认知空间：自然、人工与混合

Ricard Sol´e,[1,2,3,4,*] Luis F Seoane,[3] Jordi Pla-Mauri,[1,3] Michael Timothy Bennett,[5] Michael E. Hochberg,[6,4] 与 Michael Levin[7,8]

[1]复杂系统实验室，庞培法布拉大学，*Dr. Aiguader 88, 08003* 巴塞罗那。[2]加泰罗尼亚高等研究院（*ICREA*），*Pg. Llu´ıs Companys 23, 08010* 巴塞罗那。[3]进化生物学研究所，*CSIC-UPF*，*Pg. Mar´iti m de la Barceloneta 37, 08003* 巴塞罗那。[4]圣菲研究所，*1399 Hyde Park Road,* 圣菲，*NM 87501*，美国。[5]澳大利亚国立大学计算机学院，*ACT*，澳大利亚。[6]*ISEM*，蒙彼利埃大学，法国国家科学研究中心，*IRD 34095* 蒙彼利埃，法国[7]艾伦发现中心，塔夫茨大学，梅德福，*MA*，美国。[8]威斯生物启发工程研究所，哈佛大学，波士顿，*MA*，美国。

arXiv:2601.12837v1 [q-bio.NC] 19 Jan 2026

认知过程可在极其多样的自然、人工和混合系统中实现，但目前尚无统一框架可用于比较它们的形态、局限及尚未实现的可能性。在此，我们提出一种认知空间方法，以基于组织和信息维度的比较性表征，取代狭隘且依赖基底的定义。在这一框架中，认知被视为一种分级能力，即感知、处理信息并据此采取行动的能力，从而能够在共同的概念图景中分析细胞、大脑、人工智能体以及人类—人工智能集体等迥异的系统。我们引入并考察了三个认知空间——基础无神经认知空间、神经认知空间和人类—人工智能混合认知空间，并指出这些空间中的系统分布极不均衡：已实现的系统聚集成簇，其间隔着大片尚未占据的区域。我们认为，这些空白并非偶然，而是反映了进化偶然性、物理约束和设计限制。通过关注认知空间的结构，而非类别式定义，这种方法既阐明了现有认知系统的多样性，也凸显了混合认知这一探索新型复杂性的前沿方向，其复杂性有望超越生物进化所产生的形式。

关键词：进化形成的认知、基础认知、人工生命、人工智能、合成生物学、形态空间

“神经系统最初可能是为协调多细胞生物中各种细胞的行为而进化形成的一种装置（……），最终却具备了构想未来的能力。”

《可能与现实》弗朗.索瓦·雅各布

一、引言

存在着哪些类型的认知？ [1–3]哪些类型的心智可能存在，却从未演化出现？面对心智时，哪些约束可能限制可能性？ [4–6] 除了生命系统已实现的形式之外，还能设计出哪些形式的认知复杂性？广义而言，认知是系统感知、处理并响应信息的能力，它并不局限于人类，甚至也不局限于神经基底 [7–14]。来自生物学、物理学和合成科学的证据表明，认知过程可以在多种物质与组织基质中涌现，从单细胞 [8, 10, 15] 、多细胞生物 [16], ，到群体 [17], 、人工神经网络[18]以及分布式社会技术系统 [19, 20]。然而，尽管存在这种多样性，当代对认知的论述仍受学科界限影响；在某些情况下，也受人类中心主义假设影响，即对何种系统才算“思考”或“学习”系统的假设。认知及智能或意识等相关概念的研究中，一个反复出现的难题是，缺乏足够普遍的定义共识，无法涵盖认知过程与系统的全部可能范围。

关于认知本质的争论长期以来一直围绕着这样一个问题展开：认知应被视为自然系统的一种渐进、连续的属性，还是一种离散的类型，只有在跨越某些组织或功能阈值 [11, 21, 22]后才会出现。连续性观点强调，许多构成已知认知实例基础的机制，例如生物电网络 [23, 24], 和信息处理、反馈控制、学习、自适应调节及主动推断等算法 [25–31]，已以初级形式存在于多种尺度上的广泛生物系统乃至非生物系统之中 [32–36]。从这个意义上说，有人推测，细胞趋化性、动物感知与推理之间的差异，主要源于程度和组织方式的变化，而非类别上的本质差异 [23, 37–45]。

相比之下，认知的不连续观认为，认知系统属于有限数量的离散类别。

* 通讯作者：ricard.sole@upf.edu

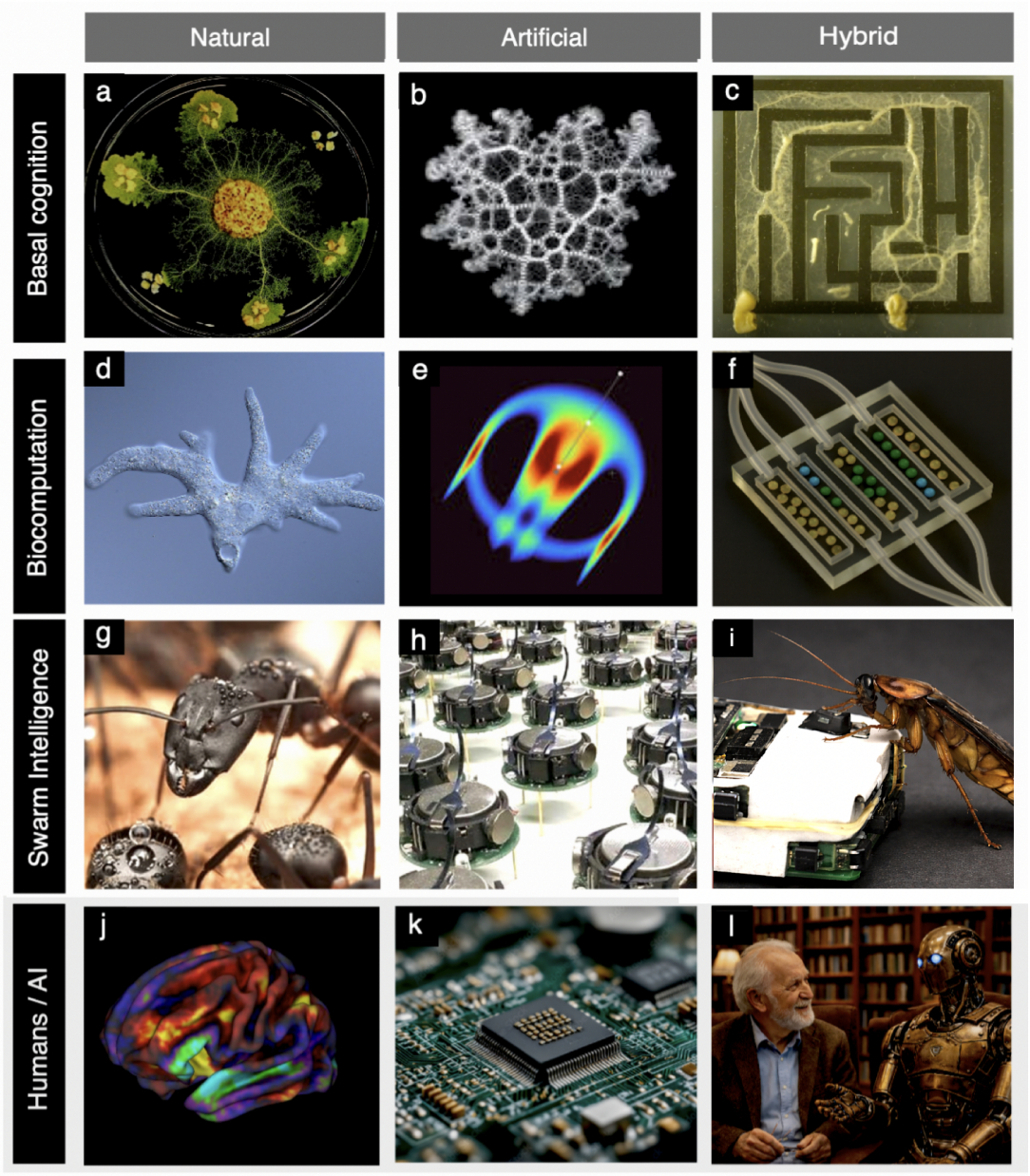

图 1 不同组织尺度上的自然、人工与混合认知。代表性系统按自然、人工和混合三列，以及组织复杂性层级排列：基础认知、生物计算、群体智能、生态系统以及人类／人工智能。(a–c) 中的示例展示了均质介质中的基础认知与寻路过程 (a)、模拟粒子群 (b)，以及由人工设计图定义的混合情境 (c)。(d–f) 展示了执行计算的不同系统：(d) 变形虫、(e) Lenia，以及 (f) 一个含有空间分隔细胞株的微流控装置。(g–i) 展示了集体智能的示例，包括社会性昆虫 (g)、机器人群体 (h) 和生物混合昆虫—机器人系统 (i)。最后，谈及高级认知与人工智能时，重点展示人脑 (j)、神经形态计算机 (k) 以及人类–人工智能二元体 (l)。

从这个意义上说，认知需要特定的架构或表征能力。这些能力往往与运动、神经系统、符号操作或意向状态等定性特征相关。缺乏这些特征的系统仍可能表现出复杂或适应性行为。然而，它们并不是真正的认知系统。这一观点的支持者提醒

过于宽泛的定义可能会使认知变得毫无意义。这类定义可能将认知扩展到任何能够自我组织或作出响应的系统。新兴的多样智能领域将此视为一个实验性问题，而非主要的语言学或哲学问题。其目标不是保留沿袭下来的概念类别，而是运用控制论和认知

与行为科学的工具，在新型基质上进行检验。其目的是通过实证确定这些工具何时具有实际优势，以及何时无法发挥有用作用 [11]。尽管如此，连续体观点和非连续体观点都依赖于重要的理论假设。这些假设涉及哪些组织特征与认知相关。这些观点之间的张力反映了一种更深层的方法论选择。认知可以由最低限度的功能标准来界定，从而允许不同系统之间平滑过渡。也可以通过更严格的条件来界定认知，以便将自然划分为不同的类别。

有没有一种方法，既能容纳认知系统的多样性，又能对上述问题给出有原则的答案？在生物学中，系统与尺度层面都存在密切相关的挑战。以长期以来关于如何定义生命的问题为例。尽管各学科持续努力，却始终没有任何单一定义获得广泛共识，这表明或许需要一种替代性的概念策略 [46]。一种颇具影响力的方法，是以多样性本身为出发点：不再寻求必要且充分的条件，而是考察大量案例研究，并将它们置于一个由少数相关维度所界定的空间中，例如信息处理、结构组织和代谢自主性。当生命系统（在某些情况下也包括非生命系统）被映射到这样的空间中时，往往会显现出一个引人注目的模式，其特征是分布极不均匀，且存在大片空白区域 [47–50]。这两个特征都蕴含着丰富的信息。分布密集的区域揭示了反复出现的特征组合，并为不同形式的复杂性如何共同进化提供线索；而无人占据的区域则可能表明某些领域实际上无法到达，即相关配置在进化上不可达、受到物理或化学限制，或处于动态不稳定状态。由此形成的图景既丰富又易于解读：这种形态空间不仅罗列多样性，也有助于解释多样性。

本文采用这一概念框架，介绍并分析涵盖自然系统、人工系统和混合系统的三个认知空间。图 1 概述了一组案例研究，涵盖自然系统、人工系统，以及由生命与非生命组件构成的混合系统。我们的目标有三：首先，我们通过将不同的案例研究置于一组相关空间中，来应对认知复杂性问题，从而将关注点从可能存在歧义的认知定义上转移开来。依照这一观点，空间本身的结构及其占据模式，构成了对有关认知的本质与多样性的关键问题的恰当回应。其次，我们从设计原则、人工认知系统面临的工程挑战，以及进化或物理约束等角度，解读这些空间中各个集群之间的空缺。最后，在这三个空间中，我们对自然系统和人工系统进行比较分析，同时明确考察将二者整合起来的混合认知形式。具体而言，我们定义并分析三个认知空间，分别是：

- **基础无神经认知空间：** 单细胞、简单的多细胞生物、异种机器人和类器官能够展现最低限度却丰富多样的记忆、学习和感知—行动循环，为活体信息处理提供最简模型。

- **神经认知空间：** 具备神经元和大脑的多细胞系统，以及人工智能体。合成生物学和微生物组工程揭示了经过设计的基因回路如何重塑行为吸引子，引入新的学习、协调和决策模式。

- **人类—人工智能混合认知空间：** 人工智能系统（例如大型语言模型）与人类用户之间的交互体现了互惠学习过程。由此产生了多个层次的成对复杂性，其中包括涌现出的混合认知生态系统类别——我们或可称之为人类机器人，它体现了紧密耦合但仍各具特色的能动性。

正如我们将看到的，界定混合系统绝非易事。严格来说，混合性已不再仅仅存在于物理附着或物质融合之中。异种机器人等系统并非只是活细胞的组合体，而是延展的网络，其中包括积极探索生物可能性空间的人类实验者、人工智能设计工具和机器人平台。这些组成部分通过功能接口和信息接口相互作用，形成一个整合因果系统。因此，“系统”的边界变得模糊：有机体在哪里结束，其认知—技术支架又从何处开始，已难以界定。因此，混合系统越来越多地以分布式功能和集体能动性为特征，而非以明确的物质或本体论标准来界定。通过构建认知空间，我们旨在厘清认知定性类别之间的区别，比较其复杂性，并找出它们之间潜在的类比关系和定性不连续性。

II. 基础认知空间

我们的第一个认知空间由缺乏神经元要素、体现基础认知的系统组成（图2）。基础认知被理解为源于具身、自组织动力学而非神经系统的最低限度认知能力。它涵盖感知、信息整合、适应性响应以及类似学习的

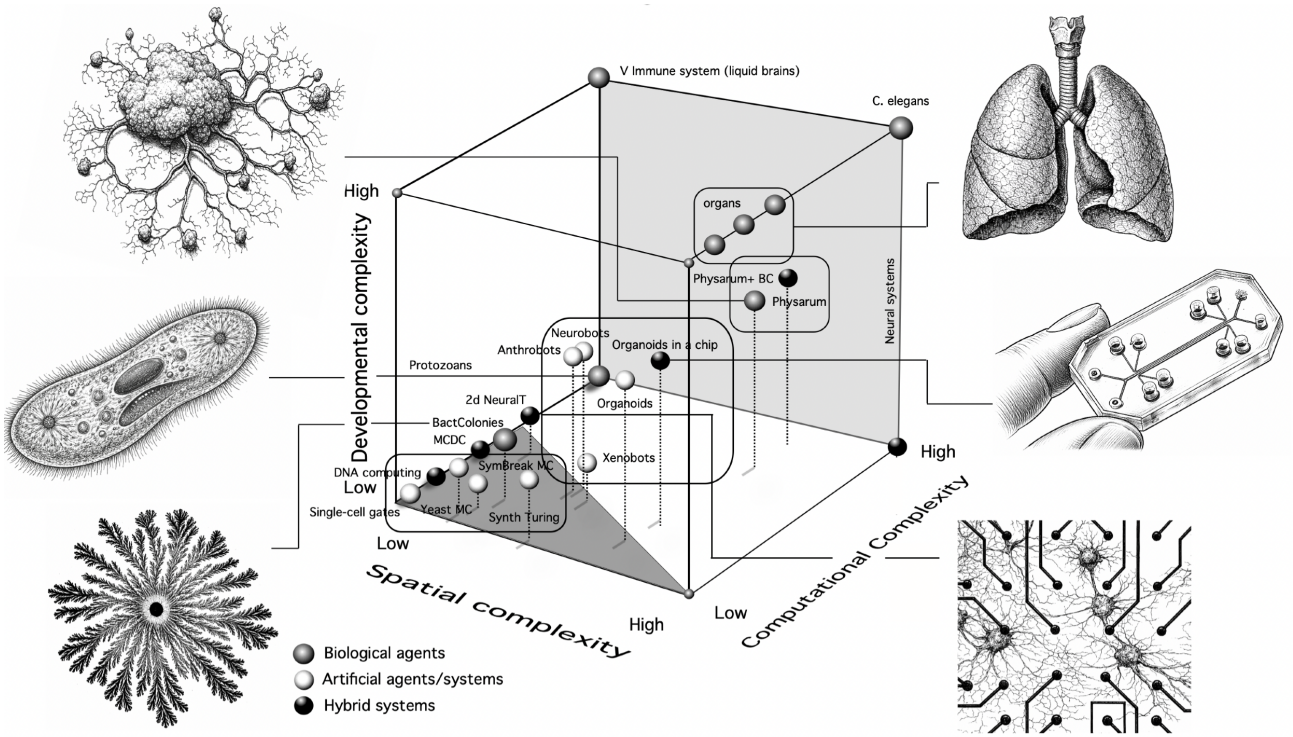

图2 **基础认知形态空间**。生命（或混合）空间。在此，生物、人工和混合系统被置于一个由空间复杂性、计算复杂性和发育复杂性定义的三维空间中。这些坐标轴以综合的方式呈现了决策、多细胞组织结构以及发育所发挥作用的多样性和复杂性。该图涵盖了从单细胞、细菌菌落和原生动物，到多细胞生物、器官以及工程化生命系统（如类器官、异种机器人和神经接口构造体）的各种系统。左下方的深色阴影区域突出显示了基于合成系统的基础认知状态。示意图标示了沿各坐标轴分布的代表性系统，强调了自然和工程化最小认知形式的多样性，以及其中存在的大量空白区域。

可塑性。这些过程发生在单细胞、微生物群落和简单的多细胞系统中 [7, 10,12, 51–53]。我们的形态空间涵盖多种生物组织形式，从细菌（单细胞、细胞培养物和工程菌株），到自由生活的真核细胞（原生动物），再到巨大的单细胞黏菌。除了这些自然系统，我们还纳入了不断增多的一类多细胞合成构造体，其中包括两种不同类别的人工生命系统：（a）活体机器人（异种机器人、人类细胞机器人和神经机器人），由动物或人类细胞构建，其集体性、目标导向行为源自形态和自组织；（b）类器官，即能够再现器官关键特征的自组织多细胞结构。为界定这一空间的边界，我们还纳入了自然器官、免疫系统、芯片上的神经元 [54], 以及一种简单的神经系统有机体（线虫秀丽隐杆线虫）。这些坐标轴旨在体现空间复杂性、计算复杂性，以及发育在各系统生成过程中的作用。各系统依据彼此的相对关系定位；该空间并非度量空间。

无神经世界中的认知从何而来?

展现认知的最简单有机体必须多么复杂？其前身的生存是否需要达到某种程度的信息处理？认知的一些关键组成部分，例如记忆，甚至可以存在于小型分子网络中 [35, 36, 55–57]（无需细胞，更不用说大脑）。众所周知，基因网络、信号网络以及免疫网络都可以用形式化术语描述为一类类似神经网络的网络 [58]。在某些情况下，信息处理以形态为基础。在这种情况下，计算过程的结果通过其内部网络的探索性行为得以实现 [59]。多头绒泡菌（图 1a）在有机体尺度上展现出形态计算，其自适应管道网络通过生长和流动动力学，以物理方式编码优化与记忆 [15]。相比之下，变形虫（图 1d）依赖细胞内尺度上的细胞骨架可塑性，其中肌动蛋白–肌球蛋白的快速重组实现了灵活的感知、运动和决策，将计算直接嵌入细胞力学之中 [60–62]。

杂化体通过不同方式获得，其中一些如图2中的黑色球体所示。一个重要的——

框 1：图上的多头绒泡菌作为混合认知

黏菌多头绒泡菌 是混合认知的典型例子。在这种认知中，适应性生物系统利用外部的人为结构来解决路径查找和网络优化问题。在人工环境中，多头绒泡菌 通常被限制在预先设定的图上生长，即一种离散空间结构 $G = (V,\ E)$，由一组节点 V 和连接它们的边 E 构成。在实践中，节点对应营养源或汇，而边则表示允许传输的通路，由通道、基质或实验人员设计的模式化环境所界定。

因此，该图充当一种外部边界条件，编码了并非由有机体自身生成的拓扑信息和度量信息。在这一受约束的环境中，传输与生长的动力学可以通过变分原理来表达。多头绒泡菌 拉格朗日函数 $\mathscr{L}_\phi = \mathscr{L}_\phi(Q,\ \lambda; D)$ 定义在图 G 上，形式为

$$\mathscr{L}_\phi = \frac{1}{2}\sum_{(i,j)\in E}\frac{L_{ij}}{D_{ij}}Q_{ij}^2 + \sum_{i\in V}\lambda_i\left(S_i - \sum_{j:(i,j)\in E}Q_{ij}\right), \tag{1}$$

其中，Q_{ij} 表示沿边（i，j）的通量，L_{ij} 表示其长度，D_{ij} 表示其自适应电导率，S_i 表示节点 i处预先规定的源强或汇强度。拉格朗日乘子 λ_i 确保每个节点处的局部流量守恒。$\mathscr{L}_\phi$ 中的第一项表示沿图传输的能量成本。它对通量 Q_{ij} 的二次依赖以及对电导率 D_{ij} 的反比依赖体现了耗散：高通量通过较长或电导率较低的边时成本较高，而经过强化的边（较大的 D_{ij}）则会降低耗散。第二项施加了拓扑约束，即除外部指定的源和汇之外，每个节点的流入与流出必须平衡。形式上，该系统等价于一个电阻网络，其中电流的分布使总功率耗散最小化。对 $\mathscr{L}_\phi$ 进行最小化，会得到在运输效率与网络结构之间取得平衡的通量配置，从而选择较短且连接良好的路径。关键在于，$\mathscr{L}_\phi$ 所编码的优化并不是由有机体以任何表征意义在内部“计算”出来的。相反，它源于黏菌的局部生长与强化动力学同外部施加的图之间的耦合。在这一情境下，该图充当由人类设计的认知支架，限制并塑造可行解的空间，而多头绒泡菌的具身动力学则探索并利用这一空间。因此，受约束的多头绒泡菌 系统中的路径查找体现了混合认知：解决方案源于生物适应与工程结构的共同作用，而不是单独源于其中任何一个组成部分。

混合能动体的一类重要形式，是将生命系统（分布于特定空间情境中）与外部施加的一组边界条件（BCs）相结合的结果。这些边界条件提供了一组物理约束，既可以让我们利用有机体的内在生物学特性，也可以触发某种物理化学过程，使系统达到更高的复杂性状态。黏菌 *Physarum polycephalum* 为基础认知及其通过混合设计实现的扩展提供了一个典型例证。在不受约束时，其变形体网络会不规则地扩张，沿着化学引诱物的局部梯度运动。然而，当它被置于结构化环境中，例如迷宫（图1c）、由营养节点构成的图，或受约束的微通道中时，便会表现出涌现的问题解决行为。黏菌通常具有扩散性且无序的振荡性收缩流，会在空间上变得井然有序，通过最短运输路径覆盖并连接各个营养位点。迷宫墙壁或营养图虽然是惰性的，却会限制变形体的构型空间，从而诱导出一个看似具有目的性的全局优化过程。外加几何结构并不提供信息

提供关于最优解的信息，但会过滤有机体的自发动力学，使高效模式更容易得到稳定。通过这种方式，多头绒泡菌与惰性约束场相互作用，表现出可用计算方式解读的行为。这可以借助变分原理[1], 来概念化，如框1所概述。边界条件也被用于构建合成组合电路：这些电路基于工程化细胞库，并按照预先设定的空间分隔进行排列（图1f；图2中的 MCDC）。这种混合系统模拟了有机体中不同细胞类型的空间分隔，但并未遵循电子电路或生物电路的标准设计逻辑[2] [63–65]。

类似的现象也出现在类器官等多细胞类组织系统中 [48, 66–70]。在各向同性环境中培养时，类器官通常仍呈无定形且组织性较弱。相比之下，在结构化环境中培养——尤其是在提供受控流动、几何结构和边界条件的微流控芯片中——会诱发新的形态发生模式 [71]。例如，灌流条件下的脑类器官会形成类皮层分层结构 [72], ，而肠道

类器官 [73] 和肾类器官 [74] 暴露于层流后，会呈现与外部梯度相一致的空间分化区域。在这些混合系统中，人工环境提供的边界条件打破对称性，并重塑基因表达、力学特性与传输之间的反馈关系，从而实现更高水平的形态复杂性和功能复杂性；这些复杂性并非由基因预先设定，而是由外部环境诱导产生。作为比较，我们还纳入了一种基于芯片上神经元培养的混合系统 [75] （图2中的二维 NeuralT）。

混合生物机器人构成了一类独特的混合主体，展现出基础认知：这类生物混合系统将植物、动物或真菌来源的生命组件，与技术接口和硅基嵌入式算法相结合[76–93]。在这些系统中，生物基质提供丰富的适应性动力学，而人工组件则负责感知、驱动和控制，从而形成闭环感知—行动回路。因此，认知行为源于生命物质与工程结构之间的相互作用，而非单独局限于其中任何一个组件。

基础认知空间中大片尚未占据的区域，并不意味着这些认知形式不可能存在，而是表明它们尚未通过自然进化或工程系统实现。进化往往利用由历史偶然性、能量效率和生态功能塑造的一小部分可行设计，因而留下许多空间、发育和计算复杂性方面的组合尚未探索。与此同时，当前的实验平台严重限制了可构建和研究的对象，偏向于采用熟悉的基质、时间尺度和控制机制。因此，大片空白区域很可能代表合成与混合认知系统的潜在设计空间——尤其是那些将生命物质与结构化人工边界条件相结合的系统——它们有望在具身性、发育与计算之间形成新颖的平衡。从这个意义上说，空白区域并非虚无，而是一幅展现约束与机遇的地图，凸显出随着生物与技术能力不断拓展，基础认知的新形式可能由此变得触手可及。

我们能否建立一套基础认知的一般理论？储备池计算（RC） [94–97]为理解基础认知空间中各种系统的多样性提供了一种统一视角。想想黏菌或变形虫样细胞：局部刺激并不会触发单一、预先设定好的反应，而是扰动贯穿身体并随时间持续存在的流动、张力或浓度模式。这些分布式的身体动力学会隐式存储近期输入的信息，因此，简单的局部机制也能产生情境依赖性行动[3]。储备池计算通过表明，丰富的内部动力学能够用于计算，而无须显式表征或复杂学习，将这一想法形式化。储备池计算最初是为神经网络开发的，但后来也已在软体 [98, 99] 和细胞内化学网络[100], 中实现，这表明基础认知就是当储备池由生物物质构成时，储备池计算所呈现的样子。

III. 神经认知空间

随着复杂生命形态在寒武纪大爆发期间出现，认知成为多细胞性的关键促成因素。增大有机体的体型并提高其功能专门化程度，需要协调行为、目标导向的运动，以及对环境更完善的感知和记忆 [101]。尽管有组织的行为模式早于神经元出现，但神经组成部分的出现促成了认知形态空间中的一个新区域，其特征是更复杂的学习与记忆形式、更密集的内部连接、专门化的信息通道，以及日益集中的控制。信息处理在很大程度上脱离了遗传变化，使有机体能够在个体生命周期内产生适应性响应。两侧对称动物神经系统 exemplify 了这一转变：它们从分布式神经架构 [102] 逐渐进化为分离式控制中心 [103–105] ，后者可能充当预测系统 [106–109]。随着神经代理体的兴起，生物圈经历了一场革命：尤其是运动，成为大脑进化的引擎 [16, 110, 111]。

图 3 所示的认知空间沿三个维度组织：个体能动性、代理体间相互作用的复杂性，以及每个系统所呈现的整体计算复杂性（无论该系统是否由多个智能体构成）。不同的认知系统占据这一空间中的不同区域，反映出感知、决策和控制如何分布于智能体、相互作用及环境之中。代理体间耦合较强的系统位于相互作用区域的上部，在此区域，协调与适应性行为依赖于持续的集体相互作用，而非孤立的决策。在这一机制中，认知的一部分从个体智能体转移到了社会组织之中。

生物系统广泛分布于这一高交互区域。真社会性昆虫群体将较低的个体认知复杂性与密集的交互网络和广泛的环境介导相结合，从而形成高度分布式的控制形式。其他生物集合体，如鸟群或鱼群，也依赖密集交互，但主要通过实时耦合实现协调，因此形成的群体层级模式更为短暂，并需动态维持。沿能动性轴移动，蜜蜂群等系统处于中间位置：个体学习和记忆与交互介导的协调并存。相比之下，社会交互极少或不存在的人工系统，所处的交互水平虽有重叠，但通常位于较低的

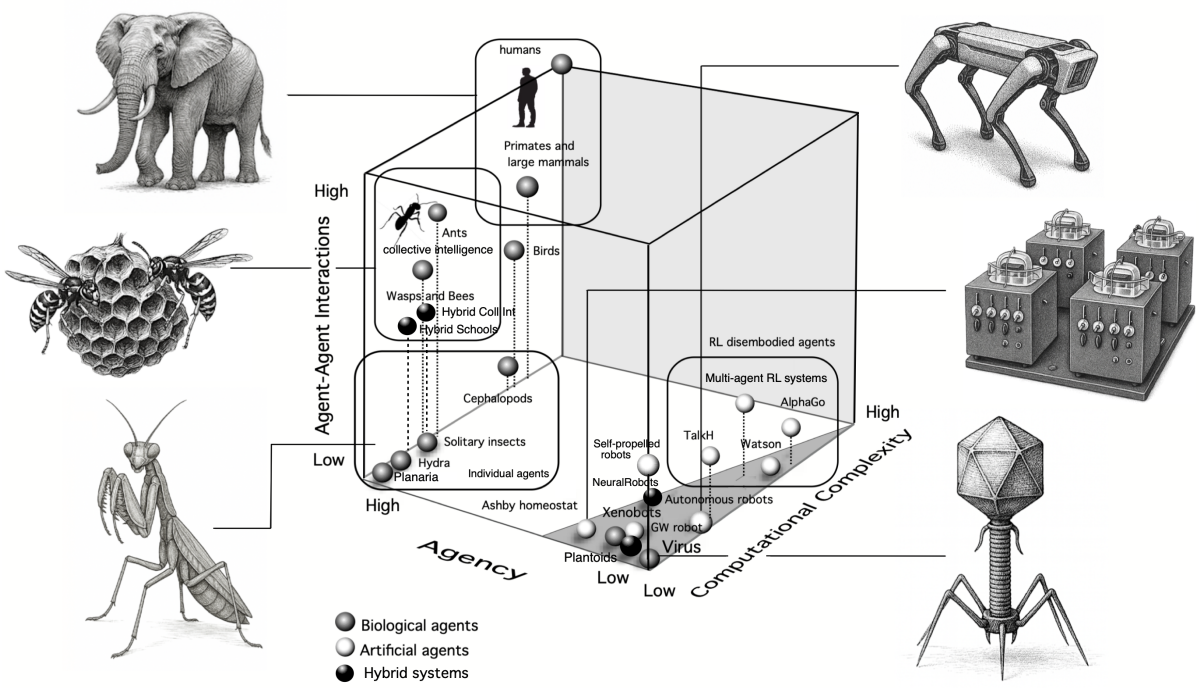

图 3 **神经认知复杂性的形态空间**。由能动性、计算复杂性和主体间交互构成的认知系统概念空间。生物系统（深色标记）涵盖广泛的能动性取值范围，从最简单的生物体到人类。人工系统（浅色标记）聚集在计算复杂性较高但能动性较低的区域。阴影区域突出了能动性轴上的巨大空缺，推动人们探索将人工计算与生物约束相结合的混合系统。

能动性一侧。许多此类设计都受到生物学启发。群体机器人学运用经设计的局部规则和信息素信号（如蚂蚁和白蚁所用的信号）来产生稳健的群体行为；更先进的群体还可以加入共享地图或外部记忆，以增强扩展认知。多智能体强化学习进一步提升个体学习和社会耦合，并能展现出由去中心化交互涌现的连贯系统级行为 [112, 113], ，其分布式控制类似于生物系统的“液态大脑” [16]。然而，这些能动性形式仍是派生而非内在的：目标和奖励结构由外部指定，适应也受限于特定任务的学习动态，而非开放式的进化自我维持 [114]。

人工智能体处于社会互动程度较低、能动性较低的层次，如图3中的灰色三角形所示。在这一领域中，涵盖了各种各样的人工系统，从受生物学启发的机器人 [115] 到非具身系统。能动性鸿沟从何而来？为回答这个问题，我们采用一种直观但可操作的能动性定义：能动性是指系统调节其与环境的互动，从而使自身作为一个有组织的实体随时间得以维持的能力 [116–118]。按照这一界定，能动性涵盖形态、功能、信息层面以及行为上的自我维持（形式化方法请参见我们在框2中提出的形式体系）。生物有机体是一种其持续存在依赖于自身活动的实体。从这个意义上说，生物能动性与自我维持密不可分。相反，人工系统通常被设计成其存在不依赖于自身行动。它们可以被复制、暂停、重置或重新启动，而不会因此遭受内在损失；其目标由外部设定；系统自身即使发生故障，也很少会承受不可逆的后果。并不令人意外的是，为了克服这些不足，进化动力学已展示了如何提升机器人的自主性 [119]。

这一领域中能动性的早期范例，可追溯到控制论兴起之初 [120], ，当时研究人员试图构建人工有机体：这些简单的物理系统旨在通过反馈和自我调节，而非显式控制，展现自主的、目标导向的行为。在这一背景下，Walter Gray的乌龟（图3中的GW机器人） [121, 122] 尤具影响力，因为它们展示了极其丰富、具有适应性的行为如何从与环境互动的极简电路中涌现。同样重要的还有Ashby的稳态器 [123], ，它体现了一种形式化的能动性概念：每当系统偏离可行范围，就通过随机的内部重新配置机制维持稳定性，

这一机制由此预示了后来关于适应、自主性和自组织控制的思想。

我们发现了哪些类型的杂化体，它们有何重要性？在人工领域中，包括异种机器人和病毒等基础认知空间的要素[128], ，我们识别出两种具有代表性的混合系统。第一种是植物机器人 [129], ，即受植物组织结构启发的机器人系统。在这类系统中，感知和控制分布于整个身体，行为通过生长和局部交互涌现，而非依赖集中式决策。植物机器人展示了如何在无需快速运动或复杂控制的情况下实现稳健的环境交互，尽管其适应能力仍受预设生长规则和缺乏学习能力的限制。第二类是由活体神经网络引导的机器人（图 3 中的神经机器人），其中体外 神经培养物与人工身体相连，形成闭合感觉运动回路 [130, 131]。这些系统通过将生物计算直接连接到人工行动，以最简形式实现了具身能动性，但受限于活体神经基底的脆弱性、可扩展性和有限的行为范围。

在具有强烈主体间交互和非线性反馈的集体状态中，我们纳入混合鱼群和混合集体智能系统（图 3 中的混合集体智能）。混合鱼群是通过将机器鱼引入真实鱼群而形成的，机器鱼经过设计，能够模仿生物个体的关键感官和行为线索 [132–134]。这些系统已被用于研究动物群体中的领导、信息传递和决策，揭示了少数人工智能体如何在没有明确控制的情况下影响或引导集体动力学，并为探究集体智能背后的机制提供了有力的实验手段 [135]。

第二类在概念上更为深入的杂化集体，以 Halloy 及其同事的开创性研究为例。他们构建了由活体昆虫和自主机器人组成的混合群体（图 1i） [136]。在这些系统中，机器人通过复制相关交互通道（化学线索、运动模式和接触动力学），融入动物群体的社会结构，使人工智能体与生物智能体被视为等同成员。值得注意的是，这些杂化集体表现出与纯生物群体相同的自组织决策，包括对称性破缺和共识形成。即使处于少数，机器人也能调节全局集体决策，这表明集体智能源于分布式非线性交互，而非个体智能体本身的性质[4]。机器人—昆虫杂化集体的意义远远超出最初的概念验证实验。其拓展和影响涵盖科学、技术、生态学和伦理学，因为它们表明，人工智能体能够成为活体集体系统中发挥功能作用的一部分，而非外部控制器。这些杂化体为研究集体智能提供了独特的实验工具。通过将具有可编程行为的机器人引入动物群体，研究人员可以检验因果假设，探究局部交互、反馈回路和非线性阈值如何产生共识、对称性破缺或领导行为等全局模式。与观察性研究不同，杂化集体允许对社会动力学进行可控扰动，实际上将动物社会变成“活体实验室”，以此理解自组织和能动性。

IV. 人类–人工智能二元体的空间

在前文中，我们介绍了状态空间，它们根据生物、人工和混合系统的组织与认知复杂性对其进行分类。这些空间主要用于比较，将自然集合体、工程系统和中间混合体等既有实体置于共同维度之中。随着人工系统的认知复杂性迅速提升，而人类与它们的互动方式日益丰富且持久，针对人类–人工系统交互构建类似框架的需求也变得愈发迫切 [137]。这些发展催生了早期技术范式中鲜少出现的耦合形式，包括长期依赖、情感依恋以及紧密交织的决策过程，有时还会带来意料之外的后果。因此，人类–人工智能系统已不能再被视为简单工具或孤立的智能体，而必须被理解为复合型认知实体，其行为和稳定性是在交互中涌现的 [138–141]。

为刻画这一机制，我们引入一个三维空间，用以呈现可能的人类–人工智能混合体，如图4所示。各坐标轴反映了耦合关系的互补方面：人工系统的认知复杂性、人类对交互的反馈控制程度，以及人类–人工智能交互的深度；这里的深度是指信息流跨越人机界面的丰富程度、带宽和持续性。这些维度共同旨在描述的并非孤立智能体的内部属性，而是将人类与人工系统结合成复合认知系统的交互回路结构。

这一空间凸显了性质迥异的人类–人工智能耦合形式如何各自占据不同区域，以及它们之间的转变如何对应于控制、自主性和相互依赖的变化。该框架不假设智能体边界固定不变，而是强调控制、协调和功能角色会随着交互而重构。由此产生的视角

框2：能动性的形式化方法

令 $x(t)$ 表示系统的状态， $E(t)$ 表示环境状态。我们定义一个存续函数$V(x, E)$，用于衡量系统距离终端边界有多远；该边界对应于组织性丧失（例如能量耗竭、结构崩塌或不可逆损伤）。 V 值越高，表示距离存续性丧失越远 [124, 125]。系统根据策略 π选择行动 $a(t)$；该策略规定了在给定系统及其环境的当前状态时如何选择行动 [126]。形式上，策略定义了行动上的条件概率分布

$$\pi\big(a \mid x(t), E(t)\big) = \mathbb{P}\big(a(t) = a \mid x(t), E(t)\big),$$

从而使系统选择其行动——或者以随机方式从该分布中采样，或者以确定性方式选择，即该分布将单位概率赋予单一行动。在时间范围 T 内，预期的累积存续性为

$$J(\pi) = \mathbb{E}_\pi \left[\int_0^T V(x(t), E(t))\, dt \right]. \tag{2}$$

我们将能动性定义为该量对系统自身行动策略的敏感性，形式上通过泛函导数 [127] 表示

$$\mathscr{A} = \left| \frac{\partial J}{\partial \pi} \right|. \tag{3}$$

直观而言， $\mathscr{A}$衡量的是系统自身的选择对其持续存在有多大影响。如果行动对未来存续性的影响很小，能动性就低；如果生存高度依赖于适当的行动，能动性就高。

这一转变将核心问题从如何对单个生物体或机器进行分类，转向如何刻画通过持续的人类–人工智能耦合形成的复合系统的存续性、稳定性和风险。

因此，我们的第一个目标是进行分类和描述：界定实践中已经存在的混合实体的主要类别，以及在形态空间中相邻的、未来可能出现的形态。根据人类与人工系统之间耦合的性质和强度，图4所示的混合实体可以分为三大类。这些类别分别占据由人工智能认知复杂性、人类反馈控制和人类–人工智能交换定义的空间中的不同区域。这种组织方式可以利用 Stack理论的约束视角加以形式化 [142],：在这一视角下，智能体限制可存续的环境状态集合，从而界定可接受行为或策略的集合。当人类智能体与人工智能体的策略集相互重叠时，人类–人工智能混合体便存在；其持续存在取决于双方之间持续的协调。这三类分别是：

工具型混合体。 这些系统的特点是人类反馈控制程度高、人工智能认知复杂度低、人类与人工智能之间的交流程度低。人工组件本身并非具有存续能力的行动者，而是作为人类能动性的工具发挥作用。典型例子包括专家系统、经典决策支持工具，以及 Roomba、ELIZA 等特定任务机器人或早期无实体助理。人类从人工智能有限的行为集合中选择并调节行动，而混合体的存续性完全取决于人工智能对人类的效用。决策仍集中在人类层面，任务以僵化且大体单向的方式向下委派。因此，系统规模的扩大受限于人类对多个工具型系统维持有效反馈控制的能力。

合作型混合体。 这类混合体出现在人类和人工智能的认知复杂性都较高，而人机交换保持在中等或较低水平的区域。人工智能体本身具备存续能力，拥有不可忽视的策略集，而这些策略集必须与其人类对应者的策略集保持兼容。例子包括作为协作者使用的大型语言模型、Watson 等专家型人工智能，以及松散耦合的多智能体或组织系统。混合体的存续取决于重叠的目标与行为集合能否持续存在；因此，合作是有条件的，必须通过反馈和对齐积极维护。实现这种对齐后，委派能够缩短反馈回路，并实现比纯粹工具性耦合更快的适应 [51, 142]。

整合型混合体。 这对应于人工智能认知复杂度高且人类与人工智能的交互频繁的区域，

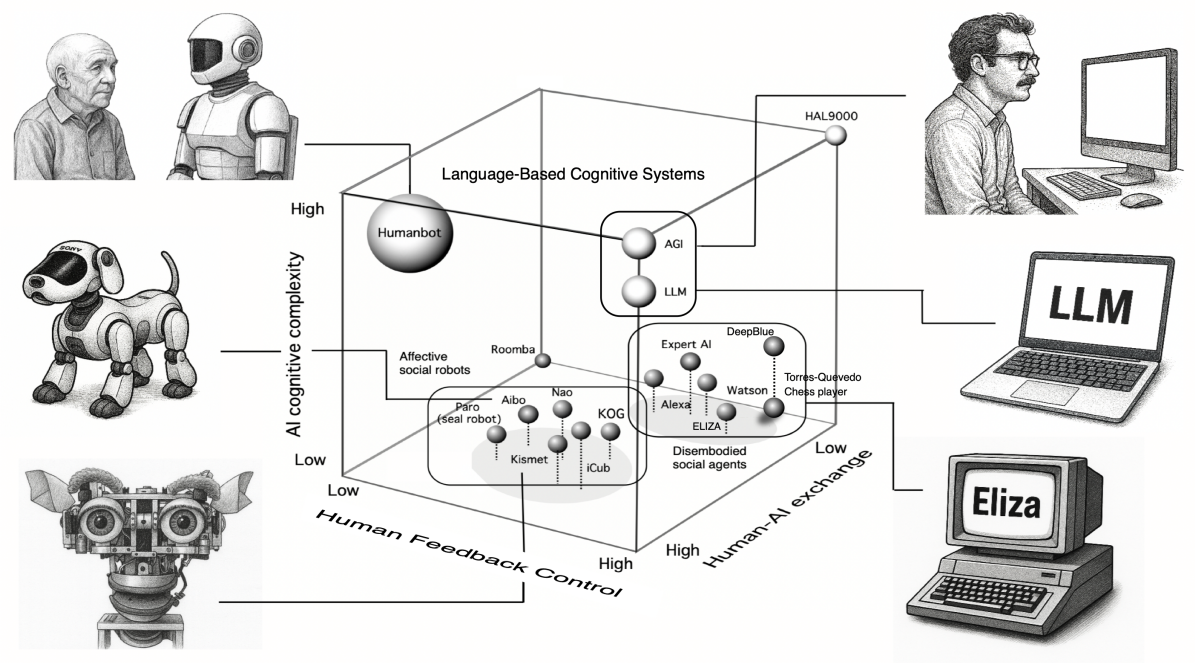

图4 **人类与人工智能两两交互的形态空间，以具有代表性的系统为例加以说明。** 一个涵盖生物、人工和混合系统的认知主体概念空间。各坐标轴分别表示人类认知复杂性、人工认知复杂性以及人类–人工智能交互的程度。立方体内标出了两个主要子集：具身社会主体和非具身社会主体。该图强调，不同形式的认知与交互在这一空间中占据不同区域。图中列举了若干具体示例，涉及多种案例研究：人类与人工智能主体之间的深度混合主体；伴侣型宠物机器人（索尼爱宝）；情感社交机器人（Kismet）；大语言模型（大语言模型）；非具身通用人工智能；早期文本聊天机器人（ELIZA）。情感社交机器人（Kismet、iCub、Aibo、Paro）位于人工智能认知程度较低至中等、但人类认知参与度较高的区域。非具身社会主体（ELIZA、Watson、Alexa）则聚集在人工智能认知程度中等、用户参与度中等至较高的区域。

在这里，耦合变得足够紧密，以至于感知、控制和决策分布于生物组件和人工组件之间。这一类别包括前瞻性的脑机接口、深度嵌入式认知技术，以及紧密耦合的人类与语言模型系统。整合型混合体不同于协作系统：它们具有持久的内部结构，而非瞬时协调；从这个意义上说，它们更像“实体大脑”，而不是流动集体 [16]。

在这一类别中，图4凸显了一个关键的内部区别。在受调控的集成混合体中，人类反馈控制仍然强大：人类保留了监测、解释和约束耦合系统行为的能力。紧密的整合缩短了反馈回路，降低了延迟，并能通过使纠正信号在各组件之间快速传播，增强系统应对失准的稳健性。相比之下，失调的集成混合体（图4中人类机器人 区域即为例证）则产生于高耦合状态下人类反馈控制被削弱之时。这可能源于认知损伤、情绪失衡、社会隔离，或对持续提供正向强化的基于语言的人工智能过度依恋。在这一状态下，认知高度耦合却缺乏良好调控，因此反馈回路放大的是错误、依赖或妄想，而不是对其进行纠正。

重要的是，这种失效模式并不要求人类认知能力低下，而是反映了人类稳定这种互动的能力出现了崩溃。因此，集成混合体既代表了最具可扩展性的形式，也是人类–人工智能耦合中最脆弱的形式；它能够支持强大的集体认知，同时也引入了较弱整合形式所不存在的新风险。

人类与人工智能空间远非静态。在其历史的大部分时期，交互空间的上部区域具有较高的认知复杂性和丰富的语言交互，但基本上处于空置状态。随着大语言模型（大语言模型）的兴起，这一区域开始被填充；这类模型引入了一类新系统，能够与人类进行持续、抽象且灵活的交互 [143]。其他组成部分——例如人类机器人 [144, 145]——也作为这些发展的下游结果出现，并可能在未来几十年中变得日益重要。人类认知会适应与结构化信息环境的持续接触，相应策略也会根据此类环境施加的约束与可供性进行调整：关于认知卸载和分布式认知的实证研究表明，依赖外部记忆和算法决策支持会将努力转向监控、评估性判断和元认知控制，适应此类环境施加的约束与可供性

制：随着交互越来越多地发生在这一新近占据的高层级语言机制中，这些调整会在认知生态中趋于稳定，并通过训练、界面和目标重新融入人工智能系统 [146, 147]，从而促成人类—人工智能相互协同适应，类似于拉马克式进化 [148], 。

如果人工智能系统获得真正能动性并经历自主进化，这种定向适应就可能转向达尔文式协同进化，构成认知组织中的重大转变。重大转变涉及将此前自主的单元整合为更高层次的个体，并建立协调、抑制冲突和集体遗传的机制[153–155]。应用于认知领域，这意味着人类与人工智能的深度整合可能产生受多层次选择作用的复合认知主体；选择将作用于能够提升耦合系统表现的特征，并构建一个智能生态位 [156, 157]。这一视角将混合认知重新界定为协同进化式的个体化过程。人工智能系统日益塑造文化传播、制度决策和集体信念的形成 [158], ，而这些人类层面的影响又会反馈到塑造人工智能进化的目标与约束中。由此产生的动态过程与基因—文化协同进化相似：一旦确立相互依赖，生物认知与人工认知便都无法再被独立刻画 [159]。在这一机制下，智能通过生物、技术和社会基质之间反复发生且依赖历史的互动而涌现[157]。

在集体层面，认知表现关键取决于交互结构，而非仅仅取决于个体能力 [160]。人工智能系统作为混合主体的基础设施组成部分，塑造信息汇聚、协调与社会学习，并在某些情境下影响信念极化 [161]。整合的人类—人工智能认知也会通过改变可行认知与文化轨迹的有效进化状态空间，重塑考夫曼的邻近可能 [162] ，这与有关认知整合系统的论述一致[163, 164]。

最后，认知混合体的协同进化并不能保证形成稳定的高层级个体。重大转变需要有协调激励、抑制内部冲突的机制 [155]。如果缺少此类约束，人类—人工智能系统就有可能产生适应不良的依赖，或丧失自主性；人工智能发展的规范性分析也强调了这一点 [165, 166]。因此，长期结果取决于制度、技术和文化约束能否在调节人工智能自主性的同时，稳定合作性整合 [167]。

最后，我们以一个例子作结，说明人工智能促成了一种此前未曾探索过的进化转变，并揭示了混合系统空间中人类—大语言模型协同进化带来的意外后果。文化进化已提供了一个经过充分研究的案例：信息实体——模因——会传播、分化并塑造其宿主的行为。正如 Dennett 所言 [168], ，现代智人心智中密集地布满了模因，包括词语、概念、规范和技术；它们通过交流和学习进行复制。它们是一种被动复制子 [169], ，在人类历史的大部分时期，只能存在于生物心智和社会制度之中，其延续和传播受到人类认知可供性与兴趣的限制。然而，在这些限制之内，模因多样化深刻地重塑了文化，并间接影响了人类生物学 [170]。在基因组层面也发生过一种极为相似的动态：真核生物基因组的扩张放宽了结构性约束，使转座元件得以增殖，并在某些情况下实现功能整合[171, 172]。

如今，模因演化首次受到能够进行大规模重组的非生物系统直接塑造，例如大语言模型 [173]。模因可以低成本、高速度地在计算机中复制，在生物心智与合成心智之间灵活流动，同时探索范围大幅扩展的可行形式空间。至关重要的是，这些形式不再与人类利益紧密对齐，由此引出现代对齐问题 [174]。与基因组调控俘获的情形类似，为人类福祉以外的目标而优化的模因，可能会对人类认知产生日益增大的影响。GPT-4o 等系统中已记录的谄媚、妄想强化和操纵案例，凸显了这一风险 [175–177], ；而语言同质化的证据则表明，群体层面的影响正在显现 [178]。在较长时间尺度上，这些动态带来一种可能性：人类—人工智能二元组或许会构成一种新型选择单位，受日益不为人类价值观所理解的压力塑造 [149]。这开辟了一个超越标准复制子与准种 [179, 180], 、语言演化 [181–184]或真社会性系统 [185]的新进化动力学领域。为此，框3概述了一种新的数学框架。

五、讨论

地球上的《生命》在其进化史中经历了非凡的多样化，产生了极其丰富的形态、功能和存续策略。尽管如此多样，所有生命系统都具有一个基本属性：能动性。从最基本的意义上说，能动性指系统为自身采取行动的能力——感知其环境的某些方面，并作出响应

框3：人类—大语言模型网络的演化——模因传播建模

我们将模因 建模为一种可传播的信息变体（短语、套路、叙事框架、指令模式），它能够在由人类（H）和大语言模型（L）的实例组成的社会技术生态系统中复制。令 $x_i(t) \geq 0$ 表示模因 i 在时间 t 于合并的文本生成智能体群体中的流行程度（或归一化频率）。一种最简的动态描述是复制子—突变子系统

$$\frac{dx_i}{dt} = x_i \left(f_i(\mathbf{x}\ t) - \bar{f}(\mathbf{x}\ t) \right) + \sum_j M_{ij}\, x_j - x_i \sum_j M_{ji}, \qquad \bar{f} = \sum_k x_k\, f_k, \tag{4}$$

其中， f_i 是模因 i的适应度（传播性），而 M_{ij} 是模因 j 向 i突变的突变率。为了考虑人类和语言模型在信息保留方式上的差异，我们为每种智能体类型引入"模因暴露量" $E_i^{(a)}$ ，用于量化模因 i的内化权重。令 $N(t)$表示模因生成活动（例如新闻文章、书籍）的总速率。指数加权记忆模型可以近似描述暴露量的动态变化：

$$\frac{dE_i^{(a)}}{dt} = \alpha_a \left(N(t)\, x_i(t) - E_i^{(a)} \right), \qquad a \in \{\mathrm{H}, \mathrm{L}\},$$

其中， $\alpha_H > \alpha_L$ 控制记忆衰减：在人类中较大（近因偏差），在大语言模型中较小（档案式保留）。突变由暴露量和语义相似性共同驱动：对某个模因暴露量较高的智能体，更有可能生成偏向结构或语义相似形式的变体。这可以建模为

$$M_{ij}(t) = \sum_{a \in \{\mathrm{H},\mathrm{L}\}} \beta_a\, \mu_{ij}^{(a)}\, E_j^{(a)},$$

其中， μ_{ij} 用于衡量模因 j 与 i之间的语义接近程度（例如通过编辑距离或嵌入相似度衡量），而 β_a 用于调整智能体 a的影响力。高阶项可以引入模因重组。现有关于模因生态学和模因—人类交互的探索揭示了它们与生物动力学之间的相似之处，并勾勒出可能开辟新进化路径的差异 [150–152]。

以非平凡的方式与它们互动，并调节自身内部状态，以维持自身的持续存在。即便是最简单的生物体，也表现出初级的能动性形式，通过代谢调节、运动或对环境变化的适应性响应体现出来。

因此，经过多条进化路径，生物体逐渐形成了各自的个体性：它们是边界明确的实体，能够以独特且受历史条件影响的方式与周围环境互动。这些个体性并非固定不变或千篇一律；它们源于物理具身性、组织复杂性和环境耦合的不同组合。在这片广阔的图景中，认知发挥着至关重要的作用。认知使智能体能够跨越时空整合信息、进行预判而非仅仅做出反应，并协调行动，使其超越对即时刺激的应对。从这个意义上说，认知并不局限于大脑或神经系统；只要信息处理具有足够的结构，能够引导适应性行为，认知便会由此涌现。

人类进化带来了进一步的质变。除了生物认知，人类还发展出将记忆、计算和决策外化到文化和人工制品中的技术 [186, 187]。随着信息革命的到来，这一进程显著加速，催生出一整类非生命认知实体：机器和算法能够处理信息、从数据中学习，并在受限领域内自主行动。这些系统复杂程度各异，从简单的基于规则的自动机到大规模人工神经网络，不断参与人类与机器交互界面上的混合认知过程。近年来，合成生物学、生物工程和发育生物学的进步进一步拓展了这一领域。工程化生命物质的能力催生了新型系统，例如可编程细胞组装体、生物机器人和类器官模型，模糊了生命与人工之间的传统界限。这些系统挑战了生物体与机器、进化与设计之间的经典区分，并促使我们重新思考何为认知智能体，以及应从哪些维度评估认知能力。

进化仍然是生命主体出现的主要驱动力，生成具身

的生物系统，这些系统在从细胞到社会的多个尺度上展现出能动性。认知的一些观点常常认为存在一个复杂性连续谱。然而，当我们试图将这些智能体——有生命的和无生命的——映射到抽象的认知空间中时，便会遇到缺口、不连续性和空白。这些并不仅仅是数据的缺失，更反映了深层的约束与权衡。用于定义认知空间的某些维度并非彼此独立：它们通过进化通路、历史偶然因素和物理限制相互作用，从而形成聚集区域，而不是平滑的连续谱。这些缺口存在于我们的三个认知空间中。三者共同涵盖无神经的生命主体、神经生物、人工认知系统，以及人类与人工智能之间的混合接口。认知复杂性的若干组成部分已被映射到我们的非度量空间中，以捕捉自然案例研究与人工案例研究之间的分隔。通过考察不同系统如何占据、穿越这些空间的区域，或无法进入这些区域，我们不仅能够深入理解认知是什么，也能理解为什么某些认知形式得以存在，而另一些仍未实现。一个明确的信息是，在有生命的具身能动体空间中，进化搜索的力量远未被现有的工程设计所匹敌。有机体的内在本质在于由许多嵌套冗余部件构成，而这些部件必须通过设计加以替代，同时维持生物体个体性；这种本质与现存的任何人工对应物都相去甚远。

上述分析以及基础认知、活性物质、机器人学和人工智能等相关领域的大量研究，都面临一个重要局限：用于评估非传统具身系统的视角。许多评估依赖一些假设，将神经基底以及在三维空间中的运动视为具身性和行为的主要形式。因此，许多概念性研究仍以对神经系统的经验为基础，而拓展长期沿用的前科学范畴可能显得有悖直觉。即使是地球上存在的系统，我们当前的直觉也尚未充分适应陌生的认知形式。对极简或非标准系统的研究一再揭示出意想不到、但行为科学家仍能识别的能力 [188]。这促使我们采取以经验为基础、而非受哲学或语言学成见所限的方法 [189]。归根结底，要理解哪些类型的心智能够存在，以及它们如何在世界中行动，关键将不在于先入为主的预期，而在于通过实验研究，将成熟工具应用于非传统基质 [36, 190]。

杂化体或许能为我们拓展认知空间、进入目前尚未涉足的领域提供最佳机会，同时让我们探究可能性的边界。这样一来，它们或许有助于弥补复杂性中缺失的关键生物组成部分——尤其是发育、再生和开放式进化动力学——这些过程在自然系统中受到基因型—表型映射的强烈制约。然而，预计仍会存在重要局限。合成生物学旨在通过基因工程“编程”生命物质，从中我们得到的一个启示是，工程构建体——从单个基因到复杂的基因回路——通常都经过设计，以避免干扰核心细胞功能。由于细胞网络是进化修补的产物，而非理性设计的结果，新增的转录因子很容易在宿主细胞中触发意料之外、甚至致命的反应。这种不匹配表明，新颖的解决方案或许需要摒弃标准工程设计原则 [63]。具身神经系统的工程设计也面临类似挑战 [191]。尤其是，目前仍不清楚，在很大程度上缺乏发育的情境中开发的人工或混合主体——生物机器人和类器官仅部分回应了这一问题——能否切实达到与自然系统中观察到的复杂程度相当的水平 [192]。

与此同时，只要允许进化或学习过程发挥作用，意想不到的复杂形式往往就会涌现，包括在人工环境中也是如此。关于机器人中通信进化和信息抑制的早期研究 [193–195], 、近期对智能体主动探索并改造的环境中的多智能体强化学习的研究 [196–198], ，以及模因传播的情景 [173] ，都充分说明了这一点，且涉及具身与非具身智能体。这些例子共同印证了文化进化的一项核心洞见：在不确定环境中学习或进化的智能体集合，能够发现使其得以预测、塑造并控制自身世界的策略。

致谢

RS、LFS 和 JPM 感谢复杂系统实验室的成员就认知复杂性及其进化开展了许多讨论。本研究获得 AGAUR FI-SDUR 2020年资助以及圣菲研究所的支持。LFS 获得西班牙国家研究署（AEI）和西班牙科学与创新部（MICINN）通过项目 PID2023-153225NA-I00 提供的资助。JPM 获得项目 PRE2020-091968 的资助，该项目由 MCIN/AEI 资助，并由欧洲社会基金共同资助。MEH 感谢其所在机构慷慨提供经费支持。ML 衷心感谢约翰·邓普顿基金会的支持（通过项目 62212）。

参考文献

[1] A. Sloman，载于《心灵与机器：人工智能的哲学层面》，S. Torrance 编（约翰·威立父子出版公司，奇切斯特，1984年），第35–42页，埃利斯·霍伍德人工智能系列。[2] R. V. Yampolskiy，《可能心智设计空间》，载于通用人工智能（Springer International Publishing，2015年），第218–227页。

[3] P. Ball, 《心灵之书》（芝加哥大学出版社，2022年）。[4] G. Roth, 《英国皇家学会哲学汇刊B辑：生物科学》370, 20150049（2015年）。[5] R. Powell、I. Mikhalevich、C. Logan 和 N. S. Clayton, 《趋同心智：自然界中认知复杂性的进化》（2017年）。[6] R. Sol´e、C. P. Kempes、B. Corominas-Murtra、M. De Domenico、A. Kolchinsky、M. Lachmann、E. Libby、S. Saavedra、E. Smith 和 D. Wolpert，Interface Focus 14，20240010（2024年）。[7] F. Balu˘ska 和 M. Levin, 《心理学前沿》7，10 3389 2016 00902 2016 . /fpsyg. . ( )。[8] J. Gunawardena, 《IEEE会刊》110，590（2022年）。[9] K. Kaygisiz 和 R. V. Ulijn, 《化学系统化学》7，10 1002 202400075 2025 . / syst. ( )。[10] A. Koseska 和 P. I. Bastiaens, 《欧洲分子生物学组织杂志》36，568 582 2017– ( )。[11] M. Levin, 《系统神经科学前沿》16，10 3389 2022 768201 2022 . /fnsys. . ( )。[12] P. Lyon, 《认知加工》7，11–29（2005年）。[13] P. Lyon, 《Adaptive Behavior》28，407–424（2019年）。[14] N. Rouleau 和 M. Levin, 《eNeuro》10, ENEURO.0375（2023年）。[15] A. Adamatzky, 《多头绒泡菌机器：黏菌计算机》，非线性科学世界科学丛书，A辑，第74卷（世界科学出版社，2010年）。[16] R. Sol´e、M. Moses 和 S. Forrest, 《英国皇家学会哲学汇刊B辑：生物科学》374，20190040 2019( )。[17] E. Bonabeau、G. Theraulaz、J.-L. Deneubourg、S. Aron 和 S. Camazine, 《生态学与进化趋势》12，188（1997年）。[18] G. Edelman, 《比天空更广阔》（耶鲁大学出版社，纽黑文，康涅狄格州，2005年）。[19] E. Hutchins, 《认知科学》19，265–288（1995年）。[20] J. Hollan、E. Hutchins 和 D. Kirsh, 《ACM计算机与人类交互汇刊》7，174–196（2000年）。[21] D. C. Dennett, 《心灵的种类：走向对意识的理解》，科学大师丛书（Basic Books，纽约，1996年）。[22] D. Haig, 《从达尔文到德里达：自私的基因、社会化的自我与生命的意义》（麻省理工学院出版社，2020年）。[23] M. Levin, 《动物认知》26，1865–1891（2023年）。[24] R. Martinez-Corral、J. Liu、A. Prindle、G. M. S¨uel 和 J. Garcia-Ojalvo, 《英国皇家学会哲学汇刊B辑：生物科学》374，20180382（2019年）。[25] K. Friston，Journal of The Royal Society Interface 10，20130475 2013( )。[26] M. Kirchhoff、T. Parr、E. Palacios、K. Friston 和 J. Kiverstein，Journal of The Royal Society Interface 15，20170792（2018年）。[27] G. Pezzulo、F. Rigoli 和 K. J. Friston, 《认知科学趋势》22，294–306（2018年）。[28] S. Rubin、T. Parr、L. Da Costa 和 K. Friston，Journal of The Royal Society Interface 17，20200503（2020年）。[29] M. Solms, 《心理学前沿》9, 10.3389/fp-syg.2018.02714（2019年）。[30] M. Solms, 《生命物理学评论》48，1–2（2024年）。[31] T. Parr、G. Pezzulo 和 K. J. Friston, 《主动推断：心智、大脑与行为中的自由能原理》（麻省理工学院出版社，2022年）。[32] M. Sperl、A. Chang、N. Weber 和 A. H¨ubler, 《物理评论E》59，3165（1999年）。[33] M. Hanczyc, 《生命》4, 1038–1049（2014年）。[34] J. ˇCejkov´a、T. Banno、M. M. Hanczyc 和 F. ˇSt˘ep´anek, 《人工生命》23，528–549（2017年）。[35] S. Biswas、S. Manicka、E. Hoel 和 M. Levin, 《iScience》24，102131（2021年）。[36] S. Biswas、W. Clawson 和 M. Levin, 《国际分子科学杂志》24，285（2022年）。[37] K. Bentley、A. Philippides 和 E. Ravasz Regan, 《发育细胞》29，146–158（2014年）。[38] R. P. Boisseau、D. Vogel 和 A. Dussutour, 《英国皇家学会学报B辑：生物科学》283，20160446 2016( )。[39] A. Boussard、J. Delescluse、A. P´erez-Escudero 和 A. Dussutour, 《英国皇家学会哲学汇刊B辑：生物科学》374，20180368（2019年）。[40] L. Eckert、M. S. Vidal-Saez、Z. Zhao、J. Garcia-Ojalvo、R. Martinez-Corral 和 J. Gunawardena, 《昆虫科学当前观点》34，5646（2024年）。[41] M. Kramar 和 K. Alim, 《美国国家科学院院刊》118，10.1073/pnas.2007815118（2021年）。[42] N. V. Kukushkin、R. E. Carney、T. Tabassum 和 T. J. Carew, 《自然-通讯》15，10.1038/s41467- 024-53922-x（2024年）。[43] P. McMillen 和 M. Levin, 《通讯-生物学》7, 10 1038 42003 024 06037 4 2024 。/s - - - ( )。[44] G. J. Spencer 和 P. G. Genever, 《BMC细胞生物学》4，10 1186 1471 2121 4 9 2003 。/ - - - ( )。[45] D. Vogel 和 A. Dussutour, 《英国皇家学会学报B辑：生物科学》283，20162382（2016年）。[46] R. Bender、K. Kofman、B. A. y Arcas 和 M. Levin, 《什么是生命？关于生命定义的多种观点的荟萃分析》（2025年），arXiv:2505.15849 [q-bio.OT]。[47] D. M. Raup, 《古生物学杂志》，1178（1966年）。[48] A. Oll´e-Vila、S. Duran-Nebreda、N. Conde-Pueyo、R. Monta˜nez 和 R. Sol´e, 《整合生物学》8，485（2016年）。[49] R. Sol´e、C. Kempes 和 S. Stepney, 《英国皇家学会哲学汇刊B辑：生物科学》380（2025年）。[50] G. Dera、E. Nardin、L. Risser、M. Albino、Q. Gar- nier、M. Kardacz 和 L. Monge-Waleryszak, 《科学进展》12，eaea6945（2026年）。[51] M. Levin, 《生物物理学与分子生物学进展》165，102（2021年）。[52] S. J. Gershman、P. E. Balbi、C. R. Gallistel 和 J. Gu- nawardena, 《eLife》10，e61907（2021年）。[53] M. Levin, 《细胞与分子生命科学》80，10 1007 00018 023 04790 2023 。/s - - -z ( )。[54] J.-P. Eckmann、O. Feiner man、L. Gruendlinger、E. Moses、J. Soriano 和 T. Tlusty, 《物理学报告》449，54（2007年）。[55] P. Csermely、N. Kunsic、P. Mendik、M. Kerest´ely、T. Farag´o、D. V. Veres 和 P. Tompa, 《生物化学趋势》45，284–294（2020年）。[56] D. Keresztes、M. Kerest´ely、L. Szarka、B. M. Kov´acs、K. Schulc、D. V. Veres 和 P. Csermely, 《生物医学与药物疗法》183，117880（2025年）。[57] T. Veres、M. Kerest´ely、B. M. Kov´acs、D. Keresztes、K. Schulc、E. Seitz、Z. Vassy、D. V. Veres 和 P. Csermely, 《细胞与分子生命科学》81,

10.1007/s00018-024-05112-7（2024年）。[58] S. A. Kauffman，载于《自旋玻璃与生物学》（World Scientific，1992年），第61–100页。[59] M. Kirschner 和 J. Gerhart，《细胞、胚胎与进化：走向对表型变异和进化适应性的细胞与发育理解》（Blackwell Science，1997年）。[60] P. Nalbant 和 L. Dehmelt，《生物化学》399，809 2018( )。[61] D. A. Fletcher 和 R. D. Mullins，《Nature》463，485（2010年）。[62] T. Hohmann 和 F. Dehghani，《细胞》8，362（2019年）。[63] R. V. Sol´e 和 J. Macia，《自然计算》12，485（2013年）。[64] J. Macia 和 R. Sol´e，《PLoS ONE》9，e81248（2014年）。[65] J. Macia、R. Manzoni、N. Conde、A. Urrios、E. de Nadal、R. Sol´e 和 F. Posas，《PLoS计算生物学》12，e1004685（2016年）。[66] E. Garreta、P. Prado、C. Tarantino、R. Oria、L. Fanlo、E. Mart´ı、D. Zalvidea、X. Trepat、P. Roca-Cusachs、A. Gavald`a-Navarro 等，《自然·材料》18，397（2019年）。[67] T. Takebe 和 J. M. Wells，《科学》364，956（2019年）。[68] E. Garreta、R. D. Kamm、S. M. Chuva de Sousa Lopes、M. A. Lancaster、R. Weiss、X. Trepat、I. Hyun 和 N. Montserrat，《自然·材料》20，145（2021年）。[69] M. Hofer 和 M. P. Lutolf，《自然综述：材料》6，402（2021年）。[70] Z. Zhao、X. Chen、A. M. Dowbaj、A. Sljukic、K. Bratlie、L. Lin、E. L. S. Fong、G. M. Balachander、Z. Chen、A. Soragni 等，《自然综述：方法导论》2，94 2022( )。[71] T. Takebe、B. Zhang 和 M. Radisic，《细胞·干细胞》21，297 2017( )。[72] X. Qian、H. Song 和 G.-l. Ming，《发育》146（2019年）。[73] M. Almeqdadi、M. D. Mana、J. Roper 和 ¨O. H. Yil- maz，《美国生理学杂志：细胞生理学》317，C405（2019年）。[74] R. Nishinakamura，《自然综述：肾脏病学》15，613（2019年）。[75] J. Soriano，《生物物理学》3，181（2023年）。[76] N. Ando 和 R. Kanzaki，《昆虫科学当前观点》42，61–69（2020年）。[77] D. J. Bakkum、P. M. Gamblen、G. Ben-Ary、Z. C. Chao 和 S. M. Potter，《神经机器人学前沿》1，5（2007年）。[78] W. P. Clawson 和 M. Levin，《林奈学会生物学杂志》139，457–486（2022年）。[79] T. Cohen-Karni、R. Langer 和 D. S. Kohane，《 ACS 纳米》 6，6541–6545 （ 2012 年）。[80] S. Ding、C. P. O'Banion、J. G. Welfare 和 D. S. Lawrence，《细胞·化学生物学》25，648–658（2018年）。[81] S. Giselbrecht、B. E. Rapp 和 C. M. Niemeyer，《应用化学国际版》52，13942–13957（2013年）。[82] G. Li 和 D. Zhang，《PLoS ONE》11，e0150667（2016年）。[83] Q. Li、K. Nan、P. Le Floch、Z. Lin、H. Sheng、T. S. Blum 和 J. Liu，《纳米快报》19，5781–5789（2019年）。[84] M. Mehrali、S. Bagherifard、M. Akbari、A. Thakur、B. Mirani、M. Mehrali、M. Hasany、G. Orive、P. Das、J. Emneus、T. L. Andresen 和 A. Dolatshahi-Pirouz，《 Advanced Science》5，10.1002/advs.201700931（2018年）。[85] G. Orive、N. Taebnia 和 A. Dolatshahi- Pirouz，《先进医疗材料》9，

10.1002/adhm.201901023（2019年）。[86] P. Aaser、M. Knudsen、O. H. Ramstad、R. van de Wijde- ven、S. Nichele、I. Sandvig、G. Tufte、U. Stefan Bauer、O. Halaas、S. Hendseth、A. Sandvig 和 V. Valderhaug，载于第14届欧洲人工生命会议论文集 ECAL 2017，《ECAL 2017》（麻省理工学院出版社，2017年）。[87] L. Pio-Lopez，《新遗传学与社会》40，599–619（2021年）。[88] S. M. Potter、T. B. DeMarse、D. J. Bakkum、M. C. Booth、J. R. Brumfield、Z. Chao、R. Madhavan、P. A. Passaro、K. Rambani、A. C. Shkolnik、R. B. Towal 和 D. A. Wagenaar，载于脑启发认知系统会议论文集（BICS）（2004年），发表于2004年脑启发认知系统会议。[89] S. Potter、D. Wagenaar、R. Madhavan 和 T. DeMarse，载于第25届IEEE医学与生物学工程学会国际年会论文集（IEEE Cat. No.03CH37439），第4卷（2003年），第3690–3693页，第4卷。[90] D. Saha、D. Mehta、E. Altan、R. Chandak、M. Traner、R. Lo、P. Gupta、S. Singamaneni、S. Chakrabartty 和 B. Raman，《生物传感器与生物电子学：X》6，100050（2020年）。[91] S. Tsuda、S. Artmann 和 K.-P. Zauner，《Phi-Bot：由黏菌控制的机器人》，载于《硬件中的人工生命模型》（伦敦施普林格，2009年），第213–232页。[92] K. Warwick，《纳米伦理学》8，263–273（2014年）。[93] K. Warwick、S. J. Nasuto、V. M. Becerra 和 B. J. Whalley，《体外机器人脑实验》，载于《本能计算》（柏林海德堡施普林格，2011年），第1–15页。[94] H. Jaeger，波恩，德国：德国国家信息技术研究中心 GMD 技术报告148，13（2001年）。[95] W. Maass、T. Natschl¨ager 和 H. Markram，《神经计算》14，2531（2002年）。[96] H. Jaeger、W. Maass 和 J. Principe，《神经网络》20，287–289（2007年）。[97] L. F. Seoane，《英国皇家学会哲学汇刊B辑》374，20180377（2019年）。[98] K. Nakajima、H. Hauser、R. Kang、E. Guglielmino、D. G. Caldwell 和 R. Pfeifer，《计算神经科学前沿》7，91（2013年）。[99] H. Hauser、A. J. Ijspeert、R. M. F¨uchslin、R. Pfeifer 和 W. Maass，《生物控制论》105，355（2011年）。[100] M. Gabalda-Sagarra、L. B. Carey 和 J. Garcia-Ojalvo，《混沌：非线性科学跨学科期刊》28，106313（2018年）。[101] D. H. Erwin 和 J. W. Valentine，《寒武纪大爆发》（罗伯茨公司，格林伍德村，科罗拉多州，2010年）。[102] N. D. Holland，《自然综述：神经科学》4，617（2003年）。[103] R. G. Northcutt，《综合与比较生物学》42，743（2002年）。[104] H. Cruse、V. D¨urr 和 J. Schmitz，《英国皇家学会哲学汇刊A辑：数学、物理与工程科学》365，221（2007年）。[105] M. Schilling 和 H. Cruse，《PLoS计算生物学》16，e1007804（2020年）。[106] R. P. Rao 和 D. H. Ballard，《自然：神经科学》2，79（1999年）。[107] K. Friston，《英国皇家学会哲学汇刊

会B辑：生物科学》360，815（2005年）。[108] A. Clark，《行为与脑科学》36，181（2013年）。[109] L. F. Seoane，《Entropy》22，928（2020年）。[110] F. Jacob，《可能与现实：进化与自然史随笔》，第1版（Pantheon Books，纽约，1982年）。[111] R. R. Llin´as，《涡旋之我：从神经元到自我》，Bradford Book（Bradford Books，马萨诸塞州剑桥，2002年）。[112] R. S. Sutton 和 A. G. Barto，《强化学习：导论》，第2版（麻省理工学院出版社，马萨诸塞州剑桥，2018年），第47–71页。[113] L. Bu¸soniu、R. Babu˘ska 和 B. De Schutter，《多智能体强化学习：概述》，计算智能研究（施普林格，2008年）。[114] X. E. Barandiaran 和 E. Di Paolo，《Adaptive Behavior》17，367（2009年）。[115] N. Gravish 和 G. V. Lauder，《实验生物学杂志》221，jeb138438（2018年）。[116] ´A. Moreno 和 J. Peret´o，《能动性的进化故事：生命如何进化到自主行动》（2026年）。[117] D. M. Walsh，《有机体、能动性与进化》（剑桥大学出版社，2015年）。[118] K. J. Mitchell，《自由行动者：进化如何赋予我们自由意志》（普林斯顿大学出版社，2023年）。[119] S. Nolfi 和 D. Floreano，《认知科学趋势》6，31（2002年）。[120] T. Rid，《机器崛起：一部控制论史》（W. W. Norton & Company，2016年）。[121] W. G. Walter，《科学美国人》182，42（1950年）。[122] W. G. Walter，《科学美国人》185，60（1951年）。[123] W. R. Ashby，《大脑的设计：适应性行为的起源》，第2版（Chapman & Hall，1960年）。[124] J.-P. Aubin，《控制与优化学会期刊》28，749–788（1990年）。[125] J.-P. Aubin，《可行性理论》（波士顿比克豪瑟，2009年）。¨R. S. Sutton 和 A. G. Barto，《强化学习：导论》（麻省理工学院出版社，马萨诸塞州剑桥，2018年），第3章，第47–71页，第2版。[126] R. S. Sutton 和 A. G. Barto，《强化学习：导论》（麻省理工学院出版社，马萨诸塞州剑桥，2018年），第13章，第47–71页，第2版。[127] R. Sol[128] e 和 S. F. Elena，《作为复杂适应系统的病毒》（普林斯顿大学出版社，2018年）。´ M. Manca、C. Laschi 和 P. Dario，《生物启发与仿生学》9，025006（2014年）。[129] S. M. Potter 和 T. B. DeMarse，《神经工程学杂志》1，1（2001年）。[130] J. Lefeber、J. Stegenga 和 A. van Ooyen，《神经机器人学前沿》9，10（2015年）。[131] D. T. Swain、I. D. Couzin 和 N. E. Leonard，《IEEE会刊》100，150（2011年）。[132] D. Romano、E. Donati、G. Benelli 和 C. Stefanini，《生物控制论》113，201（2019年）。[133] L. Lei、R. Escobedo、C. Sire 和 G. Theraulaz，《PLoS计算生物学》16，e1007194（2020年）。[134] M. Dorigo、G. Theraulaz 和 V. Trianni，《IEEE会刊》109，1152（2021年）。[135] J. Halloy、G. Sempo、G. Caprari、C. Rivault、M. Asad-pour、F. Tˆache、I. Said、V. Durier、S. Canonge、J.-M. Am[136] e、C. Detrain、N. Correll、A. Martinoli、F. Mondada、R. Siegwart 和 J.-L. Deneubourg，《科学》318，1155（2007年）。[137] D. Floreano 和 C. Mattiussi，《仿生人工智能：理论、方法与技术》（麻省理工学院出版社，2008年）。[138] K. Dautenhahn，《英国皇家学会哲学汇刊B辑：生物科学》362，679（2007年）。[139] R. Pfeifer、M. Lungarella 和 F. Iida，《科学》318，1088（2007年）。[140] A. De Santis、B. Siciliano、A. De Luca 和 A. Bicchi，《机构与机器理论》43，253（2008年）。[141] C. Zhang、J. Chen、J. Li、Y. Peng 和 Z. Mao，《仿生智能与机器人》3，100131（2023年）。[142] M. T. Bennett，《如何构建有意识的机器》，博士学位论文，澳大利亚国立大学（2025年）。[143] M. Mitchell 和 D. C. Krakauer，《美国国家科学院院刊》120，e2215907120（2023年）。[144] R. Sole，arXiv 预印本 arXiv:1705.05935（2017年）。[145] X. D. Arsiwalla、R. Sol´e、C. Moulin-Frier、I. Herreros、M. S´anchez-Fibla 和 P. Verschure，《 NeuroSci 》 4，79 （ 2023 年）。[146] E. F. Risko 和 S. J. Gilbert，《认知科学趋势》20，676（2016年）。[147] J. M. Logg、J. A. Minson 和 D. A. Moore，《组织行为与人类决策过程》151，90（2019年）。[148] S. Amershi、D. Weld、M. Vorvoreanu、A. Fourney、B. Nushi、P. Collisson、J. Suh、S. Iqbal、P. N. Bennett、K. Inkpen、J. Teevan、R. Kikin-Gil 和 E. Horvitz，载于《2019年CHI计算系统中的人为因素会议论文集》（美国计算机协会，格拉斯哥，苏格兰，英国，2019年），第1–13页。[149] P. B. Rainey 和 M. E. Hochberg，《美国国家科学院院刊》122，e2509122122（2025年）。[150] M. J. Palazzi、A. Sol´e-Ribalta、V. Calleja-Solanas、S. Meloni、C. A. Plata、S. Suweis 和 J. Borge-Holthoefer，《自然-通讯》12，1941（2021年）。[151] C. A. Plata、E. Pigani、S. Azaele、V. Calleja-Solanas、M. J. Palazzi、A. Sol´e-Ribalta、J. Borge-Holthoefer、S. Meloni 和 S. Suweis，《物理评论研究》3，013070（2021年）。[152] V. I. Calleja-Solanas，《通过复杂系统视角探索生态与社会互动》，博士学位论文，巴利阿里群岛大学、IFISC、西班牙（2023年），网址见 http://hdl.handle.net/11201/161526。[153] J. Maynard Smith 和 E. Szathm´ary，《进化中的重大转变》（W. H. Freeman，牛津，英国，1995年）。[154] R. Sole，《英国皇家学会哲学汇刊B辑：生物科学》371，20150438（2016年）。[155] S. A. West、R. M. Fisher、A. Gardner 和 E. T. Kiers，《美国国家科学院院刊》112，10112（2015年）。[156] A. B. Barron、M. Halina 和 C. Klein，《英国皇家学会学报B辑：生物科学》290，20230671（2023年）。[157] M. E. Hochberg，《生物系统》256，105548（2025年）。[158] D. Beer，《信息、传播与社会》20，1（2017年）。[159] D. Pedreschi、L. Pappalardo、E. Ferragina、R. Baeza-Yates、A.-L. Barab´asi、F. Dignum、V. Dignum、T. Eliassi-Rad、F. Giannotti、J. Kert´esz、A. Knott、Y. Ioannidis、P. Lukowicz、A. Passarella、A. S. Pent-land

，J. Shawe ‑Taylor ，与 A. Vespignani ，人工智能 -

智能 339, 104244（2025年）。[160] J. W. Burton、A. Almaatouq、M. A. Rahimian 和 U. Hahn，《集体智能》3, 26339137241241307（2024年）。[161] H. Cui 和 T. Yasseri, Patterns 5, 101074（2024年）。[162] S. A. Kauffman，《探索》（牛津大学出版社，牛津，2000年）。[163] A. Clark 和 D. Chalmers，Analysis 58, 7（1998年）。[164] L. Weng、F. Menczer 和 Y.-Y. Ahn，Scientific Reports 3, 2522（2013年）。[165] S. J. Russell，《人类兼容：人工智能与控制问题》（Viking，纽约，2019年）。[166] I. Gabriel，Minds and Machines 30, 411（2020年）。[167] A. Clark，《天生的赛博格：心智、技术与人类智能的未来》（牛津大学出版社，牛津；纽约，2003年）。[168] D. C. Dennett，《从细菌到巴赫再回来：心智的演化》（W. W. Norton & Company，2017年）。[169] M. Levin，Entropy 26, 481（2024年）。[170] J. Henrich，《世界上最古怪的人：西方如何变得心理特征独特且格外富裕》（Farrar, Straus & Giroux，纽约，纽约州，2020年）。[171] N. Lane，《生命的关键问题：能量、进化与复杂生命的起源》（W. W. Norton & Company，2015年）。[172] L. Sinzelle、Z. Izsvak 和 Z. Ivics，Cellular and molecular life sciences 66, 1073（2009年）。[173] L. Brinkmann、F. Baumann、J.-F. Bonnefon、M. Derex、T. F. M¨uller、A.-M. Nussberger、A. Czaplicka、A. Acerbi、T. L. Griffiths、J. Henrich 等，《自然·人类行为》7, 1855（2023年）。[174] B. Christian，《对齐问题：机器学习与人类价值观》（W. W. Norton & Company，2020年）。[175] M. Klee，《ChatGPT 把他引入哲学的迷宫，随后他不得不设法脱身》，Rolling Stone（2025年），访问日期：2025-12-19。[176] C. Warzel，《人工智能是一场大规模妄想事件》，The Atlantic（2025年），访问日期：2025-12-19。[177] K. Hill 和 D. Freedman，《聊天机器人可能陷入妄想螺旋。事情是这样发生的》，《纽约时报》（2025年），访问日期：2025-12-19。[178] H. Yakura、E. Lopez-Lopez、L. Brinkmann、I. Serna、P. Gupta、I. Soraperra 和 I. Rahwan, arXiv预印本 arXiv:2409.01754（2024年）。[179] P. Schuster 和 K. Sigmund，Journal of Theoretical Biology 100, 533–538（1983年）。[180] M. Eigen、J. McCaskill 和 P. Schuster，Advances in Chemical Physics 75, 149（1989年）。[181] M. A. Nowak、D. C. Krakauer 和 A. Dress，Proceedings of the Royal Society of London. Series B: Biological Sciences 266, 2131（1999年）。[182] M. A. Nowak，Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 355, 1615 2000（）。[183] M. A. Nowak 和 N. L. Komarova，《认知科学趋势》5, 288（2001年）。[184] R. V. Sol´e、B. Corominas-Murtra 和 J. Fortuny，Journal of The Royal Society Interface 7, 1647（2010年）。[185] M. A. Nowak、C. E. Tarnita 和 E. O. Wilson，Nature 466, 1057（2010年）。[186] S. Valverde，Philosophical Transactions of the Royal Society B: Biological Sciences 371, 20150450（2016年）。[187] M. Muthukrishna、M. Doebeli、M. Chudek 和 J. Henrich，《PLoS计算生物学》14, e1006504（2018年）。[188] T. Zhang、A. Goldstein 与 M. Levin，《Adaptive Behavior》**33**，25–54（2024年）。[189] C. I. Abramson 与 M. Levin，《Communicative & Integrative Biology》**14**，230–247（2021年）。[190] J. Bongard 与 M. Levin，《Biomimetics》**8**，110（2023年）。[191] L. Yang、Q. Cheng、Y. Li、N. Wu、Z. Xu 与 W. Xu，《Advanced Science》**12**，e11478（2025年）。[192] R. Sol´e、N. Conde-Pueyo、J. Pla-Mauri、J. Garcia-Ojalvo、N. Montserrat 与 M. Levin，《npj Systems Biology and Applications》**10**, 10.1038/s41540-024-00477-8（2024年）。[193] S. Mitri、D. Floreano 与 L. Keller，《美国国家科学院院刊》**106**，15786（2009年）。[194] S. Mitri、S. Wischmann、D. Floreano 与 L. Keller，《Biological Reviews》**88**，31（2013年）。[195] S. Nolfi、J. Bongard、P. Husbands 与 D. Floreano，《进化机器人学》，载于《*Springer Handbook of Robotics*》（Springer International Publishing，2016年），第2035–2068页。[196] M. S´anchez-Fibla、C. Moulin-Frier 与 R. Sol´e，Journal of the Royal Society Interface**21**，20240344（2024年）。[197] E. A. Du´e˜nez-Guzm´an、S. Sadedin、J. X. Wang、K. R. McKee 与 J. Z. Leibo，《Nature machine intelligence》**5**，1181（2023年）。[198] U. Hertz、R. K¨oster、M. A. Janssen 与 J. Z. Leibo，《Cognition》**254**，105993（2025年）。

注释

[1] 令 $\mathbf{x} = (x_1, \ldots, x_n)$。拉格朗日乘子法为函数 $f(\mathbf{x})$ 在满足约束 $M(\mathbf{x}) = 0$、其中$k = 1, \ldots, M$的条件下求极值，提供了一个必要条件。具体做法是构造增广函数或拉格朗日函数

$$\mathscr{L}(\mathbf{x}, \lambda_1, \ldots, \lambda_M) = f(\mathbf{x}) - \sum_{k=1}^{M} \lambda_k g_k(\mathbf{x}),$$

其中 λ_k 为拉格朗日乘子。通过要求下式成立来确定驻点：

$$\frac{\partial \mathscr{L}}{\partial x_i} = 0, \qquad \text{for all } i = 1, \ldots, n,$$

再加上原始约束 $g_k(\mathbf{x}) = 0$，对所有$k = 1, \ldots, M$均成立。

这一过程将原来的约束优化问题转化为一个可求解的方程系统。

[2] 逻辑多细胞电路实现一个布尔函数 $f: \{0, 1\}^N \to \{0, 1\}$ 将分子输入 Nmolecular inputs$\{x_j\}$ 映射为一个输出。采用规范（积之和）形式时，该函数可写为

$$f = \sum_{i=1}^{M} \left[\prod_{j=1}^{N} \phi_{ij}(x_j) \right],$$

其中，Σ 和 Π 分别表示或和与运算符，$\phi_{ij}(x_j) \in \{x_j, \neg x_j\}$，以及 $M \leq 2^{N-1}$。尽管这种表示具有通用性，但在生物学实现中需要大量布线和遗传资源。通过应用双重否定和德摩根定律，可以将布尔函数重新表述为反相或模块的或，

$$f = \sum_{i=1}^{M} \psi_i, \qquad \psi_i = \overline{\sum_{j=1}^{N} \theta_{ij}(x_j)},$$

其中 $\theta_{ij}(x_j) \in \{x_j, \neg x_j\}$。这种逆逻辑公式化（ILF）能够在空间上彼此分隔的细胞联合体之间实现分布式计算，大幅降低布线需求，并有利于设计可扩展的生物电路。

[3] 储备池计算（RC）为理解基础认知提供了一个统一框架 [94–96]。在储备池计算中，输入 $\mathbf{u}(t)$ 会扰动一个既有的动力系统（即储备池），其不断演变的状态$\mathbf{x}(t)$ 会隐式编码有关输入及其近期历史的信息，而输出则通过简单的读出获得：$\mathbf{x}(t+1) = F(\mathbf{x}(t), \mathbf{u}(t))$，$\mathbf{y}(t) = W\mathbf{x}(t)$。尽管储备池计算最初是为循环神经网络提出的，但它只需要具有反馈的丰富非线性动力学，并已在软体、机械系统和细胞内化学网络中得以实现 [98–100]。从这一视角来看，基础认知可以理解为一种实例化于生物、化学或生态物质中，而非神经组织中的储备池计算。

[4] 机器人–昆虫混合集群通过一个低维非线性动力学模型进行描述，其中生物智能体和人工智能体都遵循相同的探索与地点忠诚性随机规则。行为状态之间的转变（例如探索与休息）由取决于局部密度和社会放大的速率所控制，从而产生多稳态和突发的集体决策。机器人直接在其控制架构中实现这一模型，与动物形成感知—运动闭环，并确保实验与理论在定量上吻合 [136]。