·人文社科与科技前沿交叉研究·

人工智能的适应性表征认知理论

□魏屹东

[摘　要] 从认知哲学探讨智能的生成问题是一个重大挑战。这需要一个说明认知机制的概念框架“适应性表征”。“适应性表征”作为不同层级自组织系统的内在机制和解释范畴，构成了人工智能走向通用性、解释性和可靠性的基础。这种关于智能生成的适应性表征认知理论包括假设、推论和原则以及不同层级结构的交互过程，旨在说明智能的生成是一个自组织实体或系统的不同层级结构通过适应性表征进行交互涌现的结果。在适应性表征视域下，物理系统表现为属性的自反应和自呈现，生物系统表现为生命的自适应和自繁殖，认知系统表现为自学习和自表达，人工智能系统表现为机器学习和自复制，这些不同的表征方式恰恰说明适应性表征是所有自组织系统的通用机制，通用智能的通用性就是适应性表征，这意味着人工智能的不同领域都具有适应性表征特性或功能，建构人工智能系统就是创造适应性表征系统。

[关键词] 认知；智能；自组织系统；人类智能；人工智能；适应性表征

[中图分类号] TP18　　[文献标识码] A　　[DOI] 10.14071/j.1008-8105(2025)-1001

引言

当代人工智能（AI），其研究策略总体上是模拟人类认知或智能做出的。20世纪，AI的三个主要认知范式是相互独立的：起初的结构模拟的联结主义（20世纪40年代），后来的功能模拟的计算主义（也称认知主义）（50年代），再后来的行为模拟的动力主义（也称行动主义，不同于心理学中的行为主义）（90年代）。21世纪出现的机器学习领域的强化学习、深度学习、强化深度学习、计算智能以及不同范式结合的混合认知或智能，包括自然语言大模型GPT系列等生成式AI，严格讲仍然是以计算表征方法为主的路数，尽管其中充满了争论（如表征主义与非表征主义、符号AI与统计AI、经验主义与理性主义等）。

这些不同研究范式和方法尽管各自独立，但它们的目标——实现类人的智能——几乎是一致的。具体来说，计算主义假设智能存在于思维和理性中，动力主义假设智能存在于感知和行动中，结构主义假设智能存在于功能模拟中，生成主义假设智能存在于部件的积木式组合中。有没有一个理论来统一这些不同的AI范式呢？当然有，基于信息科学的机制主义就是走向通用人工智能的一种有益理论尝试[1]，它假设智能存在于主客体的信息生态演进过程中，因为不论是人脑还是人工脑，都是信息加工系统，但机制主义的“主客互动的信息生态过程”[2]128~132全局模型还不足以说明生物层次的意识现象，因为意识经验（感受性）不能还原到信息层次（电脉冲信号），其中必定存在多个跨层次的质变，而且该模型中的“主客互动”已经暗含了意识主体的存在。还有，混合认知也是一种尝试，

[收稿日期] 2025－03－09
[基金项目] 国家社会科学基金重大项目（21&ZD061）.
[作者简介] 魏屹东，山西大学哲学学院教授.
[引用格式] 魏屹东. 人工智能的适应性表征认知理论[J]. 电子科技大学学报（社科版）, 2025, 27(2): 1-19. DOI: 10.14071/j.1008-8105(2025)-1001.
[Citation Format] WEI Yi-dong. A cognitive theory of adaptive representation for artificial intelligence[J]. Journal of University of Electronic Science and Technology of China(Social Sciences Edition), 2025, 27(2): 1-19. DOI: 10.14071/j.1008-8105(2025)-1001.

但这种混合严格讲不是统一的理论，而是一种“拼接”或“嫁接”（相互嵌套或结合），比如涌现论的生物符号学化，AI的生物学化以及人–机交互、机–机交互的协同化。我认为，AI的统一理论必须是基于自然认知（人和动物）的科学理论（哲学立场是科学的唯物论或科学实在论），这势必要求多学科的交叉——认知科学、计算机科学、心理学、脑科学、系统科学和信息科学等，也涉及智能科学理论的重大变革，就像牛顿力学统一了伽利略力学和开普勒力学一样。

鉴于现有AI范式有一致的目标（智能生成），都接受“目标导向”理论，它们建构的智能系统也是“目标导向”系统。这种“目标导向”理论或系统有一个共同的特征——适应性表征（adaptive representation）[3~4]，也就是说，它们都是“适应性表征”系统，或者说，适应性表征是它们的通用机制和解释框架。这是因为，结构模拟的联结主义（人工神经网络）的目标是通过模拟人脑的结构来实现类人的智能，功能模拟的物理符号假设的目标是通过符号计算模拟人脑的功能；行为模拟的动力主义的目标是通过感知–行动来实现类人的行为，如各种机器人；生成式AI通过大量预训练产生智能行为，如各类大语言模型。这些不同的智能生成过程都可视为适应性表征行为，如大语言模型中不同token（语元）之间的关联呈现，表明token关联度体现了人类语言习惯的自动提取的语言痕迹，具有语境相关的统计性质[5]。

如果这一假设成立，那就可以通过适应性表征来统一这些不同范式，再通过适应性表征来解释基于大脑的各种心理现象（意识、心智、自我等）和通用人工智能（通用机制是适应性表征）。换言之，从适应性表征方法论来看，一切心理或精神现象都是适应性表征过程，也是适应性表征的结果[6]；人工智能是人造的机器智能，是自然智能的物理实现或延伸，当然也不例外[7]。这就是适应性表征认知观，它假设智能存在于适应性表征过程中。这个过程或机制的细节需要认知科学和人工智能领域的专家来探明，这里仅从哲学的高度“顶层设计”，至于这种理念是否合理，接下来的部分将论证不同AI范式的适应性表征的哲学假设、认知原则、层级结构及其交互过程、不同系统的适应性表征功能以及哲学反思。

一、智能系统的假设、推论和原则

关于自组织系统，无论是物理的、生物的还是认知的、智能的，就其内在发生机制和外在表现来说，本文提出如下假设和推论：

1. 世界假设：自组织的物理世界（系统）和基于这种世界的精神世界（意识、心智）都是适应性表征系统，具有内在的能动性（倾向性、反应性、生命力、感知力）。因此，宇宙、地球系统、生物体、人类和人工智能体，都具有适应性（自组织、自复制、自调整）和表征性（属性呈现、语言表达、符号象征）。

2. 理论假设：研究自组织系统（物理的和精神的）理论也是适应性表征系统。因此，科学理论和AI理论这些由人类智能创造出的人工产品是人类认知的结果，其目标导向和语言表征功能更强。因此，理论是对现实的再现（表征），现实是理论的表征对象。

3. 智能迁移推论：根据上述世界假设和理论假设，自组织物理世界的适应性表征属性可逻辑地迁移到人工世界，即人类智能可迁移到人工智能。这是生成式人工智能和通用人工智能得以实现的内在逻辑。

4. 普遍性推论：如果上述假设和逻辑推论成立，那么适应性表征不仅体现在物理或自然世界，也会反映在研究物理世界的理论中，即科学理论也是适应性表征系统，进而依据这种理论创造的人工认知或智能系统，也是适应性表征系统。

上述假设和逻辑推论蕴含了两个原则：

原则1：逻辑同一性

如果物理自组织系统是适应性表征系统，那么基于自组织系统的自然认知和人工认知系统也都

是适应性表征系统。由此可进一步推知，研究这些系统的理论——自然科学、认知科学以及AI理论也应该或必须是适应性表征系统。我已阐明自然演化系统、科学认知是适应性表征系统，由人类智能衍生出的人工智能也是适应性表征系统[8]。

原则2：目标同一性

AI的目标是模拟和实现大脑功能。就智能系统（大脑、人工脑）而言，认知科学研究其运作机制（大脑是如何思维的），已有的AI范式以不同方式（计算的、感知的和认知的）模拟脑功能，它们本质上都是适应性表征系统。就AI来说，它实质上是模拟人类认知或智能的工程或技术，已有的范式（符号的和统计的以及混合的）均具有适应性和表征性（图1），特别是具身AI，更是要将生物身体的功能嵌入机器（具身性的物理实现），文化特征（信念、习俗、道德）嵌入人工系统恐怕也只是时间问题。

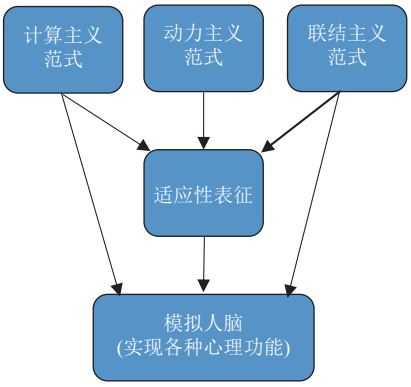

图 1　不同人工智能范式与适应性表征的关系

二、适应性表征的层级结构及其涌现机制

上述分析表明，基于自组织的适应性表征系统是一个层级结构。据笔者考证，适应性表征概念已经成为当代计算机科学和AI中的一个重要概念，用于说明人工主体或人工智能体（Artificial Agent）主动或被动地适应环境的行为。这个概念最初是海里根（F.Heylighen）对康德知识论的一种推广①，当时仅提及物理科学和认知科学中的元表征和变化适应性问题，没有意识到适应性表征可作为智能生成机制和解释框架。其实早在1975年，霍兰（John H. Holland）就通过“隐秩序”说明了适应性如何建构复杂性，探讨了自然系统和人工系统的适应性问题[9]，但没有谈及认知和表征问题，也没有推广到AI领域。

事实上，真正将适应性与表征概念结合起来是认知科学和AI领域自觉或不自觉的研究工作，但其重要性却被有意或无意地忽视了。不过，这个概念的最初洞见还是源于康德哲学。康德认为，感知只是被观察到的现象的心灵（心智）过滤器（适应性表征装置）。按照这一主张，（世界上）存在着人类心智不变和先验经验的原则，比如你可能已经在大脑中留下了空间的一个笛卡尔表征，一个时间、颜色分离的观念。海里根对康德的这种思想进行了修正，认为其中的原则不应该是不变的和必要的，相反，适应性表征中存在着经验组织的替代原则，这为心灵哲学和人类认知及人工智能研究开辟了一条新进路[10]。适应性表征为什么会成为探讨智能行为（自然的和人工的）的关键概念和方法论？接下来的小节将从物理的、生物的、自然认知和人工认知包括人工神经网络、机器学习的适应性及其相互作用机制等方面来分析。

（一）适应性表征的不同层级及其交互

适应性表征作为一个架构自然认知和人工认知的中介性范畴，有四个层次：物理的、生物的、认知的以及它们的相互作用生成的人工认知或智能，即物理系统→生物系统→自然认知或智能→人工认知或智能。物理层次是最基本的，作为物理系统，它们没有意识，是被动适应环境或目标的，如温度计、恒温器，我称之为物理适应性，如热胀冷缩、自组织演化、自由能、自相似、奇怪吸引子、量子纠缠等。生物层次是生命系统，它们的最基本组成虽然是物理的，即可分解为分子、原子甚至亚原子，但组合起来后涌现出生命甚至意识，如植物和动物，它们是本能地适应其环境的；而作为动物另类的我们人类是有意识主动适应其环境，甚至认识和改造世界，这就是人类的形而上性（抽象思维和追求真理的天性），由于形而上性的出现，人类此时的主体性凸显。主体性之所以涌现，在生物学意

义上可能是行动–感知耦合的结果，我称之为生物适应性。认知层次包括自然认知和人工认知，前者基于生物层次的神经系统特别是大脑，后者是人造的智能系统（人工神经网络），如计算机、社交机器人，这种认知系统搜索性、探索性地适应环境，我称之为认知适应性，它是基于物理和生物适应性的，甚至还包括文化适应性。

按照这种划分，适应性包括主动适应性（自主性）和被动适应性（受外部条件约束，如刺激–反应）。不管是主动还是被动，适应性显然是所有自组织系统的共同特征，只是这些系统遵循的规律有所不同。物理系统的适应性遵循物理学规律，如热胀冷缩、自组织演化。这种物理适应性一定以某种方式嵌入生物系统而产生了更强的适应性，如生物系统的自然进化。生物适应性虽然是基于物理适应性的，但遵循的主要是生物学规律，因为从物理层次到生物层次之间有许多环节（原子、分子、大分子、基因、蛋白质、神经网络），涉及从物理存在到生物进化的许多微观机制，系统科学如信息论、耗散结构论、协同学、混沌学等就揭示了其中的一些演化规律。对于涉及身体的经验或现象特征，知觉现象学和具身认知科学则给出了部分说明，即体验一定是基于身体的，认知是寓于身体的，也必然是涉身的。在我看来，生物适应性一定以某种方式嵌入认知系统从而产生认知适应性，或者说，身体的适应性一定以某种方式渗透于认知的适应性。这些还未揭示的“某种方式”是未来认知科学、脑科学（包括脑机接口）和人工智能要研究的课题。

然而，可以肯定的是，人的认知适应性一定遵循认知科学和脑科学规律，而且基于认知系统（大脑）的推理、心理–行动等，超越了物理的和生物的适应性。人的认知适应性也一定以某种方式嵌入人工认知系统，如设置语境（知识库）和编写算法，甚至通过嵌入情境–觉知系统应对意外。也就是说，人工认知系统是三种适应性的混合或融合。这里所说的混合是指，物理学、认知科学和计算机科学都介入了AI，如物理硬件、各种软件以及硬件嵌入智能体。在主体性（区别于自动性）意义上，纯粹的物理系统严格讲是缺乏自我意识的实体，没有主体性，不能以主体身份出现；有意识的生物体能够意识到自我特性，有一定的学习能力，以主体身份出现；认知系统也可作为主体出现，不过是以行动者或行为体的方式，AI中是以智能体（Agent）及其组合的方式，具有了某种能动性（Agency）。

霍兰在《自然与人工系统中的适应性》[11]中对两种系统的适应性作了深入探讨，认为遗传算法在复杂适应性系统研究中扮演着越来越重要的角色，比如从经济理论中的适应性主体到在复杂设备（如飞机涡轮机和集成电路）设计中使用机器学习技术。生物的适应性是我们最熟悉的形式，生物体通过重新排列遗传物质在它们所处环境中生存下来。霍兰提出了一种允许这种复杂相互作用的非线性数学模型，说明了该模型如何修改传统的数学遗传学观点，并用于经济学、生理心理学、博弈论和AI等领域。这意味着，适应性是一种跨越物理、生物和人工系统的普遍现象。霍兰最初将适应性概念应用于具有有限数量参数的简单定义的人工系统，后来继续探索其在广泛复杂的、自然发生过程中的应用，特别是集中于具有非线性方式相互作用的多因素系统上。在此过程中，他说明了共适应和共进化的主要影响：积木或图式的出现，它们被重组并传递给后代来提高、创新和改进。这种积木或图式的无限组合，使得AI成为可能。

明斯基在《心智社会》中描述的智能行为就是通过不同智能体的积木式组合产生的。他展示了许多不具有思维的微小部件（小程序）可以组成思维，并将这种组合称为“心智社会”，其中的小部件被称为智能体；每个智能体本身只能做一些低级智能的事情，这些事情完全不需要思维或认知，但将这些部件以非常特别的方式汇集到社区中，就会产生真正的智能[12]2。这里的“组合”，类似于德勒兹（G.Deleuze）和瓜塔里（F.Guattari）所说的“装配”（Assemblage），“装配”不是机械意义上的组装，而是强调从环境中选取一些要素，并用特定关联方式把它们组织起来形成某种智能行为，“我们将把每一个由奇异性和特征组成的合成体称为装配，这些奇异性和特征是在流动中被提取、被选择、被组织、被分层而来的，并通过这种方式使之能人为地、自然地聚集在一起。在这个意义上，一台装配就是一个名副其实的发明。”[13]406显然，不论是“组合”还是“装配”或“配置”，这些概念都突出智能体的有机特征，即“整体大于部分之和”的整体性。智能行为就是在这种整体性组合

中涌现的，因为每个单一组件（智能体）本身并不具有智能，就像单个神经元不具有意识但大脑有了思维能力一样。

按照这种设想，要解释思维是什么，就必须弄清思维是如何由无数的成分组成的。这些组件比任何有智能的生物都小得多，也简单得多。因此，组成智能的那些智能体究竟是什么？明斯基称之为“建设者”，它们是如何完成工作的取决于它们之间的连接方式，也就是组合方式。从外部看，如果不知道它们的运行机制，就会认为它们“知道”如何工作，好像有了智能；从内部看，又看不到它们有知识，只能看到一些开关以不同的方式排列，以便相互打开或关闭。也就是说，作为“建设者”的智能体是否真正“知道”如何工作的知识，答案取决于我们如何考察或取决于我们观察的位置[12]。这意味着认知现象是观察者依赖的，比如，第二代认知科学的“4E+S”研究纲领均使用英语的被动词，即认知“被具身、被嵌入、被生成、被延展、被情境化”（embodied, embedded, enacted, extended, situated），这显然是从观察者的立场考察认知的，因此认知是被组合的、被配置的以及被认识的。

霍兰在《信号与边界》[14]中认为，我们通过信号与边界可建构复杂适应性系统的“积木”，复杂适应性系统包括生态系统、政府、生物细胞和市场，其特点是边界和信号的复杂等级安排。比如，在生态系统中，生态位充当半透性边界，气味和视觉模式充当信号。尽管有大量关于不同复杂适应性系统的数据和描述，但如何“控制”这些系统仍有许多未回答的问题。在霍兰看来，理解这些系统复杂的信号/边界层次结构的起源是解答这些问题的关键。他通过生成其信号/边界层次的机制，提出了用于比较和控制复杂适应性系统的总框架，为发展智能体、生态位、理论和数学模型的框架开辟了一条道路，包括理论建构、信号处理主体、作为信号/边界相互作用表征的网络、适应性、重组与复制，并使用概率理论表征边界层次，提出将无限生成的系统作为被检验的模型绑定到一个单一框架中的方法，以及用一个简单的有限生成的多细胞有机体来发展模型。

在我看来，这有可能产生人工生命，因为人工生命就是通过模拟和合成人工介质中的类生命过程来研究生命系统的基本性质的跨学科，该领域为研究如何在由简单交互规则控制的系统中产生高级行为提供了强大的工具集。这是一种通过自然的和合成的适应性行为的新领域，涉及行为学、心理学、生态学、AI、人工生命、机器意识、机器人学等相关领域，以促进对行为和潜在机制的理解，从而使自然的和人造的智能体（人工生命）能在不确定的环境中适应和生存。目前这方面的研究集中于定义明确的模型——机器人学、计算机模拟和数学，这些模型有助于描述和比较动物和人造生命（Animats）中适应性行为的各种组织原则或架构。

在表征方面，四种系统的适应性含义稍有不同。物理系统的表征是物理属性的呈现，如热、冷、发光、固态、液态和气态；生物系统的表征是刺激–反应的功能呈现，如神经元的放电、脑电波、行动–感知耦合；认知系统的表征是符号表达，包括自然语言和抽象符号，如逻辑和数学、AI算法。概括起来，表征有两种含义：一是物理属性和生物功能的“呈现”，“呈现者”是物体、系统或生物体；二是抽象的语言–符号“再现”，“再现者”是人类和智能体。“呈现”意味着从系统内部自发地表现出来；“再现”意味着二次表达，即反映到心中的目标客体以中介形式展现出来，如概念描述、命题表达、图像展示。换句话说，就表征的内容（语义）而言，一种是属性、功能的自然呈现；另一种是使用语言、模型、算法等对内容的表达或描述。无论是哪种含义，表征本身都意味着一个观察者

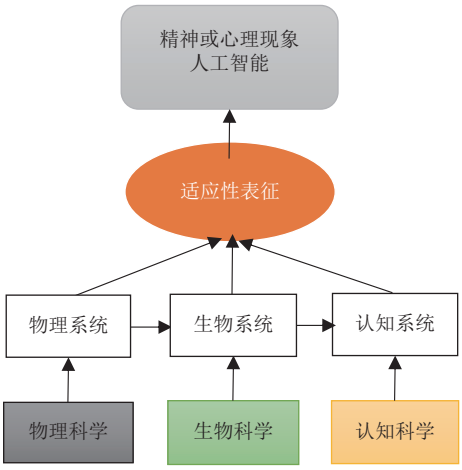

图 2　不同系统与相应学科的适应性表征关系

和解释者的存在，这就是人类。因此，广义的表征包括物理–生物层次的呈现和语言–符号层次的表达。这三个系统相互作用产生了人工认知或智能系统（图2）。物理层次提供硬件，生物层次产生人机混合系统，认知层次产生AI。AI的表征完全是符号的，其具身化的路还很长，这是现象学特别是认知现象学要研究的课题，因为感觉–知觉现象学是属人身体的，认知现象学是属人智能的、具身的，AI则是无身的或离身的，要让AI具有身体性包括意识和情感功能，就必须将AI、认知科学、脑科学和认知哲学、现象学及语言学包括语形学（语法）、语义学（意义）和语用学（实践）结合起来。至于AI最终能否具身化并产生意识和情感，或者AI无须具身化也能产生类意识和类情感的属性，这是另一个颇具争议甚至存在悖论（如具身–通用、无意识–智能）的话题，这里不作讨论。

接下来将论证，适应性表征不仅可以解释物理、生物层次上的属性“呈现”，也可以解释意识的“涌现”和认知层次上的“表征”属性。可以说，适应性表征就是生命和意识包括心智、认知、自由意志发生的机制，是连接不同层次包括物质与意识之间的机制或环节。

（二）适应性表征作为同层次属性呈现和跨层次属性涌现的机制

由上述可知，笔者将宇宙事件分为物理的、生物的、意识或认知（自然认知和社会认知）和人工认知（AI）四个层次，用系统论的术语说就是四个系统，相应地就有物理适应性表征（物理属性呈现）、生物适应性表征（行为表现）、意识或认知适应性表征（意向内容、命题态度）、AI适应性表征（知识表征、问题解决），当然认知系统如人类具有社会性，社会性就表现出适应性表征行为，各自遵循着相应学科（物理、化学、生物和神经科学等）的规律。各个层次间也是通过适应性表征相连接并相互作用的。

然而问题在于，适应性表征如何在不同层次或系统间发挥作用呢？这是一个更为深刻且不好回答的问题，因为它涉及哲学、物理学、生物学、心理学、认知科学和AI等不同领域。从自繁殖或自复制功能来考察，适应性表征在不同层次的系统中有不同的复制子（Replicator），物理系统的复制子是比特（Bit），生物系统的复制子是基因（Gene），认知（意识）系统的复制子是模因（Meme[②]），智能系统的复制子是智能体（Agent）。这些复制子通过携带信息（信号）起到传递信息（内容）的作用。

在物理系统中，“比特”一词据说最早是由图基（J.W.Tukey）提出的，由信息论创始人香农（C.E.Shannon）自行选用的，用于测量信息的基本单位，即量纲，这样一来万物就可由比特衡量，万物也因此源于比特。根据信息论，信息是物理的，但它既不是能量也不是物质，而是以熵的形式存在，即信息熵。在热力学中，熵是系统混乱度的度量，具体说是物理系统微观状态（如原子状态）的不确定程度，即处于所有可能微观状态的概率；在信息论中，熵是信息的不确定程度，即由信源发出的所有信息中的一个概率。无论是热力学还是信息论，都认为信息表征秩序，因而是负熵。比如知识作为信息，一定是有序的，无序会导致语义模糊。因此，所谓信息就是不确定性或混乱度的减少，有序度的增加，如激光就是原子有序排列的结果。在量子力学中，量子比特是最小的非平凡量子系统，有两个可能值——1和0，即两个能彼此区分的状态，也称为正交态，但同时也是量子叠加态，以不同的概率线性地组合，表现出确定性和不确定性这种混合的适应性表征状态。

在生物系统中，生物体的基因封装信息，允许信息的读取和转录，生命就是通过基因信息网络扩散的，一个生物体就是一个信息处理器，信息作为记忆不仅储存在大脑中，也储存在细胞里。根据遗传学，DNA是细胞层次的信息处理器，一种处理大量比特信息的编码。因此，DNA就是生物比特单位。生物体中的每个细胞都是生物信息网络中的一个节点，一刻不停地传递和接收信息，不停地编码和解码。在这个意义上，生物进化就是生物体与其环境持续不断进行信息交互的适应性表征过程。这意味着，我们的基因给予我们感觉和心智，感觉和心智在基因被选择的环境中具有了适应性，而表征的形成（意义）则与文化进化有关。因为“几百年来，基因从身体传送到身体，通过自然选择，使生物具有适应性。但在人类出现后，文化单位被从心智传输到心智，通过选择，使文化具有适应性”，而且“文化进化已经接替了生物进化。”[15]209这种文化进化作为比特单位被称为模因。

在认知系统中，根据自然主义，其认知能力是通过自然选择而与客观世界中那些对生物体特别

重要的方面达成协调一致的，在不断进化过程中，生物体的认知能力适应了外部世界的稳定结构，外部世界参与塑造了生物体的认知感官系统，所以我们的认知器官时时刻刻都处于与外部世界的关联中。正是这种关联性让道金斯在《自私的基因》中通过与生物基因的类比，提出文化基因的概念。这是一种不同于生物学基因的新的复制子，用于衡量一种文化传播单位或模仿单位的概念。这样看来，一个想法就如同一个基因，可以作为复制子在一个群里传播。在传播和自复制的意义上，文化不外乎是人类基因的一种扩展，或者说文化进化的能力植根于生物学的基因之中，受生物基因的制约。所以，模因也是基因的表现型效应，就像鸟会筑巢，蜘蛛会织网一样，人会制造人工制品包括非物质的文化。一种模因一旦形成，就会自复制，如流行歌曲的传播，也就是说，模因通过将自己储存到人类的记忆中保留下来，然后通过人类的行为繁殖。因此可以说，是基因与模因的复杂互动结合造就了人类的心智、理性甚至自由意志。如果说是基因塑造了人的身体，那么是模因塑造了人的心智和智能，二者的互动让人类更加智慧和聪明。从基因到模因的适应性表征，反映了人类的进步和知识的增长。但是，“我们必须记住，一旦模因引起了我们的注意，它就会设法繁殖，不管对人类是否有好处。……模因无论是技术制品还是抽象概念，都像基因一样指示我们行动。我们大部分的精力都用于选择后者繁殖它们。”[16]174现在的互联网包括新的通讯技术会不断制造出复制子，如病毒、APP，这些复制子似乎懂得利用超自然的信息传播信息，似乎是有智能的，事实上它们都是人造的“好像智能”。

对于这种人造智能（这里强调“人造”是因为这类智能是基于人类智能的，不是自然界本来就有的），其复制子是“智能体”，它们通过交换信息达到传播信息的目的，如计算机和AI系统，其最显著特征是通过操作符号来解决问题。不同的智能体可以组合成更大、功能更强的智能组，就像明斯基在《心智社会》中所描述的那样，心智社会是积木的世界，一个智能体就是其中的一块积木，组合起来形成整体的心智社会。积木的组合就像基因的自复制一样，思维、心智、自我、记忆等精神现象都是通过积木式组合形成的。AI的智能行为很大程度上就是通过这种适应性组合实现的。

质言之，上述四类复制子都是内部机制的施动者，如化学中的试剂，生物学中的DNA，AI中的agent。这些复制子就像生物卵一样，形成了一个自然–生命–意识–心智–智能连续链，不同系统之间是通过适应性表征无缝连接的，最终形成一个有机统一体。接下来的部分将通过自然系统（世界）和语言系统的比较来进一步说明适应性表征系统不同层次间的交互机制。

三、世界与语言系统的交互机制

笔者将不同层次的适应性表征与不同的语言层次作比较（图3），一方面是要说明二者整体属性的不可还原性和组成上的可还原性；另一方面要说明语义信息（意义）在跨层次的还原中如何消失了。

众所周知，我们生活于自然世界，我们会讲母语（自然语言）。自然世界创造了我们，我们创造了语言，然后我们又使用语言来描述自然世界和表达我们自己。自然世界让我们有了立足之地，语言让我们活得更精彩。也就是说，世界与语言的相互作用使我们能够认识世界并改造世界，同时也使我们认识和改造我们自己。语言的使用不仅发展了智力，也产生了海量的知识。我们之所以能将世界和语言区分开来，

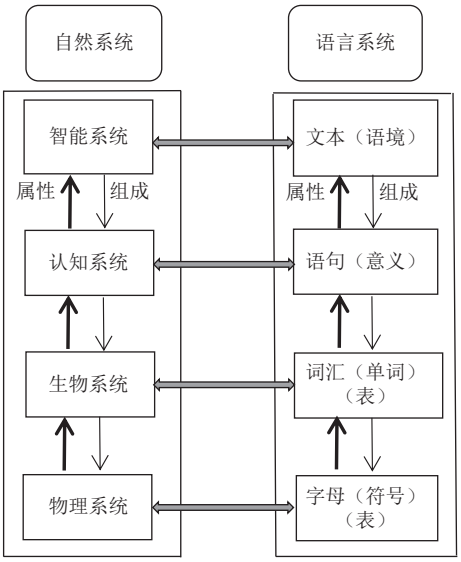

图 3　自然系统与语言系统的交互机制

（粗↑表示系统属性的不可还原性，细↓表示系统组成的可还原性）

是因为作为系统的世界和语言是有边界的。

从系统的角度看来，世界和语言是不同的系统，一个是自然系统，一个是人造系统。系统是有边界的，边界之外是其环境。自然系统的边界是物理实体、自然物，一个自然类就是一个系统，如动物、植物，明显区别于其环境。大脑是自然认知系统的边界，智能体是AI的边界。语言系统的边界在不同层次有不同形式，如字母表、词汇表、文本和语句。边界作为条件是受某个更高级法则支配的，比如力学的边界条件受机械工作原理的控制，概念的意义受其所在语句的约束。这意味着一个行为体（机器或生物体）、合语法的语句和行为事件都是受双重控制的综合体。根据波兰尼的看法，这种综合体是层级结构，其中包括两种法则，高级的和低级的，比如下棋的规则是低级法则，棋手的策略是高级法则。高级法则可以控制低级法则但不能用低级法则来说明，反过来，低级法则制约高级法则，低级法则遭到破坏，高级法则将不复存在，比如没有下棋的规则，就不会有下棋的策略。同样，没有生物体（生命）的存在，就不会有高级的心智的存在，这是生命–心智连续性论题。所以说，无论高级法则是否运行，低级法则都继续起作用，而且低级法则能终止或破坏高级法则所控制的综合体。

综合体的不同层次间是如何连接并持续运行呢？相对于综合体，不同层次是其子系统，比如我们人是由物理的、化学的、生物的、认知的和社会的不同子系统组成的一个连续综合体，但不同层次之间有各自的结构和运行规律，如身体结构和语言交流是不同的系统，其中的生物学法则不同于物理和化学法则，心智科学法则又不同于生物学法则。这意味着，基于生命的高级心智不能用低级的物理化学规律来说明。因为生物体存在于一个层级体系中，每个层级都有自己的结构法则和机体法则，而且这两种法则在生物体中相互交织。若将这两种法则用于心身关系，心智的高级法则依赖于其对有关生理学的低级法则的控制，由此波兰尼得出三个结论[17]202：（1）任何生理学上的观察都不能让我们理解心智的运作，因为从生理学对神经结构和机制过程的观察都是无生命的；（2）心智的运作永远不会干扰生理学的法则，也不会干扰其所依赖的更低级的物理和化学法则；（3）由于心智的运作依赖于更低级的身体法则的运作，所以身体上的不利变化会扰乱心智的运作，有利变化则会为心智提供新机会。根据这种看法，生命和心智现象不能还原为低级的物理–化学和生物学现象，也不会影响它们，但依赖于它们。这与查默斯的自然主义二元论一脉相承。

关于心智（心灵），贝特森提出六个标准，认为所有这些标准结合起来就可以解决心身问题，而且思想、进化、生态、生命、学习之类的现象只会发生于符合这些标准的系统中。这些标准包括[18]105：（1）心灵是相互作用的部分或组件的集合。（2）心灵各部分间的相互作用是由差异触发的。差异是一种不占据空间或时间的非实体现象；差异与负熵和熵有关，与能量无关。（3）心灵过程需要并行的能量。（4）心灵过程需要循环性（或复杂的）决定链。（5）在心灵过程中，差异的结果应被视为先前差异的转化（即编码事件）。转化的规则必须相对稳定（即要比内容更稳定），但规则本身也会经历转化。（6）对转化过程的描述和分类揭示出内在于现象的逻辑类型之层级结构。在贝特森的心智标准中，第一是整体性，第二是差异性，第三是动态性，第四是循环性，第五是转化性，第六是层级性。这些标准事实上就是生物系统和认知系统的特征，复杂性自组织理论业已揭示了其运作机制。

从意义生成来看，对于自然系统，属性或性质是物理实体的意义，如水银柱上升是温度计的意义，生命是生物体的意义，心智是身体的意义，智能是AI的意义。对于语言系统，单词是字母的意义，语句是单词的意义，文本是语句的意义。自然系统产生的是自然信息（具有客观性），语言系统产生的是人为信息（具有主观性），在解读的意义上，所有信息都是通过符号载体实现的，所有意义都是通过不同类别的适应性表征产生的。笔者发现，在这两个系统中，紧相连的层次，上层是下层的意义，或者说，下层涌现出上层的意义，比如认知系统和智能系统，智能是由认知系统给出的，即人创造了智能，而不是机器，机器只是展示智能的物理装置。一个系统的不同层次是连续的，这一连续过程存在意义整合，整合也是通过适应性表征实现的。用波兰尼的说法，意义的形成与我们的身体密

切相关，或者是通过整合我们身体内部的表征，或者是通过整合身体外部的信息，所有从外部知道的意义都源自我们看待自己身体的方式。正是通过我们的身体内化了外部事物，我们才能使外部事物成为我们所关注的目标。在这个意义上，无身的人工智能无论多么智能，都不会有意识（感受性）。

这种源于身体的意识，根据波兰尼的看法有两种：一是焦点觉知，二是附带觉知。焦点觉知是集中注意力的认知，附带觉知是在焦点觉知中无意关联其他方面的认知。两种觉知的融合会产生之前没有呈现的属性。比如我们观察一幅人物画，聚焦对象（如某人）和附带地关注（其他人）构成了整体画面。这其实是一个适应性表征过程，内在于身体的表征是我们察觉不到的（隐表征），如记忆，因为它们是整合的，外在于身体的表征是可察觉的（显表征），如身体姿态、语言表达，因为它们是演绎的，所以适应性表征是逻辑演绎与非逻辑整合的统一体。

就整体意义而言，生命、意识从组成上还原到物理-化学层次（基本粒子、原子、分子）时，生命和意识的整体属性就消失了，如生物体的死亡和尸体分解。词语被还原到字母层次时，其意义也就消失了，如word分解为w, o, r, d，汉字分解为笔画。因此，组成上可以还原的综合体，最终沦为无意义的组件或片段，比如汽车这种机械装置，当拆分为组件时，作为能发动能运动的物理装置的意义就不存在了。所以“机制，无论是人工机制或是形态发生的机制，都是支配着无生命自然界的规律的边界条件，它们自身不能还原为那些规律。作为遗传密码起作用的DNA，它所包含的有机碱基的排列模式是一个不能被还原为物理和化学规律的边界条件。就人而言，关于生命的更深层次的控制法则可以被描述为一个关于边界条件的层次系统，这些边界条件一直延伸到关于意识和责任”[17]220。总之，任何系统都是有边界的，超越边界就不成其自身了，其存在的意义也就消失了。所以，离身的意识就不是意识了，意识是具身的，人工系统的智能只是操作符号意义上的，原则上不需要意识。

四、不同自组织系统的适应性表征

这一部分重点论证不同自组织系统的适应性表征功能，旨在进一步说明AI系统是适应性表征系统，可用适应性表征方法来解释，使其成为可解释、可信赖的AI，有无意识并不重要。

（一）作为性能呈现的物理适应性表征

从系统及其演化的角度看，系统科学的各种理论——一般系统论、控制论、信息论、耗散结构论、协同学、超循环论、混沌学等[19]——所揭示的物理系统的演化规律，实质上就是适应性表征的具体机制，从而导致系统从无序走向有序（表征过程）。因此，系统科学的不同理论揭示了适应性表征系统的演化规律和运行机制。

第一，一般系统论为适应性表征提供了物理系统的整体自组织特性。根据一般系统论，系统是由若干要素以一定结构形式联结构成的具有某种功能的有机整体，是一种自组织结构。整体性、关联性、层级结构性、动态平衡性、时序性等是所有自组织系统的共同基本特征，适应性表征行为就是系统涌现出的一种整体功能。这与系统论强调调整系统结构，辖制各要素的关系，使系统达到优化目标的旨趣相一致。

第二，控制论为适应性表征提供了控制-反馈的内在运行机制。根据控制论，在控制系统运行的过程中，施控系统根据被控系统运行状态的变化，将其输出结果的一部分用以调整被控系统的输入，以改变被控系统的输出状态，并将被控系统的运行引向给定目标。这种控制方法就是反馈控制。因此，控制论强调系统的行为能力和系统的目的性，这与适应性表征的目标导向性是一致的，因为控制论主张任何系统要保持或达到一定目标，就必须采取适当的行为，如输入与输出、控制与反馈、生命与心智。

第三，信息论为适应性表征提供了承载内容（信息）的负熵依据。根据信息论，信息是“不定性的消除”或“负熵”，也就是使系统趋向有序化，而适应性表征就是趋向有序的，就是要减少或消除无序状态。这与信息论主张的信息是系统的组织程度、有序程度的观点相一致。这样看来，信息是

我们要认知与表征的内容，认知是一种信息处理过程，而表征则是信息的再现，信息因此成为构成物理世界的一种基本存在形式。进一步说，适应性表征的过程就是获得信息、加工信息、呈现信息的负熵过程。

第四，耗散结构论为适应性表征提供了非平衡演化的动力。根据耗散结构理论，所有演化系统都是耗散结构——一种从外界吸收能量的系统。耗散结构是宏观有序结构的生成，而这样一种有序结构的生成是突变的结果，需要一定的阈值。突变实质上是一种相变，在相变过程中，由于系统中分子间的相互作用而导致其原先平衡均匀的状态失去稳定性而发生。因此，远离平衡态是有序生成之源，平衡态反而是一种混乱状态。如果将适应性表征系统看作一种耗散结构，那么它一定是远离平衡态的，因为适应地表征就是一种目标导向的动态演化和稳定呈现的过程。

第五，协同学为适应性表征的发生提供了内部协调生成机制。根据协同学，有序结构的出现并不是非要远离平衡态这个条件，如超导体和电磁体这种有序结构，可以在热力学平衡下从无序状态产生，而像激光发射这种远离平衡态的系统与耗散结构意义的平衡态系统，在形成系统的有序结构的机理方面又是很相似的。因此，关键在于系统内部各子系统之间能否“协同”。这意味着系统中各个子系统的运动状态是由子系统的独立运动和子系统之间关联引起的协同运动共同决定的，协同运动必然导致系统形成一种能反映系统有序程度的宏观“序参量”，如法律和价值体系。“序参量”一旦出现，就主宰系统进入有序化过程，它通过信息反馈支配子系统的行为，使整个系统走向有序结构，而且不同序参量的合作会形成一种宏观结构，其竞争将导致只有一个模式的存在。这种合作和竞争决定着系统从无序到有序的演化过程。适应性表征无疑就是系统的一种序参量，它是演化系统进入有序状态的标志。人工智能中多智能体的协同合作何尝不是这样。

第六，混沌学为非线性随机系统的适应性表征提供了初始条件敏感和自相似原理。根据混沌学，系统对初值是敏感的，即某些确定系统的初值稍有变化，经过一段时间后，各自的差别就明显表现出来。这就是著名的“蝴蝶效应”，它揭示出系统的演化与初始条件的极微小变化密切相关，忽略次要的因素或条件就会造成结果的巨大差异。在此情况下，以实验观察系统的运动是不可重复、不可预测的，表现出“随机性”。适应性表征系统，无论是物理的、生物的还是认知的，严格说都是非线性随机动态系统。比如神经元的微小“涨落”和协同，自相似和自复制，可能导致了意识的出现，从而产生认知行为。在这个意义上，大脑（人工脑）就是一个动力学系统，从动力学探讨大脑（人工脑）的工作机制显然是一个可行进路。

物理适应性表征的一个典型例子是荧光显微图像的适应性粒子表征(Adaptive Particle Representation, APR)[20]。众所周知，现代显微镜创造了一种数据流，每秒产生千兆字节的数据。存储和处理这些数据是一个严重的瓶颈，不能完全通过数据压缩来缓解，这是因为图像被处理为像素网格。APR方法使用含噪音的三维图像，在保持图像质量的同时，适应性地表征了图像的内容，在一系列图像处理任务中实现了最佳效果，提供了一种简单有效的荧光显微图像的内容感知表征。也就是说，APR以基于图像内容定位的粒子替换像素来解决这一问题，不仅克服了存储瓶颈，也克服了内存和处理瓶颈。相比而言，荧光显微图像像素表征中的信息与数据比要低得多，并且受图像的空间和时间分辨率而不是其内容的控制。因此，一种理想的荧光显微图像表征形式与人类视觉系统有着共同的适应性和局部增益可控性。

（二）作为进化计算的基因适应性表征

相比于物理系统，生物系统更具适应性。进化生物学已经揭示了这一点，但对生物系统的分子演化机制的详细说明，艾根的“超循环论”有其独特优势。根据超循环论，生命信息的来源是一个采取超循环形式的分子自组织过程。这种循环现象有三个不同的层次：（1）转化反应循环，整体上是个自我再生过程；（2）催化反应循环，整体上是个自我复制过程；（3）超循环(Hyper-Cycle)是指催化循环在功能上循环耦合联系起来的循环，即催化超循环，其共同特征是：不仅能自我再生、自我复制，还能自我选择、自我优化，从而向更高的有序状态进化。这是典型的大分子系统的适应性表征行

为。认知系统的适应性表征也具有这些特征，可看作是一种超循环系统。

超循环论和生物学业已揭示，在基因（大分子）层次，基因的正确表达在生物体的功能中起着至关重要的作用。这个问题涉及改进基因表达程序的进化性、参数控制和并行化的适应性表征[21]，其中的基因表达编程（GEP）是一种进化线性染色体编码非线性（树状）结构的遗传算法[22]。在原始GEP算法中，基因组大小是问题特有的，是通过反复试验来确定的，其中引入突变、变换和重组算子是为了改进GEP，并产生一个具有异质结构的染色体群体，这是原始GEP算法所不支持的。这种新方法允许在通常不相容的个体间进行杂交，在一个群体中进行物种化，增加表征的可进化性，增强并行的GEP。

GEP算法的实质是进化计算问题，它是利用生物机制启发的过程来获得一个给定问题的解决方案。将进化计算算法应用于问题始于定义潜在的解决方案如何被表征，这就是“问题表征”。“问题表征”是由用于生成解决方案的输入数据类型（终端集）、所需的数目和输出类型，以及用于将输入转换为输出值的操作（函数集）来定义的。将进化计算算法应用于特定问题的一个重要步骤是指定定义问题表征和控制算法的参数。找到合适的参数值才能产生令人满意的结果，这通常需要开发出启示法或专家知识，这个寻找合适参数的过程就是适应性表征行为。在进化计算算法中，候选解群或个体的概念，被用来表征一个特定问题的可能解的集合。因此，用于表征一个解决方案的编码或基因组，取决于进化计算方法。进化计算可以像二进制代码一样简单，也可以像一种完整的编程语言一样复杂，能够实现灵活的基因组表达，赋予适应性特性，增加群体内的多样性，并增强算法的并行化。

显然，GEP进化算法的特征体现为可进化性、杂交与物种形成、分布式进化、参数调整和适应性。进化性是说，问题表征的结构在运行过程中不会改变，因为它仅限于顶部结构域长度和基因数目的初始值。这限制了算法缩小探索的范围，并降低了算法在群体中产生有意义的变化或范式转变的能力。杂交与物种形成的意思是，在GEP中，遗传操作和转化仅限于结构相同的基因组，防止不同物种或不同结构的基因组在种群内进化和竞争。分布式进化是说，并行化受到不同种群无法交互的限制，从而减缓了搜索空间的探索。参数调整和适应性是指，GEP算法本身缺乏适应性机制，因此它需要额外的时间和资源来系统地评估不同的控制参数集，并使算法受到算子偏差的影响。总之，进化计算对基因表达编程算法进行了改进，使其能够灵活地表达基因组，特别是增强了处理规范GEP中的可进化性、杂交和形态化、参数控制和并行化等问题的能力。

例如，一种DNA微阵列技术的最新进展——滑动窗−随机林（Sliding Window-Random Forest, SW-RF）方法，为监测数千个基因的表达水平提供了条件[23]。根据这种方法，研究人员将注意力集中在识别局部DNA元素，或推断基因的表达与核苷酸序列之间的关系上，提出了一种用于监督学习的新型数据适应性表征方法，以预测与生物序列相关的反应，比如机器学习中的分类序列通常被映射到学习任务的低维表征，以避免高维问题。SW-RF是一种基于特征的方法，它需要两个主要步骤来学习分类序列的表征：一是每个序列都用恒定长度的重叠子序列表征；二是在这个表征上基于树的学习者得到训练，以获得类似于表征的单词包，即每个序列在树的终端节点上的子序列的频率。在表征学习完成后，任何分类器可在习得的表征上得到训练。或者说，一套逻辑回归是在习得的表征上得到训练，旨在识别分类任务的重要模式。实验表明，该方法对合成数据和DNA启动子序列数据的准确性都有明显的提高。

（三）作为意向发生的神经网适应性表征

如果说化学分子与生物大分子层次具有适应性表征功能，那么人脑就是一种适应性预测加工系统，因为它能够对不断变化的环境做出迅速反应。我们的学习是通过修正神经元之间的联结的强度进行的。生理学的可塑性原则表明：“皮层神经元所创造的有关世界的表征并非固定不变的，而是不稳定的。在人的一生中，根据新经验、新的自我模式、外部世界的新刺激以及新同化工具等的不同，这一表征会不断调整自己。”[24]195相似地，人工神经网络的权值，也必须改变以呈现出相同的适

应性，因为它是由大量处理单元互联组成的非线性、自适应信息处理系统。比如，在监督学习的人工神经网络中，监督学习通过比较网络的表现与所析取的响应，相应地修改系统的权值，从而自发地表现出适应性。

在脑科学领域，脑科学家利用自组织神经网络试图通过竞争学习来保持输入空间的拓扑结构，这一能力被用来表征物体及其运动[25]。具体说，脑科学家使用一种自组织网络，即生长的神经元来表征沿图像序列的对象中的形变。作为适应性处理的结果，对象由拓扑表征图表示，该拓扑表征图构成了对其形状的诱导性剖分。这些映射能够在不重构学习过程的情况下适应对象拓扑的变化。这就是神经网络的适应性表征，可通过基于区间树的柱状图适应性表征来实现。也就是说，通过竞争性学习，自组织神经网络对神经元的参考向量以及它们之间的互连网络进行适配，以获得试图保持输入空间的拓扑映射。而且，即使输入新的模式，自组织神经网络也能够进行连续的重新适配处理，而不需要重置学习。这些能力已经被用来表征物体及其运动，通过增长的神经气体(the Growing Neural Gas, GNG)，自组织神经网络的学习过程比其他自组织模型更灵活。这项研究表明，GNG算法被用来表征二维物体的形变，得到一个拓扑表征图，可用于表征、分类或跟踪多个任务目标。当物体的拓扑形变很小且在图像序列中的连续框架间渐进时，可以利用先前的映射信息来放置神经元，而无须重置学习过程。利用GNG的这一特性可以实现高速表征过程，例如一种称为柱状图区间树的新数据结构可用来表征在给定的灰度范围（Gray-Level Range）内灰度等级的分布情况[26]，允许用户根据每个应用程序的需要适应地细化间隔。

对于信号表征和分类，可通过神经网络适应性生成微波模板，用于信号表征和分类[27]。这种超微波允许微波的形状适应特定的问题，该问题超越了固定形状微波的适配参数。通过对一维信号进行仿真，概念可扩展到图像，这等于将概念应用于音素和说话人的识别。微波经常被应用于表征，但很少被分类。研究证明了微波是如何使用不同的神经网络结构和最适合的能量函数，根据任一项任务适应性地计算的。因此，超微波的新概念允许对特定问题适应性地计算微波形状，而不是仅适应性地计算固定形状微波的参数。适应性微波概念在一维信号上得到了验证，这些概念不仅可用于音素和说话人的识别，而且也适用于图像识别，特别是使用超微波的膨胀来处理输入尺度变化的想法适用于一维信号和图像。总之，适应性生成最优微波特征集的思想，对于信号和图像都是一种强有力的表征方法。

在人工神经网络中，适应性机制的研究主要集中于学习运动任务期间的动力学适应性表征，比如中枢神经系统如何在不同的动力学条件下学习控制运动以及如何表征这种习得行为[28]。在外部施加力的情况下从机械环境中执行运动任务，这个环境是由机器人机械手产生的力场，被试在握住机器人的末端执行器时完成动作。由于力场显著改变了任务的动力学，与自由空间中的运动相比，力场中被试的初始运动被严重扭曲了。然而，在实践中力场中的手轨迹与在自由空间中观测到的轨迹非常相似。这表明，对于执行运动，有一个独立于动力学条件的运动学计划；在变化的机械环境中，性能的恢复就是运动适应性。实际上，适应性不是通过查找表组合的，相反，被试通过计算元件的组合对力场建模，所述计算元件的输出在整个运动状态空间上被广泛地调整。这些元素形成了一个模型，该模型在一个类似于关节和肌肉而非端点力的坐标系中外推到训练区域之外。这种几何性质表明，适应性过程的元素用传感器和执行器的固有坐标系来表征运动任务的动力学，可说明中枢神经系统是如何在不同的动力学条件下学会控制运动的，以及这种习得行为是如何被表征的。

（四）作为机器学习的适应性表征

如前所述，人工认知系统是基于物理系统和生物系统的，其具体形态是计算机、AI装置和机器人，其中机器学习是适应性表征的主要方面。那么，在机器学习中，智能体如何在语境中进行适应性表征呢？语境邦迪设置（the Contextual Bandit Setting）可说明这个问题[29]。在标准语境邦迪设置中，学习算法常常面临仅仅关于它过去采取的动作给出反馈的问题[30]。规范示例是互联网搜索的问题，其中目标是在给定一些可观测的语境(查询、地理位置)的情况下，学习用户感兴趣的那些结果。标

准语境邦迪算法特别关注在线设置，其中学习者从N个策略的一大类中反复选择，但只能根据自己的选择得到部分反馈，每个策略将语境X映射到K个动作中的一个，每个动作都在每个回合上具有潜在的不同奖励。学习者的效用是用其后悔来衡量的，被定义为最佳策略的累积奖励与学习者的奖励之间的差异。因此，在这种设置中，具有表征学习和未标记历史的语境邦迪（Contextual Bandit with Representation Learning and Unlabeled History, CBRH）在进行在线决策之前，学习者可以获得一组未标记的语境，允许学习语境表征，而不是使用原始语境。

在在线阶段，学习者根据每个语境适应地选择嵌入的目标，并根据迄今观察到的语境进行更新。有研究[31]针对CBRH问题提出了基于在线聚类（Clustering）和离线聚类的两种具体算法，将在线嵌入选择和学习与语境的汤普森抽样邦迪结合起来。在几种类型的非平稳环境中对算法进行评估，并将这些算法与标准语境邦迪以及通用(单个)嵌入进行比较。这是一种在AI中如何语境化智能体的方法。一旦人工主体能够自己主动适应语境及其变化，它就能够像人那样去思维和行动，这是一种自语境化的能力。适应性表征特别是认知层次上的，就是一种语境化能力。对于人工主体而言，初始语境是需要人类设计者为其设置的，可称为它的历史语境，在此基础上，人工主体可通过预训练根据目标任务使用和调整历史语境，以便适应性地解决问题。

对于人工主体的适应性表征学习问题，机器学习中引入了一种用于目标检测的适应性表征学习范式[32]，即一种用于对象检测的无监督域适应性方法，目的是缓解像素层适应性的不完美转译问题（the Imperfect Translation Problem, ITP）和同时进行特征层适应性的源偏倚判别性问题（the Source-Biased Discriminativity Problem, SBDP）。这个过程分为两个阶段：域多样化(Domain Diversification, DD)和多域不变表征学习(Multidomain-Invariant Representation Learning, MRL)。在DD阶段，通过从源域（历史语境）生成各种不同的移位域（新语境），从而使标记数据的分布多样化。在MRL阶段，将对抗式学习应用于多域识别器，以鼓励在域中区分特征。DD解决了源偏倚判别性问题，而MRL则减轻了不完全的图像转译问题。这就为学习范式构建了一个结构化的域适应性框架，引入一种可实施的DD实践方法。

对于机器人来说，它们面对的是多样的环境，必须适应这种多样性才能实时地行动。例如，视觉表征中的低延迟应用动态三维网格的适应性表征[33]。AI中高变形三维模型（如动态三维网格）使三维可视化表征变得越来越流行，因为它们能够真实地表征真实世界的物体−人的运动，为新的和更先进的沉浸式虚拟、增强和混合现实体验铺平了道路。在信号的多尺度适应性表征方面，一种用于表征信号的多尺度、适应性、移位不变框和双框的新框架被称为AdaFrame（适应性框架），在推理时间的计算效率方面改进了基于字典学习的技术[34]。这种新技术从编码效率的角度对微波框等经典的多尺度基进行了改进，为低层级信号处理任务（如压缩和去噪音）和高级任务（如用于目标识别的特征提取）的基于词典学习技术提供了一种有吸引力的替代方法，包括与深卷积网络（deep convolutional networks）方法的关联，具有前微波和微波框架的多尺度和计算效率，在字典学习中具有适应性。研究表明，在适应性框架和适应性双框架之间，适应性双框架由于额外的易用性而更易使用，学习过程也更容易，特别是当系统是非常冗余的情况下，学习过程可以通过学习分解和重构两个阶段分别进行。

（五）作为强化学习的适应性表征

强化学习（reinforcement learning, RL）的主体也是智能体。早在1998年，萨顿（R.S.Sutton）就将强化学习分为三个阶段[35]：1985年前是试错学习阶段，强调使用积极探索的学习器，即智能体，开发出了利用标量回报信号指定智能体的目标的核心思想（即回报假设），但这些方法通常只学习策略，一般不能有效地处理延迟回报。1985 ~ 1998年阶段的强化学习特指价值函数方法（价值函数假设），它是强化学习的核心，因为几乎所有方法都集中于价值函数的逼近方面，旨在计算最优策略。从适应性表征的视角看，逼近最优解的过程就是适应性表征过程。按照萨顿的预测，强化学习的第三个阶段即未来阶段，可能会把重点放在对价值函数进行估计的结构研究上，并与心理学相结合来积极创

造表征世界的建构主义方法。事实上，此后的许多新进展都与实现价值函数逼近的新结构相关。

萨顿的预测发生在20世纪末，当时强化学习已经成为AI中机器学习的显领域[36]。现在看来，目前的深度强化学习（DRL）就是强化学习所预测的第三个阶段。DRL是深度学习与强化学习的有机结合，一种新兴的通用人工智能算法，可看作是从弱AI走向强AI的重要一步。谷歌（Google）是这一领域的推动者和领跑者，2015年10月谷歌开发的AlphaGo就是使用深度强化学习算法，2017年1月出现的Master，就是AlphaGo的升级版。2025年初推出的DeepSeek，强化学习也是其主要方法之一（其他方法包括混合专家架构、多头注意机制、蒸馏方法等）。可以说，DRL算法在围棋领域所向披靡，在语音识别领域也达到人类的水平，比如微软2016年下半年就宣布在语音识别领域取得重大突破。深度学习的层次目前已达到150多层，如微软开发的图像识别残差网络深度高达152层。这一成就要归功于“大数据、大模型和大计算”这三大技术，它们是深度强化学习的三大支柱，适应性表征贯穿始终。可以预计，未来的深度强化学习层次可达上千层，那时的AI将很可能成为与人类智慧相匹敌的人工主体（预计耗能巨大、成本很高）。

目前，强化学习的适应性表征已经得到充分研究。强化学习中的自治主体（an Autonomous Agent）寻求一种有效的控制策略来处理序列决策任务。与监督学习不同的是，主体从不看对或不对行为的例子，而只接收作为反馈的奖励信号。现有方法的一个限制是，它们通常需要人为地为解决方案设计一个表征，如神经网络的内部结构。由于糟糕的设计选择会导致严重的次优策略，能自动调整自己表征的主体有可能显著地提高性能。惠特森（S.Whitson）提出了两种自动发现高性能表征的新方法：第一种方法综合了时间差分方法、传统的强化学习方法和演化方法，可以学习广义最优化问题的表征。这种综合是通过在线进化计算和进化函数近似来完成的，前者的大多数强化学习问题的在线性质制定了进化方法，而后者的进化是对时差方法至关重要的价值函数近似器的表征。第二种方法称为适应性分块编码，它自动学习基于块编码的表征，这种编码形成了价值函数的分段常数近似。适应性分块编码从粗表征开始，并在学习过程中逐步细化它们，分析当前的策略和价值函数，从而推断出最佳的细化表征。

惠特森给出了强化学习的一个框架：智能体→行动→环境→状态（奖励）→智能体……的循环，其中智能体采取一系列的行动操作，每个行动都会产生一个奖励和一个新的状态，形成一个循环过程[37]8。惠特森还引入了一种设计输入表征的新方法，即一种用于找到一组足以描述人工主体当前状态的最小特征集的方法，它是为应对一个被称为特征选择问题的挑战而提出的。特征选择问题即“强化拓扑的神经进化”（Neuroevolution of Augmenting Topologies, NEAT）问题，该方法是对NEAT的扩展，一种进化神经网络的方法。

事实上，在强化学习中，寻找良好的状态适应性表征是一个具有挑战性的问题，如策略树算法[38]。在机器学习中，直接表征策略的策略梯度算法通常需要较少的参数来学习好的策略，但它们通常采用可能不足以用于复杂域的固定参数表征。而策略树算法可以在一个基本策略的不同实例化上，以决策树的形式学习策略的适应性表征。策略梯度既用于优化参数，也可通过选择使策略预期返回最大局部增长的分块来使策略树生长。实验表明，策略树算法能够选择真正有用的分割，并对常用的线性吉布斯软件最大化（Gibbs Softmax）策略进行了显著改进。显然，策略树算法与参数策略梯度方法具有相同的收敛性，但在改进策略时可以采用其表征形式。这些事实表明，期望奖励的梯度是在强化学习中寻找表征的有用信号。

（六）作为大数据整合的适应性表征

在现代社会，由于信息技术的普遍使用，用户个人环境中大量可用的数字资源带来了许多问题。用户的目标是在大量的异构数字资源中进行选择，这是在活动中使用的最佳资源，如数字资源搜索结果在个人学习环境中的适应性表征[39]。传统上，用户个人环境中大量可用的数字资源是一个耗时的过程，需要用户花费大量的精力来优化参数的选择。这常常使不可利用的数字资源仅在存储库或数字图书馆中可用。基于使用语境和用户配置文件的适应性视觉表征方法，允许用户解释资源

搜索结果，这是在一种个人互动智能语境中完成的。

在大数据应用方面，计算云是大型分布式数据的数据适应性表征[40]。许多地理分布、互联站点、政府部门都有大量的本地数据，它们有兴趣协作学习这些数据背后的低维几何结构。这是一种新的字典学习算法（也称云K-SVD的分布式算法）被用于学习所感兴趣的分布式数据的子空间联合结构。这种云K-SVD实现了协作数据适应性表征的目标，而无须在不同站点之间传输各个数据样本，其适应环境的效率是比较高的。大量模拟实验表明，K-SVD算法可以促进字典的协作学习，最接近分布在各个地理区域的海量数据，具有有效性和收敛性。因此，这种算法可用于现实世界和大数据问题，例如，在各种设置下的数据适应性紧框架方法是一种强大的稀疏近似工具，一种能提供多尺度结构的数据适应性表征模型，具有与微波基础相似的缩放特性，但没有必要的自相似结构[41]。显然，适应性提供了更好的稀疏性，使用类贝索夫规范结构（Besov-like Norm Structure）既能诱导稀疏性，又有助于识别重要特征。数值实验证实，分配较低权重的恢复框架向量，对应于较大尺度和较低局部变化的图像元素，从而表明自然图像的加权稀疏性导致了自然尺度的分离。

在动态导航领域，智能体将每个感知到的障碍物看作一个局部敏感的、障碍物赋值函数，然后返回一个以引导和避障为基础的障碍表征[42]，这是智能体遇到障碍时做出的一种适应性应急反应，包括行为灵活性、计算效率和动态响应能力。通过对真实机器人的测试验证，这种新表征与以前使用的表征相比，在测试场景中的有效性、对各种场景和参数值范围的鲁棒性，以及计算效率等方面具有良好的效果。这种方法的参数值的鲁棒性对于不可预测或异质环境中的导航特别重要，提高了环境的一个组件子区域中有效的参数设置在其他环境中也是有效的可能性。更高的效率和更高的计算效率表明，动态切线表征（Dynamic Tangent Representations）和导航可能特别适合于计算能力有限的应用，如微型机器人。

在电影虚拟现实中，全向视频的内容适应性表征是一个典型例子[43]。电影虚拟现实通过呈现真实世界场景的全向视频提供了一种沉浸式的视觉体验。一个关键的挑战是开发全向视频的有效表征，以便在资源约束下最大限度地提高编码效率，特别是采样数和比特率，并将表征的选择表述为一个多维、多选择背包问题，证明所得到的表征很好地适应了变化的内容。

另外，在新闻报道领域，如何追踪突发新闻就是适应性表征行为[44]，推特（Twitter）通常是最新的消息来源，用来记录和追踪突发事件。这是一项重要的文本分析任务，由于推特很短，随着时间的推移，标准文本相似性度量常常失败。一种对文本摘要的自动评估研究表明[45]，适应性相似机制最适合于跟踪推特上不断发展的故事。在推特数据上训练的跳跃图模型能够捕捉推特文本中的语义和句法相似性。创建在自录音再现装置（Tweets）中使用的所有术语的向量表征，使得人们可以将单词与账户说明和井号(#)标签（微博、Twitter中用来标注线索主题的标签）进行比较，从而减少对实体进行预处理和执行查询扩展以保持高回忆的需要。学习向量的组合性使人们可以将术语组合起来，得出单个自录音再现装置之间的相似性度量。在滑动窗口方法中使用新的数据对模型进行再训练，通过在每个时间窗口生成新的自录音再现装置和查询的术语表征，人们可以创建一种测量自录音再现装置相似性的自适应方法和有用的辅助工具。

五、进一步的讨论

将适应性表征概念引入认知或智能系统，一方面是想要科学地说明物质与意识或存在与思维这一哲学基本问题；另一方面是要解释人工系统产生智能的机制或原理。对于哲学基本问题的说明形成了种种哲学流派或立场，诸如唯物论、唯心论、二元论及其变种，构成了形形色色的观点与争论。笔者坚持科学唯物论，即承认意识源于物质又反作用于物质。然而，这种笼统的哲学主张不足以让唯心论者信服，因为意识也可产生物质，比如AI就是人类意识的产物，所有人类知识也是意识（心智）的产物。

我们的意识经验具有私人性，也就是知道自己体验了什么，但不知道他人拥有类似的经验。这是哲学上的“他心”问题。这个问题一般可以通过“移情”或“同理心”来推知，因为我们是同类。进一步的问题是，我们能够从科学知识推出意识经验（感受性）吗？心脑同一论认为是可以的，因为我们的“感觉”“知道”“意识到”等体验是由我们的神经系统创造的，这就是知觉和认知的具身性。换句话说，主体的经验（感受性）是神经系统产生的，比如“色彩”“味道”等经验感觉，均源于神经系统。问题是，神经系统如何产生了经验呢？也就是心灵哲学家查莫斯所谓的“意识难题”——大脑的客观物理过程如何产生或涌现出主观的意识经验。

我们如何区分客观的物理过程与主观的意识经验？或者说，物理过程如何转化为意识经验？从科学上我们可以考虑信息转换或信息加工在其中的作用。这样一来，大脑中产生经验的问题就可以转化为“脑中的电信号如何转化为意识经验”的问题。这就是认知科学或脑科学的研究路数。值得关注的是，李德毅的“认知物理学”构想很可能成为实现人工认知的物理方法。这种新方法论用物理学理论和方法解释心智和精神，用物理机器拓展认知和智能，提出物质、能量、结构和时间是人类认知和机器认知的基本要素，因为激活机器的钥匙是时钟、时序和递归，认知机器用负熵能量产生时序，依靠递归思维，突破了计算智能的局限，可发展记忆智能和具身智能。具体来说，“认知物理学试图用记忆驱动的经验模式、知识驱动的推理模式、联想驱动的创造模式以及假设驱动的发现模式来形式化人的不确定性认知，实现物质硬构体和思维软构体相互纠缠”[46]。这意味着，若用机器实现人的认知模式，就要求机器要有新的硬核——抽象、联想和交互，其结果是：通过抽象得到结构，通过联想产生类比，通过交互形成反馈，将已有人工智能范式整合，最终形成一种认知螺旋，从而创造出可交互、会学习、自成长的认知机器，并与人类认知交互而迭代发展。因此，意识经验可以科学地研究，并不是神秘不可测的。从科学研究角度看，就是要研究产生意识经验的电信号是如何产生的。大量的实证研究表明，电信号是由神经纤维传导的，具体而言，沿神经纤维传导的神经脉冲（电信号波动）创造了意识经验（感受性），如看到“红色”，感到“头疼”等。而神经脉冲（电信号波动）是由带电离子（钠离子和钾离子）在神经纤维中的传导生成的，这个传导过程就是“意识经验流”，而神经纤维中的带电粒子就是“意识经验的神经相关物”。不过，对我们来说，这些传导过程是微观的，只有通过高倍的显微镜才有可能观察到，但神经纤维中离子的传导过程我们体验不到。

从适应性表征来看，神经纤维的信号传递过程是分布式表征，即特定的心理功能由分布于大脑中的多个区域的神经活动来表征。这意味着，意识经验是整个大脑不同功能区的神经活动（激活或抑制）构成的网络形成的，单一功能区（如脑皮质）不足以说明意识现象。在描述的意义上，神经纤维的信号传递的分布式表征可由功能网络表征，以便可视化。这种可视化的网络在AI中是图表征，具体来说，功能网络由节点及其连线组成，节点表示功能处理单元，如神经元、脑区等，连线表示节点间的交互关系，而且连线的长短可表示节点间的强度。因此，AI的发展有助于破解物质产生意识难题，比如通过机器意识的研究，通过AI技术介入研究过程（如基因编码）。这不仅是哲学问题，也是科学和技术问题；不仅是哲学工作者的任务，也是科学研究者和AI专家的任务。

笔者之所以将适应性表征作为心智支架和解释框架，是因为发现它是所有自组织系统的共性，当然也是精神或心理现象产生的内在机制。换句话说，适应性表征有不同形态——自组织协同、刺激反应、目标导向、自我反思，分别对应于物理系统、生命系统、意识系统和认知系统（高级智能），表现出从低级到高级的智能发展过程。从哲学上审视，物理系统体现了自组织协同性，可视为“原我”（无意识）；生命系统表现出行为适应性、感受性和意向性，可视为“本我”，具有了初级目标意识（潜意识）；意识系统彰显了目标指涉性和反身性，可视为“自我”或“自我意识”，具有心理表征（图像意识）能力；认知系统（高级智能）表现出自我反思性（自我预知性）和抽象符号表征性（符号意识），AI是这种系统的逻辑衍生物。

这些不同层次的自组织系统从更大和更高的社会水平看，应该是连续的，它们之间的界限相对而言是模糊的，比如有意识的生物体，既是物理系统，也是生命、意识和认知系统。相比而言，AI系

统与生命、意识系统是截然分明的，但在智能层次（符号处理意义上）与意识系统（人类智能）就比较模糊了（人和机器都会处理符号）。因此，在智能层次，适应性表征将会发挥更大的作用（两种智能的区分见表1）。

表 1　自然智能与人工智能的比较

属性或类别	自然智能	人工智能
哲学立场	科学实在论 生物自然主义	认知主义 功能主义
科学基础	进化生物学 认知科学、脑科学	计算机科学、人工智能、机器人学
基本组成	碳基生物体	硅基物理装置
本质属性	心理意向性 意识感受性	功能倾向性 目标指向性
身心关系	生物具身性 心脑同一性	物理功能性 心身离散性
意向关系	目的意向性	对象关涉性
理解性	有学习能力 有理解力	机器学习 无理解力
因果性	因果性	相关性
表征关系	强适应性表征 （动物的适应性强于人，但在表征方面弱于人）	弱适应性表征 （具身AI或通用AI的目标是实现强适应性表征）

六、结语

综上，适应性表征是所有自组织系统演化的一种普遍性特征。可以说，凡是有AI介入的领域合作活动，都会存在适应性表征问题，比如强化学习的适应性转变、策略树和策略梯度的适应问题。一句话，智能体这类人工主体，我们人类要想让它们执行任何我们赋予它们的认知任务，适应性表征是其中必不可少的一个关键环节，因为这个概念和方法可以合理地解释从物理层次到生物层次再到认知层次的连贯一致的联系。因此，适应性表征不仅是阐明科学认知和人工认知的核心认识论概念，也是实现智能行为的重要科学方法论。新一代人工智能或通用人工智能要打破两种智能之间的壁垒，适应性表征是一个有用的认知支架，因为它不考虑两种智能的物理成分（碳基的还是硅基的），也不考虑有无生命和意识问题，只关注两种智能的共享属性，而共享属性恰恰是通用的。

注释

① https://en.wikipedia.org/wiki/Adaptive_representation, 2019-11-22.

② 这个词源于希腊语的“mimesis”，意思是模仿，《牛津英语词典》对它的释义是“一种通过非遗传的模仿途径进行传播的文化因素”。道金斯将希腊词根“mimeme”缩短为“meme”，以便与“gene”看起来像。在他看来，meme 既与memory（记忆）有关，也与法语词Même（同样的）有关，而且念起来也与cream（搅合）合韵。这就是他采用这个词的理由。

参考文献

[1] 钟义信. 机制主义人工智能理论——一种通用的人工智能理论[J]. 智能系统学报, 2018, 13(1): 2-18.

[2] 钟义信. 统一智能理论[M]. 北京: 科学出版社, 2023.

[3] 魏屹东. 适应性表征: 架构自然认知与人工认知的统一范畴[J]. 哲学研究, 2019(9): 114-124.

[4] 魏屹东. 认知的适应性表征: 机制、特征与功能[J]. 南京社会科学, 2022(5): 17-25.

[5] 陈小平. 大模型关联度预测的形式化和语义解释研究[J]. 智能系统学报, 2023, 18(4): 894-900.

[6] 魏屹东. 适应性表征: 意识生成的内在机制和解释框架[J]. 科学·经济·社会, 2023, 41(5): 32-49.

[7] 魏屹东. 人工智能的适应性知识表征与推理[J]. 上海师范大学学报(哲学社会科学版), 2019(1): 65-75.

[8] 魏屹东. 人工智能的适应性表征[J]. 上海师范大学学报(哲学社会科学版), 2018, 47(1): 28-39.

[9] HOLLAND J H. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence[M]. Ann Arbor, MI: University of Michigan Press, 1975.

[10] HEYLIGHEN F. Representation and change: a metarepresentational framework for the foundations of physical and cognitive science[M]. Belgium: Communication and Cognition, Ghent, 1990.

[11] HOLLAND J H. Adaptation in natural and artificial systems[M]. Cambridge, MA: MIT Press, 1992.

[12] 马文•明斯基. 心智社会[M]. 任楠, 译, 北京: 机械工业出版社, 2016.

[13] DELEUZE G, GUATTARI F. A thousand plateaus: capitalism and schizophrenia[M]. Minneapolis, MN: University of Minnesota Press, 1987.

[14] HOLLAND J H. Signals and boundaries: building blocks for complex adaptive systems[M]. Cambridge, MA: MIT Press, 2012.

[15] 史蒂芬·平克. 心智探奇: 人类心智的起源与进化[M]. 郝耀伟, 译. 杭州: 浙江人民出版社, 2016.

[16] 米哈里·契克森米哈赖. 自我的进化: 第三千年心理学[M]. 朱蓉蓉, 译. 北京: 世界图书出版公司, 2019.

[17] 迈克尔·波兰尼. 认知与存在: 迈克尔·波兰尼文集[M]. 李白鹤, 译. 南京: 南京大学出版社, 2017.

[18] 格雷戈里·贝特森. 心灵与自然: 应然的合一[M]. 钱旭鸯, 译. 北京: 北京师范大学出版社, 2019.

[19] 苗东升. 系统科学原理[M]. 北京: 中国人民大学出版社, 1990.

[20] CHEESEMAN B L, GÜNTHER U, GONCIARZ K, et al. Adaptive particle representation of fluorescence microscopy images[J]. Nature Communications, 2018, 9: 5160.

[21] BROWNE N P A, DOS SANTOS M V. Adaptive representations for improving evolvability, parameter control, and parallelization of gene expression programming[EB/OL]. (2010-05-06). http://dx.doi.org/10.1155/2010/409045.

[22] FERREIRA C. Gene expression programming: a new adaptive algorithm for solving problems[J]. Complex Systems, 2001, 13(2): 87-129.

[23] CAKIN H, GORGULU B, GOKCE M, et al. A data adaptive biological sequence representation for supervised learning[J]. Journal of Healthcare Informatics Research, 2018, 2(4): 448-471.

[24] 米格尔•尼科莱利斯. 脑机穿越: 脑机接口改变人类未来[M]. 黄珏苹, 郑悠然, 译. 杭州: 浙江人民出版社, 2015.

[25] GARCIA-RODRIGUEZ J, FLOREZ-REVUELTA F, GARCIA-CHAMIZO J M. Representation of objects topology deformations with growing neural gas[C]//International Work-Conference on Artificial Neural Networks: Computational and Ambient Intelligence. [S. l.]: Springer, 244-251.

[26] YANG X D. Adaptive representation of histogram using interval tree[C]//International Conference on Image Processing and Its Applications. Maastricht, Netherlands: IET, 1992: 490-493.

[27] SZU H H, KADAMBE S. Neural network adaptive wavelets for signal representation and classification[J]. Optical Engineering, 1992, 31(9): 1907-1916.

[28] SHADMEHR R, MUSSA-IVALDI F A. Adaptive representation of dynamics during learning of a motor task[J]. The Journal of Neuroscience, 1994, 14(5): 3208-3224.

[29] LIN B, CECCHI G, BOUNEFFOUF D, et al. Adaptive representation selection in contextual bandit[EB/OL]. (2018-02-03). https://doi.org/10.48550/arXiv.1802.00981.

[30] BEYGELZIMER A, LANGFORD J, LI L H, et al. Contextual bandit algorithms with supervised learning guarantees [EB/OL]. (2010-02-22). https://arxiv.org/abs/1002.4058v2.

[31] AGRAWAL S, GOYAL N. Thompson sampling for contextual bandits with linear payoffs[J]. PMLR, 2013, 28(3): 127-135.

[32] KIM T, JEONG M, KIM S, et al. Diversify and match: a domain adaptive representation learning paradigm for object detection[J]. Computer Aided Geometric Design, 2019, 73(1): 70-85.

[33] ARVANITIS G, LALOS A S, MOUSTAKAS K. Adaptive representation of dynamic 3D meshes for low-latency applications[EB/OL]. (2019-08-01). https://doi.org/10.1016/j.cagd.2019.07.005.

[34] TAI C, WEINAN E. Multiscale adaptive representation of signals: the basic framework[J]. Journal of Machine

Learning Research, 2016, 17(140): 1-38.

[35] SUTTON R S. Reinforcement learning: past, presnt and future[C]//Simulated Evolution and Learning. [S. l.]: Springer, 1998: 195-197.

[36] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M]. Cambridge, MA: MIT Press, 1998.

[37] WHITESON S. Adaptive representation for reinforcement learning[M]. Cham: Springer, 2010.

[38] DAS GUPTA U, TALVITIE E, BOWLING M. Policy tree: adaptive representation for policy gradient[C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2015.

[39] SAWADOGO D, SUIRE C, CHAMPAGNAT R, et al. Adaptive representation of digital resources search results in personal learning environment [EB/OL]. (2015-05-10). https://hal.archives-ouvertes.fr/hal-01211605.

[40] RAJA H, BAJWA W U. Cloud K-SVD: Computing data-adaptive representations in the cloud[C]//2013 51st Annual Allerton Conference on Communication, Control, and Computing. Monticello, IL, USA: IEEE. 2013: DOI:10.1109/Allerton.2013.6736701.

[41] DOBROSOTSKAYA J, GUO W H. A data adaptive biological sequence representation for supervised learning: data adaptive multi-scale representations for image analysis[C]//2008 IEEE International Conference on Information Reuse and Integration. Las Vegas, NV, USA: IEEE, 2008: DOI: 10.1109/IRI.2008.4583007.

[42] AARON E, MENDOZA J P, NICHOLS F. Adaptive obstacle representations for dynamical navigation[C]// Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference. [S. l.]: AAAI, 2012: 11042343.

[43] YU M, LAKSHMAN H, GIROD B. Content adaptive representations of omnidirectional videos for cinematic virtual reality[EB/OL]. (2015-10-30). http://dx.doi.org/10.1145/2814347.2814348.

[44] BRIGADIR I, GREENE D, CUNNINGHAM P. Adaptive representations for tracking breaking news on Twitter[EB/OL]. (2014-03-12). https://doi.org/10.48550/arXiv.1403.2923.

[45] LIN C Y. Rouge: a package for automatic evaluation of summaries[C]//Proceedings of the ACL-04 Workshop. Barcelona, Spain: Association for Computational Linguistics. 2004: 74-81.

[46] 李德毅. 认知物理学导引[J]. 智能系统学, 2024, 19(3): 493-493.

A Cognitive Theory of Adaptive Representation for Artificial Intelligence

WEI Yi-dong

Abstract　Exploring the generation of intelligence in terms of Cognition Philosophy is a major challenge. This requires a conceptual framework that accounts for cognitive mechanisms, which is “adaptive representation”. Adaptive representation, as intrinsic mechanisms and explanatory categories of self-organizing systems at different levels, forms the basis of artificial intelligence moving towards generality, explainability, and reliability. This theory of adaptive representation for the generation of intelligence includes assumptions, inferences and principles as well as interactive processes of different hierarchical structures, aiming to show that the generation of intelligence is the result of the interactive emergence of different hierarchical structures of a self-organized entity or system through adaptive representation. Under the adaptive representation perspective, physical systems exhibit self-reaction and self-presentation of properties, biological systems exhibit self-adaptation and self-propagation of life, cognitive systems exhibit self-learning and self-expression, and AI systems exhibit machine learning and self-replication, and these different modes of representation precisely illustrate that adaptive representation is universal to all self-organized systems, and that the universal of general intelligence is adaptive representation, which implies that different domains of artificial intelligence have adaptive representational properties or functions, and constructing an AI system is creating an adaptive representational system.

Key words　cognition; intelligence; self-organizing systems; human intelligence; artificial intelligence; adaptive representation

编　辑　蒋晓