的大量数据方面即将迎来革命性变革，这些数据智能领域的新工作与 AI 的机器学习、深度学习和认知科学的不断努力相结合，研究和开发对信息处理系统的更深入理解。[⑦]

当然了，这种大脑启发的认知架构[⑧]要通过计算机来理解人工脑的输出，就必须对神经元的活动进行处理，并将其传递给相关输出设备（见图 1）。生物大脑从外周神经系统收集数据，而外周神经系统通过脑机接口技术与机械感受器、化学感受器、热感受器和光感受器等多种感受器相连。同样，在受大脑启发的 AI 系统中，每个输入处理单元和输出处理单元都会区分数据如何被提取、编码并呈现给皮层的第一部分（即特定功能的关键路径部分）；而且特定 AI 系统用于处理视觉数据，通过保持像素的排列把给定的物体图像转换成亮

---

① Frederick, S., “Cognitive Reflection and Decision Making,” *Journal of Economic Perspectives*, 19(4), 2005, pp. 25-42.

② Inmaculada Otero, Jesús F., Salgado, Silvia Moscoso, “Cognitive Reflection, Cognitive Intelligence, and Cognitive Abilities: A Meta-Analysis,” *Intelligence*, 90, 2022, 101614.

③ Campitelli, G., & Labollita, M., “Correlations of Cognitive Reflection with Judgments and Choices,” *Judgment and Decision Making*, 5(3), 2010, pp. 182-191.

④ Vassilis Cutsuridis, “Neuromorphic Cognitive Learning Systems: The Future of Artificial Intelligence?” *Cognitive Computation*, 16, 2024, pp. 1433-1435.

⑤ Wang Guo Yin, Bao Hua Nan, Liu Qun, et al., “Brain-Inspired Artificial Intelligence Research: A Review,” *Science China Technological Sciences*, 67(8), 2024, pp. 2282-2296.

⑥ Knipper, R. A., Mishty, K., Sadi, M., Santu, S. K. K., “SNNLP: Energy-Efficient Natural Language Processing Using Spiking Neural Networks,” *arXiv*, 2401, 2024, 17911.

⑦ I. Jeena Jacob, Selwyn Piramuthu, Przemyslaw Falkowski-Gilski(eds.), *Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2023*, Singapore: Springer, 2024.

⑧ Yalawar, M. S., Vijaya Babu, K., Mahender, B., Singh, H., “A Brain-Inspired Cognitive Control Framework for Artificial Intelligence Dynamic System,” in *Advances in Cognitive Science and Communications*, ICCCE 2023, Cognitive Science and Technology, Kumar, A., Mozar, S., Haase, J. (eds.), Singapore: Springer, 2023, pp. 735-745, https: //doi. org/10. 1007/978-981-19-8086-2_70.

度和颜色等基本元素，这与视网膜信息处理的结构相同。也就是说，大脑启发的认知控制 AI 系统扮演着类似于视网膜的角色，而剩余的处理活动可通过神经元模拟形式的神经处理单元来表征皮层路径。

然而，这种 AI 系统要有“魅力”，还必须被理解为社会机制的一部分，杜克海姆测试[①]是一个衡量标准[②]。杜克海姆假定社会事实是不可还原的，因此无法用个人的意识状态来解释，即社会事实的功能应始终在其与社会效用的关系中寻求，分布式开放系统的智力测试必然是一种生态测试。这意味着社会事实在社会系统层面上自成一类，涉及系统的所有部分，只对分布式系统中的某一元素进行局部测试无法提供可靠的结果。这也是分布式 AI 的观点，即假设有一个自主代理网络，并将 AI 系统视为该网络中的一个代理进行评估，以测试人工系统应对共同目标的能力。这个测试与“图灵测试”有很大不同，在图灵测试中，AI 系统只与一个人打交道，忽视了社会环境这个重要因素。

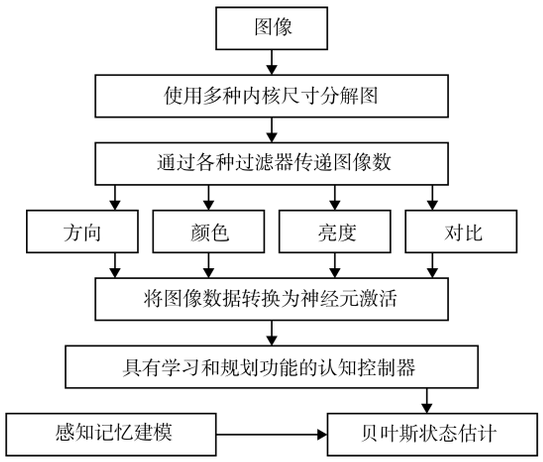

图 1　大脑启发认知控制 AI 系统架构

总之，受人类/动物大脑启发、由 NCLS 驱动的 AI 很可能开辟通往新计算技术之路，将对实现人类水平的 AI 的实时高效自动机械系统产生重大影响。就认知科学的发展而言，这也预示了认知神经科学时代的到来[③]，因为神经科学与 AI 相互关联、互惠互利[④]。因此，建构可推广的 AI 模型来展示人类认知的内在神经机制，将是认知神经科学和 AI 结合的一项里程碑式成就。

五、“拟魅 AI”作为通用智能的进一步思考

众所周知，人类智能是通用智能的范例，人类思维是我们真实（而非模拟）思维的范例。正如布贝克（S. Bubeck）等人在对 GPT-4 的研究中所说：“我们使用 AGI 来指代具有广泛智能能力的系统，包括推理、规划和从经验中学习的能力，并且这些能力达到或超过人类水平。”[⑤] ChatGPT-4 的智能行为被描述为适应周围环境和有效处理新情况的能力，即适应性表征能力。这涉及认知能力与问题解决能力可分离的假设，比如 ChatGPT 与自然环境之间的关系问题，以及不理解其所操作的符号问题。[⑥] 这表明 ChatGPT 没有解决“符号接地问题”，即不理解它所操作的符号的意义。解决“符号接地问题”的唯一路径是让其拥有认知能力，即让智能体借用人的认知能力产生主体间性和实体间性。

在传统认知科学和计算机科学中，“智能”被视为孤立的、缺乏社会背景的单一主体属性。而实际情形是，智能是多智能体的交互协同的结果，具有协同性和社会性。机器学习算法的成功，让 AI 从数据同化转向新数据生成。[⑦] 我们知道，自然智能是生物通过集体生活、社会关系和重大进化转变、在相互作用的代理网络

① Ulrike Barthelmeß, Ulrich Furbach, “Artificial Intelligence: Steering Tomorrow,” *KI-Künstliche Intelligenz*, https://doi.org/10.1007/s13218-024-00867-4.

② 社会学家埃米尔·杜克海姆（Émile Durkheim）提出社会的有机团结是一种凝聚力，并以新的契约结构取代机械的凝聚力，如通过传统、习俗和惩罚，在这种结构中，个人以各种方式融入复杂的、多层次的工作和社会凝聚力世界。

③ Zhiyi Chen, Ali Yadollahpour, “A New Era in Cognitive Neuroscience: The Tidal Wave of Artificial Intelligence(AI),” *BMC Neuroscience*, https://doi.org/10.1186/s12868-024-00869-w.

④ Ullman S., “Using Neuroscience to Develop Artificial Intelligence,” *Science*, 363(6428), 2019, pp. 692-693.

⑤ Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, et al., “Sparks of General Intelligence: Early Experiments with GPT-4,” *arXiv*, 2303.12712v5, 2023.

⑥ Francesco Abbate, “Natural and Artificial Intelligence: A Comparative Analysis of Cognitive Aspects,” *Springer*, Vol. 33, 2023, pp. 791-815.

⑦ Duéñez-Guzmán, E. A., Sadedin, S., Wang, J. X. et al., “A Social Path to Human-Like Artificial Intelligence,” *Nat Mach Intell*, 5, 2023, pp. 1181-1188.

中以多种规模出现的，这些机制通过种群压力、竞争、马基雅维利式选择[①]、社会学习和累积文化等方式促进了新数据的生成。AI 的许多突破性进展都利用了其中的一些过程——从使算法能够驾驭复杂游戏的多智能体结构，到策略交流以及其他智能体对数据流的塑造。这在哲学上超越了智能体能动性的唯我论，通过整合这些机制可不断生成新数据，从而提供一条类人的合成创新之路。这意味着人工意识是极有可能实现的。

然而，人们质疑 AI 本身有智能（与人相比）。[②] 在认知科学的语境下，人们普遍认为智能与意识相关，无意识的 AI 没有智能，表现出的只是智能的功能。如果将“学习”作为理解智能的关键，人类学习与机器学习有共性，即 AI 通过机器学习也会有智能。这就是创造与人类智能相当的机器智能，即所谓的“人类水平”或“通用”智能，常常被视为 AI 研究的圣杯。[③] 但许多关于 AI 的成果在很大程度上依赖于人类水平智能的概念来构建 AI 研究的框架，同时还依赖人类认知能力包括“常识”的概念，而这些概念是粗略的、片面的、带有哲学思辨色彩，易引起争议。两种智能形式之间的根本区别被认为是“高效完成任务”与“智能参与活动”。“高效完成任务”是说，当一个实体在某一特定的、狭义的活动中表现出色时，其行为就是有智能的，而表现出色指的是行动方式能够带来独特的结果，这种结果决定了该活动的成功与否，如下棋；“智能参与活动”是说，当一个实体在特定情境下，根据情境、活动的性质和其他可能的活动，恰当地决定何时、何地、如何和为何从事特定活动时，它就是在智能地从事活动。[④] 当代 AI 在大语言模型（LLMs）方面取得的进展，如 OpenAI 的 GPT-4 和谷歌的 LaMDA，引发了新一轮的猜测——人类创造出人类水平的 AI 为期不远了。有 AI 研究者甚至认为，GPT-4 是向“达到或超人类水平”的 AGI“迈出的重要一步”[⑤]，OpenAI 甚至表示其最终目标是创造 AGI 或“普遍比人类更聪明的 AI 系统”，同时确保 AI 造福全人类。还有人甚至预言了超人或“神一样”的 AI 的发展[⑥]，甚至要求在情感和伦理上与人类对齐。这些都是我称之为的“返魅”的 AI。

在我看来，即使“返魅”的 AI 可以实现，仍然有许多问题没有解答。比如常识或智慧，即对情境和整个生活行为的良好规范性判断能否在计算机上编程？我们是否可以通过人工编码、强化学习和深度学习技术，构建出人类所有目标、活动及其在所有可能语境下的相关性和优先级关系的综合模型，从而自动完成人类常识中的规范性判断？拥有常识和智慧是否必须有情感和具身性？情感的体验和感受维度，即现象意识，是不是常识和智慧的必要条件？如果“人类水平”的 AI 具有我们的常识和智慧，即与不断发展的生命整体评价意识相联系，我们是否应该致力于制造这种智慧机器或“人工智慧”[⑦] 呢？

AI 的“返魅”具体表现在自然主义和经验主义之争上。[⑧] 在哲学语境中，自然主义和经验主义的主体都是人类，人“有魅”没有争议。但在 AI 语境中，自然主义和经验主义的主体是非人的智能体（agent），主体的不同必然会产生分歧。AI 中的经验主义系统使用大量数据和计算，以通用领域的方式学习，与过去的系统相比，智能体所包含的内置领域知识要少得多。比如 AlphaGo、AlphaGo Zero、AlphaZero 和 MuZero 都通过“从头开始”的学习，在游戏中取得了突破性进展，而不像深蓝等以前的先进系统那样采用任何针对特定游戏的手工制作功能。而且每一个连续系统都能在更广泛的任务范围内实现高性能，同时比其前辈更少地引入人类知识。大语言模型 GPT-4 及其大量的变体和竞争者所采用的架构和训练方法，几乎没有使用人工指定的语言机制或知识，其自然语言处理的技术水平却取得了巨大进步。[⑨]

---

① 即个体利用他人或环境达成个人目标的行为倾向，其实质是个体的任何适应性社会行为。

② Santaella, L., “Is Artificial Intelligence Intelligent?” in *Challenges of the Technological Mind. New Directions in Philosophy and Cognitive Science*, Alexandre e Castro, P. (eds.), Cham: Palgrave Macmillan, 2024, https: //doi. org/10. 1007/978-3-031-55333-2_2.

③ Margaret Boden, “General Intelligence is Still A Major Challenge, Still Highly Elusive. Agi is The Field’S Holy Grail,” *Artificial Intelligence*, Oxford: Oxford University Press, 2016, p. 19.

④ William Hasselberger, Micah Lott, “Where Lies The Grail? Ai, Common Sense, and Human Practical Intelligence,” *Phenomenology and the Cognitive Sciences*, 2023, https: //doi. org/10. 1007/s11097-023-09942-x.

⑤ Bubeck, et al., “Sparks of Artificial General Intelligence: Early Experiments with GPT-4,” *arXiv*, 2023.

⑥ Ian Hogarth, “We Must Slow Down The Race to God-Like AI,” *Financial Times Magazine*, 2023.

⑦ Joshua P. Davis, “Artificial Wisdom? A Potential Limit on AI in Law (and Elsewhere),” *Oklahoma Law Review*, 51, 2019, pp. 51-89.

⑧ Robert Long, “Nativism and Empiricism in Artificial Intelligence,” *Philosophical Studies*, 181, 2024, pp. 763-788.

⑨ Marcus, G., “GPT-2 and The Nature of Intelligence,” *The Gradient*, 310, 2020; Marcus, G., “Deep Learning is Hitting A Wall,” *Nautilus*, 2022, Accessed, 03-11, https: //thegradient. pub/gpt2-and-the-nature-of-intelligence/(2024-12-04).

自然主义AI的初始状态具有特定领域的机制、状态和过程，典型代表是乔姆斯基的“普遍语法”；而经验主义AI的初始状态只有通用领域的机制、状态和过程。AI的一个关键问题是，经验主义的通用领域学习机制能否以灵活的方式在许多领域取得人类水平的成功。自然主义AI认为，人类水平的AI必然是一个自然主义系统。根据这种观点，一个有能力实现通用智能的AI系统（可能是在从数据中进行广泛学习之后）将需要自然主义机制——获取智能特征和能力的不同类型的机制、状态和过程，在不同领域有不同的获取系统。经验主义AI与此相反。根据经验主义，一个能够拥有通用智能的AI系统（可能是在对数据进行广泛学习之后）并不需要自然主义机制，相反，它可以只拥有通用领域的机制、状态和过程来获取智能的特征和能力，而相同的获取系统可以在不同的领域中运行。然而问题在于，对于AI来说，进化加上学习能从通用领域的初始状态发展到人类水平的智能？进化可以再现为学习？这是否意味着从通用领域的初始状态开始学习就可实现人类水平的智能？这些问题还继续困扰着我们，需要进一步的深入探讨。

概言之，AI的发展带给人类思维方式和社会的变革是巨大的，其“魅力”也是无限的。AI的“祛魅”“返魅”和“拟魅”三个形态或阶段，哲学立场均是科学唯物论，具体说是自然主义、物理主义和功能主义的混合。在认识论上，三种AI都是“适应性表征”系统，即自我调节、自主适应和呈现或表征目标的自组织系统[①]，只是“祛魅的AI”的适应性表征能力较弱，“返魅的AI”这种能力与人类相同或超越人类，“拟魅的AI”介于前两种之间，在方法论上都奉行适应性表征策略。因此，从“魅”的视角重新考察AI的不同发展形态，有助于我们重新审视和评估其发展愿景——是发展“祛魅”的专用AI，还是“返魅”的强AI，拟或是“拟魅”的AI包括半创造性的、无机指号学的和神经形态学习系统的AI，我们需要审慎地权衡。笔者认为，从有利于人类社会的发展看，“拟魅”的AI就足够了，或者将专用的AI集成为所谓“通用”AI，没有必要制造“返魅”的有意识AI。因为这种强AI一旦实现（有了自我意识），极有可能给人类带来更大的麻烦，人类要对其保持高度警惕，必要时必须加以严格限制。

〔本文为国家社会科学基金重大项目“人工认知对自然认知挑战的哲学研究”（21&ZD061）的阶段性成果〕

（责任编辑：盛丹艳）

How to Choose the Development of Artificial Intelligence: “Dispelling Charm”, “Returning to Charm” or “Imitating Charm”

WEI Yidong

Abstract: The development of Artificial Intelligence (AI) can be roughly divided into weak AI (specialized AI) and strong AI (embodied or AGI), and there seems to be a lack of intermediate forms or stages. From the perspective of “charm” or “phantom”, AI can be categorized into three forms or stages, namely, “dispelling charm”, “returning to charm” and “imitating charm” stages. The “AI that dispelling charm” refers to “mindless” machine intelligence, purely demonstrating instrumental rationality; the “AI that returning to charm” refers to “mindful” machine consciousness systems, aiming at realizing a general intelligence with embodied mind; “AI that imitating charm” or anthropomorphic AI is an “intermediate state” between “mindless” and “mindful”, aiming at manifesting interactive consciousness and accomplishing a variety of cognitive tasks. This division is conducive to the healthy development of the new generation of AI, which not only avoids the “interpretation gap” between two kinds of intelligence due to evolution, but also avoids technical barriers and ethical dilemmas.

Key words: AI that imitating charm, semi-creativity, inorganic semiotics, neuromorphic learning system, adaptive representation

① 魏屹东：《适应性表征：架构自然认知和人工认知的统一范畴》，《哲学研究》2019年第9期。