

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)07-2339-19

论文引用格式: Bao H, Liang T J, Zheng Y. 2026. The new-generation artificial neural networks in the 21st century — embodied cognitive neural networks: a review. Journal of Image and Graphics, 31(7):2339-2357(鲍泓, 梁天骄, 郑颖. 2026. 21世纪新一代人工神经网络—具身认知神经网络. 中国图象图形学报, 31(7):2339-2357)[DOI:10.11834/jig.260112]

21世纪新一代人工神经网络—具身认知神经网络

鲍泓^{1,2}, 梁天骄^{3*}, 郑颖²

1. 北京联合大学北京市信息服务工程重点实验室, 北京 100101; 2. 北京联合大学机器人学院/人工智能学院, 北京 100027;
3. 中国科学院大学计算机科学与技术学院, 北京 100049

摘要: 面向21世纪, 为构建一种具有自适应结构、可解释、泛化强和能效高的人工神经网络(artificial neural network, ANN), 对新一代ANN的基础理论、模型和架构进行探索。按时间线, ANN从20世纪40年代首次提出, 发展至今已80多年, 并将延展至21世纪中叶, 本文按5个维度特征将ANN划分成5个时代。当前, 处在第4代ANN, 其主要特征是数据拟合、深度学习、注意力机制和Transformer结构, 以大语言模型为基础的ChatGPT(chat generative pre-trained Transformer)为代表, 通过对话式图灵测试, 但属于“黑箱”测试, 并局限于离身智能形态的涌现。根源是基于规模扩展缩放定律的大语言模型本质上不对称性, 缺乏对现实世界物理规律的理解, 多模态、多任务、多形态输出智能呈锯齿型且能效低; 而具身智能形态机器人的神经网络还缺少自主智能, 只能按预设程序完成规定动作, ANN在离身智能和具身智能之间出现巨大鸿沟。为解决第4代ANN的这些重大缺陷, 需要新的理论、模型和架构支撑。当前, 围绕下一代ANN的发展方向和技术路线, 出现了很多争论和分歧。本文追溯前4代ANN主流理论、模型和架构的发展, 重点分析了第4代ANN的特征, 评述了面向第5代ANN的世界模型与联合嵌入预测架构、认知螺旋模型与智痕元胞网络架构。最后, 以认知物理学理论和驾驶脑认知技术实践为基础, 提出一种具有第5代ANN核心特征的具身认知物理神经网络(embodied-cognitive physics neural networks, E-CoPNN)轻量化架构, 基于具身认知架构研发的全球首个跨具身视觉语言基础模型, 在跨域的自动驾驶12项基准与具身智能17项基准均达到SOTA, 量化验证了统一具身认知模型的双向正向迁移能力。当今, 构建面向21世纪的新一代ANN, 在哲学上, 将从身心二元论转向具身知觉一元论; 在基础科学上, 将从20世纪的生物物理学升级到21世纪的认知物理学; 在数学上, 从量的关系拓展到形的关系; 在模型上, 将推动ANN研究范式从数据拟合转向结构重构; 在应用上, 将填补ANN发展中离身智能与具身智能的鸿沟, 为实现会学习、自成长、自纠错以及可交互的具身认知机器人广泛应用奠定基础; 在代际演进上, 将从第4代ANN跃升到具有类脑认知和动态自适应结构等特征的第5代ANN; 支撑认知为融合先导的四大科技“纳米—生物—信息—认知”聚合发展, 提高人类认知能力, 迎接新一轮的认知革命。

关键词: 人工神经网络(ANN); 第5代ANN具身智能; 离身智能; 具身认知; 认知物理学; 具身认知物理神经网络(E-CoPNN); 统一场论; 智能场; 人类注意力机制; 选择性机制; 驾驶脑认知; 综述

The new-generation artificial neural networks in the 21st century — embodied cognitive neural networks: a review

收稿日期: 2026-02-28; 修回日期: 2026-03-29; 预印本日期: 2026-04-05

* 通信作者: 梁天骄 liangtianjiao23@mails.ucas.ac.cn

基金项目: 国家自然科学基金项目(91420202, 61932012); 中英两国政府联合科学创新基金(牛顿基金)项目(UK - CIAPP/324)

Supported by: National Natural Science Foundation of China(91420202, 61932012); UK-China Joint Science and Innovation Fund(Newton Fund) Project(UK-CIAPP/324)

Bao Hong^{1,2}, Liang Tianjiao^{3*}, Zheng Ying²

1. Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China;

2. School of Robotics/School of Artificial Intelligence, Beijing Union University, Beijing 100027, China;

3. School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: For the 21st century, this study is dedicated to exploring the fundamental theories, models, and architectures of next-generation artificial neural networks (ANNs), with the objective of constructing high-performance ANNs that featuring adaptive topology, interpretability, strong generalization capability, and high energy efficiency. Since the initial proposition of ANNs in the 1940s, the field has witnessed over 80 years of development. Extending to the mid-21st century, ANNs can be categorized into five generations based on five core dimensions, and such generational evolution constitutes the core trajectory of neural network research and advancement. This study defines five generations of ANNs (abbreviated as xG-ANNs) from five perspectives: neuronal unit, information coding, network structure, learning mechanism, and Turing test. 1G-ANNs: Threshold logic networks, represented by the McCulloch-Pitts model and the perceptron; 2G-ANNs: Continuous activation networks (e. g. , sigmoid, tanh), typified by the classical back-propagation network; 3G-ANNs: Represented by spiking neural networks 4G-ANNs: Deep neural networks, represented by AlexNet, ResNet, Transformer architectures, and the attention mechanism, which have passed the conversational Turing test for disembodied intelligence; 5G-ANNs: Cognitive neural networks, with five defining characteristics: 1) cognitive units that integrate memory, reasoning, and human-like attention; 2) hybrid coding of semantic symbols and distributed representations; 3) modular cognitive architecture with dynamic reconfigurable topology; 4) meta-learning, causal reasoning, and lifelong learning; and 5) the embodied Turing test that remains as a breakthrough. A core academic consensus has been formed regarding 5G-ANNs. That is, such networks integrate neural computing, symbolic reasoning, and cognitive architecture with low-power consumption, support dynamic topology, memory cognition, neuro-symbolic fusion, and embodied/world models. They exhibit intrinsic advantages, including adaptive structure, few-shot generalization, interpretability, low energy consumption, and embodied cognition. At present, the field is in the era of 4G-ANNs, which are characterized by data-driven fitting, deep learning, the attention mechanism, and Transformer frameworks. Represented by large language model-based ChatGPT, 4G-ANNs have passed the conversational Turing test, but such validation is a “black box” assessment that is restricted to the emergence of disembodied intelligence. The root causes lie in the inherent asymmetry of large language models built on the scaling law of parameter expansion, their lack of comprehension of physical laws that govern the real world, and critical drawbacks, including jagged multimodal and multiform intelligent output and inferior energy efficiency. By contrast, neural networks that are deployed in embodied intelligent robots lack autonomous intelligence and can only execute predefined actions per programmed instructions, leading to a huge gap between disembodied intelligence and embodied intelligence in 4G-ANN systems. Addressing these critical limitations of 4G-ANNs calls for the support of novel theories, models, and architectural designs. At present, extensive debates and divergence persist over the developmental orientation and technical routes of next-generation ANNs. This study analyzes and summarizes the developmental progress of mainstream theories, models, and architectures of the first four generations of ANNs. It focuses on the characteristics of several typical 4G-ANN models and their enhanced variants, and reviews representative architectures for 5G-ANNs, including world models, joint embedding predictive architecture, cognitive spiral models, and intelligence-trace cellular network frameworks. Finally, grounded in theory of cognitive physics and the practical technologies of driving brain cognition, this work proposes a lightweight architecture, called embodied cognitive physics neural network (E-CoPNN), which incorporates the core characteristics of 5G-ANNs. In particular, E-CoPNN features dynamic topology (three-layer nested structure and dual systems), memory cognition (three categories of memory), neuro-symbolic fusion (integration of Euclidean and non-Euclidean data, combination of lexemes and concepts), and embodied cognition/world models (fusion of disembodied and embodied intelligence, integration of cognitive space and physical space). The proposed architecture embodies the typical properties of 5G-ANNs: an adaptive structure, few-shot generalization, interpretability, low power consumption, and embodied cognition. Targeting brain-like general intelligence, 5G-ANNs regard dynamic topology, autonomous memory, cognitive reasoning, and structural evolution as core attributes, shifting the ANN research paradigm from data fit-

ting to structural reconstruction, and thus, representing a pivotal direction for achieving machine autonomous intelligence and continuous learning. Conclusion and Significance: At present, the development of next-generation ANNs for the 21st century will drive a series of transformative revolutions: philosophically, a paradigm shift from mind-body dualism to embodied cognitive science based on embodied perception monism; theoretically, a disciplinary expansion from 20th century biophysics to 21st century cognitive physics; model-wise, a fundamental transition of ANN research from data fitting into structural reconstruction; in application, bridging the long-standing gap between disembodied intelligence and embodied intelligence in ANN development; generationally, an evolutionary leap from 4G-ANNs to 5G-ANNs characterized by brain-like cognitive capabilities and adaptive topological structures. This work lays a solid foundation for the widespread application of embodied cognitive robots with learning, self-development, self-correction, and human-robot interaction capabilities. Furthermore, it supports the convergent development of nano-bio-info-cogno, or the four cutting-edge technologies led by cognitive integration. Finally, this work enhances human intellectual competence and ushers in a new round of cognitive revolution.

Key words: artificial neural network (ANN); fifth generation ANN embodied intelligence; disembodied intelligence; embodied cognition; cognitive physics (cognitophysics); embodied cognitive physics neural network (E-CoPNN); unified field theory; intelligence field; human attention mechanism; selective mechanism; driving brain cognition; review

0 引言

智能的英文 intelligence 由拉丁语 intelligentia 演化而来,古希腊的 nous 是其思想原型,指人最高级的理性直观、洞察本质的能力。19—20 世纪,智能从哲学高级理性降到人的认知 (cognition) 能力,核心是“超越感官经验、把握事物本质的认知过程”,此时智能脱离出自然生命体的“聪慧”(“智慧”),成为人类心智中能够进行抽象思考、逻辑判断和概念把握的核心部分,这部分的边界不包括自然生命体中的意识、直觉等感性的因素。

1) 智能与认知:自然与人工的边界

1948 年,图灵在对智能本质研究后,将智能再扩展到机器智能,提出《智能机器》报告(内部版)(Turing, 1969),将机器的智能形态分为离身智能和具身智能。1950 年发表的《计算机器与智能》(Turing, 1950)提出振聋发聩之问“机器能思维吗”,并给出判断机器智能的“图灵测试”,宣告了人工制造的机器(非自然生命体)也有认知能力,能够产生像人一样的智能。

1956 年,达特茅斯会议研讨机器如何思维,有不同路线,但取得共识的是提出了 artificial intelligence,简称 AI,含义是人工或人造的非自然生命体生成的智能,中文直译为人工智能,目前尚无中文的简称,因此本文简称仍用 AI。人工智能彻底脱离自然生命体的“人”,用人造的机器模拟人的感知、记

忆、推理、学习和决策过程及其能力,因此,智能是自然生命体和非自然生命体都可以认知的能力的度量。在不同场合,自然/非自然生命体也称为有机/无机生命体或碳基/硅基生命体。

认知是人的思维过程,机器认知就是机器的思维过程,智能是认知的结果,是在认知过程中学习、理解和思考能力的度量。显然,认知和智能有区别也有直接关联,二者一体两面(鲍泓等, 2025):一体就是人或机器本体,两面是认知看过程,智能看目标。科学注重过程、方法和可验证机制,哲学关注目标、价值和终极意义。

2) 人工神经网络:首次提出与代际划分

1943 年,神经生理学家 Warren S. McCulloch 与数理逻辑学家 Walter Pitts 共同发表的研究论文 (McCulloch 和 Pitts, 1943) 中给出人工神经网络 (artificial neural network, ANN) 的定义:由互联的形式化神经元(阈值逻辑单元)构成的网络,可执行逻辑计算并模拟生物神经系统的信息处理。这是首次将 ANN 作为计算模型正式提出与定义的,并给出第一个人工神经元模型(M-P 模型),奠定了 ANN 的数学基础。

ANN 提出已 80 多年,其代际演进是反映神经网络研究和发展的核心脉络,本文在 Samsonovich (2022) 对 ANN 代际划分 (generational divide) 基础上,从神经元单元、信息编码、网络结构、学习机制和模型测试 5 个维度定义了 5 代 ANN (以下每代缩写为 xG-ANN)。1G-ANN: 阈值逻辑网络,代表: M-P 模

型、感知机;2G-ANN:连续激活网络(如 sigmoid、tanh 等),代表:经典反向传播 BP(back propagation)网络;3G-ANN:脉冲神经网络,代表:脉冲神经网络(spiking neurons network, SNN)(Maass, 1997);4G-ANN:深度神经网络(deep neural network, DNN),代表: AlexNet、ResNet(residual network)、Transformer, GPT-3(generative pre-trained Transformer 3)大语言模型产生涌现,通过对话式图灵测试;5G-ANN:认知神经网络,其多维特征是:(1)认知单元(记忆+推理+类人注意力)符合人的认知结构;(2)语义符号+分布式双编码;(3)模块化认知架构(动态拓扑);(4)元学习、因果推理、终身学习;(5)统一具身认知基础模型,通过具身图灵测试。核心特征归纳为:神经计算+符号推理+认知架构+低功耗,工程应用上具有自适应结构、小样本泛化、可解释、低功耗和具身认知等特征。

3) ANN 基础模型:数学同构、物理同源

早期 1G-ANN 的 M-P 模型提出后,各代 ANN 模型的能力不断增强,随之其数学基础不断拓展。当前 4G-ANN 的基础模型中,大语言模型(large language model, LLM)本质上就是一个函数,其中, $\mathbf{x}$ 是输入向量, $\mathbf{w}$ 是权重, $\mathbf{b}$ 是偏置向量, $\mathbf{w}$ 、 $\mathbf{b}$ 是模型参数。若要让函数具有非线性输出,则需加入激活函数,如 ReLU(rectified linear unit)、sigmoid 等。DNN 就是由许多单层神经网络叠加形成的复合函数。大模型的深度网络除去输入层和输出层,中间可达几十到上百层。一个有 L 层的网络,其中间层第 i 层有 n 个神经元, $i-1$ 层有 m 个神经元,则第 i 层 $\mathbf{w}$ 总共有 $m \times n$ 个参数, $\mathbf{b}$ 共有 n 个参数。这个有 L 层、带权重并加入激活函数的神经网络复合函数可表示为

$$f(\mathbf{x}) = \sigma(\mathbf{W}_L(\mathbf{W}_{L-1} \cdots \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) \cdots + \mathbf{b}_{L-1}) + \mathbf{b}_L) \quad (1)$$

大语言模型(LLM)在巨量的文本数据预处理阶段形成百亿甚至千亿级以上参数,从本质上看,一个 LLM 是一个条件概率函数,它学习的是文本序列的联合概率分布,模型的核心任务是预测下一个 token(中文称为词元)出现的概率,它就是一个多层嵌套的深度复合函数,由嵌入、位置编码、多层 Transformer 块和输出层构成,通过自回归生成智能,本质是一种依赖序列上下文、循环迭代的生成逻辑。

4G-ANN 可以看成是一个数据驱动的函数拟合

过程:输入数据 $\rightarrow$ 经过权重矩阵(向量) $\boldsymbol{\theta} \rightarrow$ 输出预测。但只有正向过程并不能保证函数拟合的准确性,因此网络中要有反向传播机制,反向传播可以算出每个权重(参数)应该变大还是变小、变多少,让模型预测更准。反向过程的数学关系是:用损失函数 $L(\boldsymbol{\theta})$ 算出每个权重(参数)的残差,残差是依赖于 $\boldsymbol{\theta}$ 的标量场,因此,这个 $L(\boldsymbol{\theta})$ 是个泛函,泛函的梯度就是自动微分的反向传播,用梯度更新权重,让损失变小,所以整个网络训练就是用有限维的权重去优化一个泛函。正向到反向的闭环:权重决定预测 $\rightarrow$ 预测产生残差 $\rightarrow$ 残差构成损失 $\rightarrow$ 损失算出梯度 $\rightarrow$ 梯度修正权重。如果在权重上有物理方程的约束,如偏微分方程(partial differential equation, PDE),则 ANN 也可当成方程求解器,如图 1 所示。

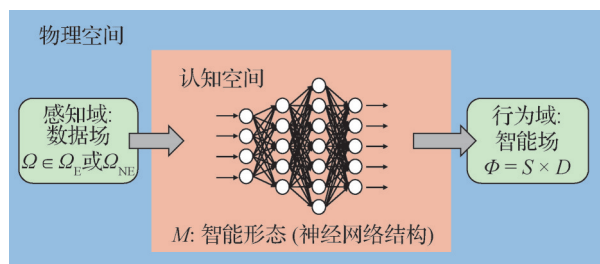


图 1 人工神经网络可以看成是函数或方程求解器

Fig. 1 ANN can be regarded as a function approximator or an equation solver

图 1 中,机器认知有两个空间:物理空间(蓝色)和认知空间(黄色)。ANN 是在认知空间网状结构。认知过程是神经网络在这两个空间中“感知—思维—行动”以及反馈的循环往复的活动。ANN 从物理空间中感知现实环境,感知域中的数据场(data field)用 Ω 表示,用以表示数据的时空结构和数据间的相互作用。数据有两类:欧氏数据(Euclidean data)和非欧数据(non-Euclidean data),分别用 Ω_E 和 Ω_{NE} 表示。数据的颗粒度由细到粗可分为像素(pixel)级、词元(token)级和概念(concept)级。各级数据向量的维度由高变低,一般用高维向量空间表示,其数学基础是线性代数和概率统计。如上所述,ANN 各层中的每个神经元(图中 $\circ$)的建模精度和信息编码方式体现在权重和偏置向量(参数)之中,数据正向经过多层神经网络的运算和学习,输出(涌现)整体智能到行为域,整体智能是智能场(intelligent fields)的形式,图中用 Φ 表示,智能场就是具有

时空连续性的智能行为域(作用域)。图中几个概念及其物理意义和数学关系如下:

Φ : 整体智能, 分为离身智能与具身智能、强智能与弱智能、通用智能与专用智能等, 物理意义: 智能场;

M : 智能形态(架构/模态/连接方式), 物理意义: 智能的结构/相态;

C : 资源(数据、参数、算力、时延等);

S : 智能强度, $S = f(C)$, 物理意义: 智能的功率/强度;

D : 智能密度, $D = S/C$, 物理意义: 智能的浓度/效率。

数学关系: 智能形态 M 决定 S 和 D 能达到的上限, 相变关系是 M 决定智能相的结构类型(不同结构的 ANN), S 决定智能相的强弱, D 决定智能相的效率。智能形态 M 是结构算子, 决定智能强度 S 与智能密度 D 如何被组织成整体智能 Φ , $\Phi = S \times D$ 。

综上所述, 4G-ANN 的数学基础是微积分、线性代数、函数、泛函和偏微分方程, 5G-ANN 支持结构表示、概念关系学习和几何先验等核心特征, 其数学基础要拓展到微分几何学、拓扑学和图论等, 数学上从量的关系拓展到形的关系。物理支撑: ANN 的智能涌现可类比为物质相变, ANN 的结构决定智能强度的增长速度, 资源决定智能强度的上限, 而智能密度反映结构对资源的利用效率, 总之, ANN 与物理同源, 与数学同构。

4) ANN 的智能相变: 模型参数的规模扩展与智能涌现的关系

ANN 的结构决定其智能强度 S 和智能密度 D , 并随参数量变化。例如, 4G-ANN 中有 3 种不同架构的神经网络: Transformer、混合专家模型(mixture of experts, MoE)和循环神经网络(recurrent neural network, RNN), 以大模型规模扩展的缩放定律(scaling laws)为基准(Kaplan 等, 2020), 它们的 S - D 相变关系如图 2 所示。图 2 中, 智能相空间(S - D space): 以智能强度 S 为横轴, 智能密度 D 为纵轴。智能相边界: 模型从“无智能”到“涌现智能”的相变分界线(图中深色区间的边沿)。智能轨迹: 模型扩参时, 在相空间中走的路径 = 强度上升、密度下降。最优智能点: $S \times D$ 最大 $\rightarrow$ 强度与密度的最佳平衡点。Transformer、MoE 和 RNN 对应不同颜色曲线; MoE 多专家稀疏结构在同等强度下获得更高密度, 可用于万亿

级参数超大规模多模态与多任务大模型; Transformer 自注意力机制的全局并行编码器—解码器架构, 强度和密度均衡适中, 可用于千亿级参数大模型, RNN 是按时间步递归循环串行传递序列的结构, 无法进入强智能相, 只能用于轻量低密的中小模型。

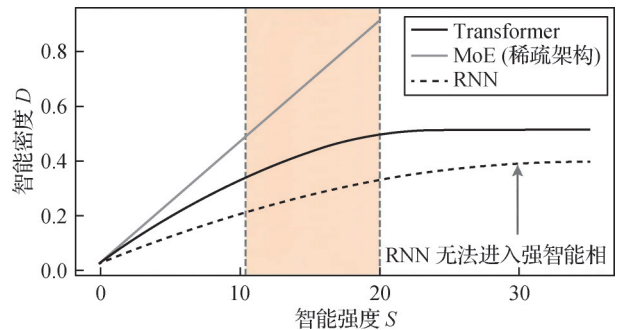


图2 不同智能形态的智能相空间 S - D 轨迹对比

Fig. 2 Comparison of S - D trajectories in the intelligent phase space across different intelligence forms

5) 行为域智能场的锯齿状相变: 通向通用 AI 的鸿沟

大模型 + Transformer 为代表的 4G-ANN 存在一个根本缺陷就是其智能涌现输出产生锯齿状相变, 智能场输出相变是指大模型在控制参数连续变化时, 输出行为发生突变、进入全新宏观相的现象, 不是渐变, 是阶跃式质变。锯齿状智能本质就是大模型在多模态、多任务空间里的“多临界点共存 + 跨模态跨任务相变跳变”, 是相变理论中的一种高度非均匀、非平衡和弱耦合的特殊形态, 也是造成离身智能和具身智能之间“鸿沟”本质原因。

在离身智能大语言模型中这种锯齿状的智能输出本质是对文本、图像和视频等多模态数据的训练和学习的不对称, 对多领域任务中专业知识理解的不平衡, 例如, 出现下棋、语言写作和编程等能力达到专家级的水平, 但在其他专门领域达不到基本专业水平, 甚至不具有儿童的视听觉的空间感和常识性的认知能力, 如房间中放置的一杯水, 儿童看到后知道是否能拿起来喝, 但大模型回答会过度谨慎, 说一堆“不确定”的“废话”, 不敢直接给结论, 有时被文字绕晕, 忽略最朴素的常识, 一本正经地“胡说”。大模型的锯齿状智能更不能用在具身智能要求稳定的智能输出的环境中, 具身智能(机器人、自动驾驶、无人机、操控手)有一条铁律: 控制回路必须连续、光滑、稳定、可微。而锯齿智能的本质是: 能力不

连续、输出跳变、阶跃突变、不可微,这两者天生互斥。例如,当前具身智能机器人的认知能力甚至还达不到2~3岁儿童搭乐高积木的泛化、想象和创造力水平,而行动力却能按预定程序做到某些专业技能(如机器臂操控)的能力。

总之,4G-ANN离输出均匀、平衡和稳定的通用智能还有很大距离,锯齿型输出明显任务执行中的能效也很低,在离身智能和具身智能之间还存在鸿沟,需要下一代的5G-ANN填补这一鸿沟。

6)“四要素说”和“统一场论”:人和机器认知的同源性和同一性

面向人工智能的基础理论问题,李德毅院士将人工智能的发展置于物理学、认知科学与文明演进的三重坐标系中加以审视。21世纪初首次提出认知物理学(cognitophysics)概念(李德毅等,2003),提出云模型、云变换和数据场等认知形式化理论,经过20多年的研究与实践,形成了认知物理学完整的基础理论框架(李德毅等,2023,2024a,b,c,2025),指出:智能或者智力只是人们相对于体力、体能的类比说法。人也好,认知机器也好,对一个特定问题认知产生的智能,其整体智能包括在认知空间的计算智能、记忆智能,也包括在物理空间中的感知智能和行为智能。世界的本质是物质、能量、结构和时间,因此,机器认知是由物质、能量、结构和时间4个基本要素组成(李德毅等,2023),智能会随着认知体物质、能量和结构的变化,随着时间推移,朝向最小能耗逼近。

人只能感知现实世界的三维空间加时间,更高维度空间真实存在,只是看不见,人类可通过微积分和线性代数(高维向量空间)等数学—物理方法将高维转换为三维空间加时间认知世界。从人类认知和机器认知的第一性原理出发看智能(李德毅等,2025),智能的量纲不是牛顿,也不是焦尔,它应该是无量纲的。人在完成一个个特定问题形成解决方案的过程中需要能量支撑,从人诞生的时间到当前时间 T 形成的生命智能,用智能方程表示为

$$I = \int_0^T f[M(t), S_h(t), S_s(t), E(t)] dt \quad (2)$$

式中, I 代表智能, t 代表时间, M 代表物质, S_h 代表硬构体, S_s 代表软构体, E 代表能量,其中物质(M)、能量(E)、结构(S)和时间(T)是人的认知和机器认知的4个基本要素, I 是前述的整体智能 Φ 加上时间 t

变量后持续输出的智能,用连续场 $\Phi(x)$ 描述系统有序程度,用“负熵”来衡量,这样,可用统一场论定义整体智能:智能场 $\Phi(x,t)$,智能强度场 $S(x,t)$ 、智能密度场 $D(x,t)$,因此,场论为相变提供了微观到宏观、物理空间到认知空间的统一语言,通过PDE作为物理约束计算出智能方程的解。例如,物理ANN可解麦克斯韦方程组(一阶PDE)的电动力学问题、薛定谔方程(二阶PDE)的量子力学问题。

“四要素说”是统揽人类认知和机器认知的第一性原理,用它可以更深入地理解事物之间的关联、物理规律和生命的复杂过程,解释智能的产生,实现智能的度量,尤其是理解人的认知和机器认知的同源性和同一性,本文提出神经网络结构决定智能场的相变,将物理的统一场论扩展到认知空间的数据场(输入)和智能场(输出)。李德毅院士断言:“人的智能和机器智能、物理同源、数学同构、时间同序”,堪称21世纪认知/智能机器发展的系统性宣言。

7)方法和结构

本文用溯源、分析、归纳、演绎及案例实证方法,从5个维度(模型、结构、第一性原理、费曼学习和实践)回答下一代ANN的核心特征、发展趋势和路径等问题。核心是遵循“本源锚定→历史归纳→规律演绎→趋势验证”的逻辑闭环,按时间线筛选“影响趋势的关键变量”,以智能和认知本质研究为前提,以ANN提出80年来跨世纪跨学科的具有标志性和里程碑性的人物、事件和研究成果为主线,包括2个纲领性报告、5篇AI领域最有影响力的论文、一项图灵奖和二项诺贝尔奖,以及20世纪创建的生物物理学和21世纪创建的认知物理学,综述20世纪中期到21世纪中叶五代ANN的核心特征、发展脉络,探索新一代ANN的研究方向和发展趋势。

1 20世纪的ANN

1.1 早期探索:追本溯源,研究发散,1G-ANN时代(20世纪40—70年代)

1.1.1 ANN萌芽时期—ANN诞生

20世纪40年代,受生物神经学和数学研究的启发,1943年,麦卡洛(McCulloch)和皮茨(Pitts)根据人脑神经元的工作原理提出人工神经元模型(M-P神经元模型)(McCulloch和Pitts,1943),1944年,给

出了基于数学的神经网络理论,他们描述的神经网络是一个单一神经元的数学模型(麦卡洛克—皮茨模型:McCulloch-Pitts model, 2018年被正式确立为生物物理学名词),具有阈值和权重,是ANN的原型,但它们不是逐层的结构,也没有指定任何特定的训练机制,只是一种概念性理论假设。在当时M-P模型也没有非自然生命的物理载体作为依托构成的人造神经组织,产生不了智能,无法验证。

此后,图灵机模型和冯诺依曼体系结构的电子数字计算机的发明,制造出了计算机。图灵(1948, 1950)给出了早期判定机器智能的测试方法和指标,称为图灵测试。后来,冯·诺依曼也对ANN进行了研究,ANN被考虑作为计算机器的组织结构的方案之一。

1.1.2 ANN雏形的形成(20世纪40—50年代)

通用数字计算机诞生后,图灵开始思考“机器能否表现智能行为”,为驳斥当时普遍的否定态度,分析其源于人类优越感、宗教观念、早期机械局限等偏见,1948年,图灵在英国国家物理实验室(national physical laboratory, NPL)内部做了一个《智能机器》(Turing, 1969)的报告,报告中提出了智能形态与智能判定方法。智能形态有二种:具身智能和离身智能。智能判定方法:用下棋场景验证机器智能测试(类图灵测试)。

报告的核心要点归纳有以下几点:1)智能本质论证:机器可展现智能行为,反对“人类专属智能”的偏见,是后来强人工智能可行性的基础。2)机器分类体系:提出离散/连续、控制/主动机器分类的计算机架构,是现代计算机体系结构的雏形。3)无组织机器模型:A型是与非门(NOT-AND gate, NAND)随机连接的最简网络;B型加入连接修改器,支持“干扰/教育”式权重调整,是后来联结主义和ANN的雏形。4)学习与优化机制:机器可通过经验修改自身指令实现学习,并提出遗传搜索思想优化机器结构,是后来强化学习、遗传算法的理论源头。5)智能测试与形态:提出图灵测试和多模态智能任务划分,其中A为早期模仿测试(类图灵测试),以博弈验证智能;B区分具身智能(机器人)与离身智能(博弈、翻译等)。6)哲学与工程视角:认为智能具有情感属性,需兼顾存储与效率的工程实现,为AI伦理、算力与算法协同设计提供支撑。

当时该报告未公开发表,直到1969年才得以公

开,是后来联结主义、强化学习的奠基文本。对后来AI发展影响最大的是图灵1950年发表在《Mind》上的论文《计算机器与智能》(Turing, 1950)。该文开篇就提出了振聋发聩的问题“机器能思维吗”,并给出了一个被称做“模仿游戏”的概念测试(属于离身智能范畴)。图灵还预言:到2000年,将会出现足够强大的计算机程序,能够在5分钟的文本对话中,让至少70%的人类评判者无法准确区分它与真实人类,从而通过他提出的“模仿游戏”,即后来被称为图灵测试的方法。《智能机器》报告堪称20世纪智能机器的系统性宣言,而论文《计算机器与智能》也承载着无数人初识智能本质的启蒙意义,成为AI历史上最有影响力的5篇论文之一。

时代的局限性:20世纪前,笛卡儿为代表西方哲学传统中身心二元对立论对科学研究影响很大;图灵1948年将智能形态划分为离身智能和具身智能;在论文(Turing, 1950)中统一称为计算智能,其智能能力的测试仍属于离身智能范畴,测试方法属于“黑箱”测试,忽视了具身智能,也忽视了记忆智能。后来,哲学家希尔勒用“中文屋”的思想实验(Searle, 1980)也对这种以纯符号、程序和语法产生的计算智能的判断提出了质疑。

1.1.3 “人工智能”诞生,1G-ANN成为联结主义的代表(20世纪50—70年代)

围绕智能机器怎么产生智能,1956年6到8月,在达特茅斯会议上,多个学科的十多位著名科学家进行为期8周的研讨,在概念上统一称为“人工智能”(简称AI),但对实现AI的研究路线有很大分歧,后来形成符号主义(逻辑程序和语义网络,属离身智能)、联结主义(感知机和认知机,是1G-ANN的代表)和行为主义(自适应控制机,属具身行为智能)三大学派。直到70年代,联结主义的ANN仍不是AI研究的主流,以下介绍几个对ANN早期发展有贡献的研究。

1)20世纪40年代末,现代计算机之父、博弈论之父冯·诺依曼就关注并开始研究计算机与人脑的关系,1948年9月在加州理工学院希克森研讨会“大脑行为机制”上做了《自动机的一般逻辑理论》演讲,后于1951年发表(von Neumann, 1951),首次系统比较计算机与神经系统,分析神经元的“混合特性”(轴突脉冲的数字性与突触传递的模拟性),探讨神经网络的并行处理与计算机串行处理的本质区别。之后

的研究还首次给出了图灵机与 MCP (McCulloch-Pitts neural network) 神经网络等价性的证明:神经网络原则上可以计算任何功能,就像任何数字计算机一样。可惜他 1957 年因病过早去世,没能继续他的这一研究,后人将他的这些研究成果分别写成《计算机与人脑》和《自复制自动机理论》二本著作,分别在 1958 和 1966 年出版,影响了更多来自神经生物学领域而非计算机科学的学者,认为有必要假设人类大脑可以被视作计算设备,提出了感知机(perceptron)和认知机(cognitron)的神经网络。

2) 感知机—感知器神经网络。1958 年,弗兰克·罗森布拉特根据 M-P 模型开发了感知器神经网络(Rosenblatt, 1958)。感知机是首个可学习二分类模型,掀起 ANN 的首波热潮,其模型层级化结构启发现代卷积神经网络设计。但 1969 年, Minsky 证明单层感知机无法解决异或(非线性)问题,也就是模型存在无法处理异或问题的局限性, ANN 又遇冷。但其线性阈值概念为深度学习技术所延续。

3) 认知机—神经认知机(neocognitron)。1975 年,日本学者福岛邦彦提出认知机自组织神经网络模型,是“认知机”这一技术模型的明确提出者,1980 年代进一步提出神经认知机(Fukushima, 1980),优化了认知机模型,解决视觉识别中的位移不变性问题,为卷积神经网络(convolutional neural network, CNN)奠定基础。Fukushima(1981)首次提出以“认知机”命名的类脑神经网络模型,模拟视觉认知过程。神经认知机是以生物物理和神经生理学为核心基础构建的模型,认知机的核心贡献在于将生物视觉系统的结构与功能,系统性地转化为可数学建模的神经网络架构,其强化“生物启发”的神经网络建模核心思路,形成从“生物神经机制中提取模型设计规则”的主流研究范式。对后来的 Hopfield 网络、脉冲神经网络(SNN)和卷积神经网络(CNN)都有启发。

1.2 中期奠基:理论奠基,研究收敛,2G-ANN/3G-ANN 时代(20 世纪 80—90 年代)

1.2.1 Hopfield 神经网络

1982 年,受生物物理学的启发,物理学家霍斯菲尔德提出 Hopfield 神经网络模型(递归联想记忆模型)(Hopfield, 1982),引入了“计算能量”概念。1984 年,又提出连续时间 Hopfield 神经网络模型(Hopfield, 1984),为神经网络和智能机器注入物理

理论根基,奠定了深度神经网络的生物和物理理论基础,成为后来获诺贝尔物理学奖的里程碑成果,也为联结主义确立了以生物物理学为基础的 ANN 研究导向,此后的 ANN 研究开始收敛。

1983 年,辛顿受 Hopfield 网络启发,与 Sejnowski 提出玻尔兹曼机雏形,用统计物理工具实现无监督学习(Ackley 等, 1985)。1986 年,辛顿等提出反向传播算法(BP)(Rumelhart 等, 1986)使多层网络的可训练性得到解决,该文是 AI 历史上最有影响力的 5 篇论文之一,推动联结主义从边缘成为 AI 主流,直接催生现代大模型时代。是 2G-ANN 的奠基成果。

1998 年, LeNet-5 (LeCun) 展示了首个实用卷积神经网络(CNN)在手写数字识别中的可用性。但当时的算力与数据规模不足,其整体效果在多任务上仍不及支持向量机(support vector machine, SVM)等方法, ANN 研究随后一度降温,但这一阶段(2G-ANN)为后续深度学习的爆发的和 4G-ANN 的形成奠定了关键的算法与架构基础。

1.2.2 脉冲神经网络(SNN)—3G-ANN(20 世纪 90 年代至今)

2G-ANN 是连续数值计算的神经网络,而脉冲神经网络(SNN)是一种类脑神经网络模型,它模仿生物神经元脉冲编码与传递机制。1997 年正式提出 SNN 概念(Maass, 1997),由于它不是 2G 的 CNN 架构,因此将 SNN 独立划分为 3G-ANN。

SNN 以脉冲序列而非连续数值作为信息载体,采用漏积分放电(leaky integrate-and-fire, LIF)简化神经元模型,神经元仅在膜电位超过阈值时发放脉冲,更贴近生物神经元的“全或无”放电机制。其核心特点与应用价值为:1)事件驱动:仅在脉冲发生时激活,计算稀疏,能耗远低于 ANN,适合边缘/低功耗场景。2)时序编码:天然处理动态/事件数据(如 DVS 相机、语音、动作序列),时间信息是核心载体。3)生物拟真:连接结构与学习规则更接近大脑皮层,为类脑智能与脑机接口提供路径。

但 SNN 提出以来一直处于研究阶段,直到近年在应用中才有突破,如在果蝇全脑模拟研究中, SNN 作为介观全脑神经连接的仿真工具近年取得重大进展, Dorkenwald 等(2024)首次在 10 万神经元级别实现果蝇全脑身体闭环仿真,超越线虫(302 神经元),进入复杂脑仿真时代,也证明生物具身仿真的可

行性。

2 21世纪的ANN(21世纪初—中叶)

2.1 AI进入爆发期:深度学习、大语言模型(21世纪初—20年代,进入了4G-ANN时代)

2.1.1 21世纪科技发展宣言:NBIC报告发布、人类认知组计划(HCP)的启动

刚进入21世纪,为制定21世纪科技纲领性文献,美国国家科学基金会(NSF)与美国商务部(united states department of commerce, DOC)汇聚了多个学科70多位科学家参与编写了《聚合四大科技,提高人类能力》报告,四大科技是指:纳米、生物、信息和认知,简称NBIC(nano-bio-info-cogno),2000年启动研究,2002年发布报告,2003年正式出版(Roco和Bainbridge,2003);NBIC报告奠定会聚理论框架,四大科技领域:纳米技术(物质基础)、生物技术(生命系统)、信息技术(数据处理)和认知科学(智能与脑机制)。会聚机制与路径:以纳米尺度物质统一性为基础,强调NBIC技术间接口与跨学科协同,将认知科学确定为融合先导,给出了“物理为基础—认知为先导”的科技发展路线:会聚逻辑:纳米制造→生物实现→信息控制→认知导向。科技会聚核心是认知科学为先导,NBIC四领域协同融合,重塑人类能力与社会发展。

NBIC报告还提出了人类认知组计划(human cognome project, HCP),将HCP列为21世纪前50年最高优先级工程,对标人类基因组计划(human genome project, HGP),目标是破译人类心智、认知、思维与意识的运行原理,通过NBIC(纳米、生物、信息、认知)技术聚合,全面提升人类感知、学习、记忆、决策与沟通能力。

2.1.2 人类认知组计划(HCP)研究计划

NBIC报告和HCP计划是纲领性文件,之后各国投入巨资制定了国家级的实施计划。2008年,在中国科学家李德毅、郑南宁院士等推动下,中国国家自然科学基金委员会(National Natural Science Foundation of China, NSFC)启动了“视听觉信息的认知计算”重大研究计划(2008—2015年),其学科交叉需求是整合计算机科学、神经科学、认知科学和机器学习等,解决感知特征提取、多模态融合和动态决策等跨领域难题;用无人驾驶车、脑机接口物理载体验证

和应用牵引,开展AI(具身智能)基础理论研究和关键技术的突破,推动了中国认知计算在具身智能(机器人)和脑机制(脑机接口)领域的发展。美国的脑计划BRAIN Initiative(2013)、欧盟的人脑计划Human Brain Project(2013)以及日本的脑、心智计划Brain/MINDS(2014)等相继启动。2018年,由中国科学家蒲慕明、骆清铭院士倡议发起“全脑介观神经联接图谱国际大科学计划”,2021年中国科技部正式设立科技创新2030—“脑科学与类脑研究”重大项目,这些计划中共同将脑结构图谱与神经连接机制研究推向战略制高点。确立了ANN在AI领域的主导地位。

2.1.3 为什么以纳米尺度统一物质结构

20世纪90年代初,著名物理学家杨振宁为《科技导报》撰写的《二十世纪的物理学》(杨振宁,1991)指出:我们一般把前几世纪的物理学称为宏观物理学,把二十世纪的物理学称为微观物理学。NBIC中物质基础:纳米尺度统一物质结构,是跨领域集成的共同起点的结论,是物理学经几个世纪发展得出的。纳米尺度物质结构决定了材料的性质,例如,AI算力之争的关键是芯片,而芯片的纳米极限之争,本质是物理极限、技术垄断、产业霸权和算力主权的四重博弈,是全球科技与国力竞争的核心战场;再如,纳米孔测序技术可实现单碱基分辨率,这一机制既依托纳米尺度的结构设计,如AlphaGenome推动了基因组学从“序列读取”向“功能解读”跨越。当前,量子技术的发展为纳米尺度的物质控制与功能实现注入新原理、新方法与新范式,是对纳米尺度为基础的深化与外延拓展。2025年图灵奖自1966年设立以来,首次将奖项授予量子信息科学领域的研究成果,该成果开创性地将量子力学原理与计算机科学、密码学深度融合。

2025年物理诺贝尔奖颁给“发现电路中的宏观量子力学隧穿效应和能量量子化”的3位量子物理学家,证明了量子世界的奇异特性(如隧穿效应、能量量子化)并非微观粒子专属,可在宏观系统中具象化。验证了从微观到宏观结构决定功能的规律:原子结构决定了分子级的功能;分子结构决定了细胞的功能;神经细胞的结构决定了神经网络的功能;神经网络的结构决定认知过程的涌现(智能场)。如引言所述,场论为智能相变提供了微观到宏观、物理空间到认知空间的统一语言。

传统脑研究长期局限于微观分子与宏观脑区尺度,对介于两者之间、决定大脑信息处理与认知功能的介观神经联接图谱缺乏系统性解析,而这一尺度正是理解神经元环路工作机制、揭示脑疾病发病机理、发展类脑计算模型的关键基础。随着高通量脑成像、单细胞标记、神经环路示踪与大数据分析技术的快速进步,跨物种、高精度、细胞类型特异性的介观连接图谱绘制已具备可行性,成为全球神经科学共同面临的重大科学工程。

2.1.4 认知作为融合先导的哲学和现象学基础:具身认知理论的建立

20世纪前的几个世纪以来,以哲学家、数学家笛卡尔(1596—1650)为代表的西方哲学传统中根深蒂固的身心对立二元论影响广泛,笛卡儿的“我思故我在”是身心二元对立的,主张思(心灵)主导,身体是机械,重在理性和科学。同时代,中国思想家、教育家、心学家王阳明(1472—1529)倡导的“知行合一”,主张认知与实践统一、道德与行动合一,身心一体不分,重在实践与道德。“知行合一”思想闪耀了人类智慧光辉,是人类高级生命体在知识、经验和元认知基础上,整合理性、情感、意识、道德和社会理解,形成的深层洞察、价值判断、长远决策与伦理抉择能力,只是对近代的科学领域影响较小。

20世纪中叶,法国哲学家梅洛—庞蒂(Maurice Merleau-Ponty, 1908—1961)对身心对立二元论进行了彻底反思,提出具身知觉一元论(概括一句是“我在故我思”),以现象学视角消解了身心二元对立,他在1945年撰写的《知觉现象学》(Merleau-Ponty, 2001)中系统阐述了身体作为认知媒介的核心思想。该理论强调心智、身体与环境不可分割,认知根植于身体与世界的互动体验中。通过与笛卡儿身心二元论的对比可以清晰地看到,近代以来西方占据主导地位的离身认知范式,本质上是身心二元论思维在认知科学与人工智能领域的延伸,将智能简化为无身体、无情境的符号运算,割裂了认知与生命、主体与世界的原初关联。而梅洛—庞蒂所揭示的具身知觉逻辑,则回归了认知的的生活世界本源,证明智能从一开始就是身体性、情境性与行动性的统一体。

在具身知觉一元论的视野下,具身智能与离身智能并非相互对立的两种范式,而是同一智能结构的基础层面与高阶层面,可统一到具身认知理论:具身智能是一切理性、思维与符号能力的发生学根基,

离身智能则是在具身经验基础上形成的抽象化、理论化能力。这一结论对人工智能研究具有重要价值:未来通用智能的实现,不能仅依赖算力的无限提升,而必须重建机器的身体性、知觉性与情境嵌入性,让智能重新回到与世界互动、与环境共生的真实结构之中,让有具身认知的机器人“智”、“体”均衡发展,并与人类的“德”对齐。梅洛—庞蒂的具身知觉一元论为当代认知科学领域的具身认知转向提供了根本性的哲学奠基与理论支撑,使具身认知从一种经验研究假说上升为具有坚实存在论基础的研究范式。

当今,中国科学家李德毅院士从第一性原理出发提出“四要素说”以及人工智能看哲学、教育、智慧和意识,相关系列论文汇编成《人工智能的边界—李德毅学术思想文选》,由武汉大学出版社出版(2026)。该书明确区分了智慧与智能,并给出它们的关系表示:智慧 = {意识, 欲望, 情感, 智能, ...}。人工智能的边界,在于将智能从人的自然生命体智慧中剥离出来,使之成为机器智能或人工智能;而机器输出的智能又必须与人类价值和伦理对齐。这方面的对齐任务不能由智能机器自行完成,而应由人类依据规则加以约束。人类文明的发展,应当让人发挥人的智慧,机器发挥机器的智能,各智其智,智人之智,智智与共。

2.1.5 认知作为融合先导的跨学科理论支撑:从生物物理学到认知物理学的跨越

杨振宁在《21世纪的物理学》中还提到:物理学的进步孕育了许多跨学科的研究,诸如天体物理学、生物物理学等。20世纪50年代初,DNA(deoxyribonucleic acid)双螺旋结构的发现,催生了物理学与生物学的结合,产生了生物物理学,这个新学科可解释生命和进化论。1958年,著名实验生物学家、细胞生物学家、教育家贝时璋先生创建了中国科学院生物物理研究所。1990年跨学科跨国的伟大工程“人类基因组计划(HGP)”启动:美国牵头,英、法、德、日随后加入,1999年中国参与(是唯一发展中国家),承担1%测序任务,2000年6月,完成任务“工作框架图”(覆盖97%基因组),2003年4月,完成任务精细图(覆盖92%),提前2年达标。这时美国国家科学院及时出版了《Bio2010:转变本科生教育以培养未来研究型生物学家》报告,倡导生命科学教育量化改革,旨在推动生物学与物理科学的深度融合,同

年,美国理论物理学家菲利普·纳尔逊教授总结20世纪中期以来物理—生物—信息”融合研究成果,出版了著作《生物物理学:能量、信息、生命》(Nelson, 2003),后改编为教材,成为经典的畅销教科书。

2001年开始,中国学者李德毅首次提出不确定性人工智能和认知物理学,开启了融合认知科学、物理学和信息技术的AI研究之先河。近年,又围绕“物理—生物—信息—认知”的融合路径撰写了认知物理学的核心论文,即机器认知三部曲系列论文(李德毅等,2024a,b,c)。认知物理学旨在解释进化论尚难充分解释的现象,提出了经验、推理、创造和发现4种认知模型,构成一个动态演化的“认知螺旋”。这一分类超越了传统AI三大学派(符号、连接、行为)的割裂,实现了认知层面真正的融合,并给出了认知机器的智能方程,以实现认知机器可解释的图灵测试。李德毅在撰写认知物理学著作时表示:“我更愿意写《认知物理学:物质、能量、结构、时间》,因为这是二十一世纪的物理学”。

从20世纪中叶生物物理学的创建到HGP计划的完成,标志着生命科学从“观察科学”走向“精确科学”的研究范式转向。21世纪的认知物理学也将推动HGP计划的实现,使AI研究从“不可解释”的图灵测试向“可解释”的图灵测试研究范式跃迁。

2.1.6 认知作为融合先导的具身认知论和生物神经学基础

具身认知论(Varela等,1991)是20世纪末以来认知科学领域兴起的重要研究范式,其核心主张在于从根本上突破传统认知主义的离身性预设,将身体、行动与环境置于认知活动的核心位置。该理论认为,认知并非发生于大脑内部的符号表征与逻辑计算过程,也并非可以脱离身体而独立存在的信息加工活动,而是根植于生命体的物理结构、感知运动系统以及与外部环境持续互动的动态过程。具身认知学的核心内容:人的神经系统包括大脑、小脑对应的神经系统(分属中枢神经系统)、身体各部分(包括视觉、听觉、触觉、嗅觉和味觉等)的神经(属于周围神经系统和脊髓神经系统)组成,具身认知过程是大脑、小脑通过脊髓神经通路紧密连接、协同调控机体全身的神经活动过程。遍布全身的神经网络是产生离身智能与具身智能的动力和源泉。婴幼儿的认知能力成长过程是先通过周边神经在不断感知环境时

学习和互动以改变和重塑全身神经结构,逐步增长大脑认知能力,而并非先有了成人大脑的结构和能力再成长的,其中,脊髓神经系统起到重要作用,它若损坏将影响大脑和小脑对身体周边神经的调控和身体的行为能力。

生物神经学用神经影像与电生理技术揭示大脑可塑性规律,为神经康复干预提供量化依据(Boyd等,2007),强调行为干预是驱动正向可塑性的关键。2010年后,生物神经医学、脑机接口(brain-computer interface,BCI)技术的进展对具身认知理论进行了科学验证。在“全脑介观神经联接图谱国际大科学计划”中首次提出在细胞类型分辨率层面系统绘制跨物种介观神经连接图谱(Zhong等,2021),填补从微观神经元到宏观脑区之间的关键认知空白,为揭示大脑信息处理、学习记忆和认知决策等高级功能的神经机制提供了前所未有的数据基础。最近,果蝇全脑连接组图谱测绘完成,并上传了“数字果蝇”(Shiu等,2024),该研究的核心价值在于突破了全脑仿真的技术瓶颈,验证了“结构即智能”的颠覆性猜想,为脑科学、人工智能等领域提供了全新研究范式。

2.2 ANN爆发期:通过对话式图灵测试,4G-ANN时代(21世纪初—20年代)

2.2.1 深度神经网络和深度学习

2006年,数据驱动、深度学习逐层预训练等思路在一定程度上缓解了深层网络训练困难。Hinton等提出的深度卷积神经网络AlexNet在2012年ImageNet图像分类竞赛中显著提升视觉识别性能,一举夺冠,并发表了这一研究成果的论文(Krizhevsky等,2012),是AI历史上5篇最有影响力的论文之一。

这一时期,图形处理器(graphics processing unit,GPU)算力与大规模数据共同推动深度学习进入快速发展阶段。随后,CNN、RNN、长短期记忆网络(long short-term memory network,LSTM)与生成对抗网络(generative adversarial network,GAN)等模型在视觉、语音与机器翻译等任务中取得系列突破。阿尔法围棋(AlphaGo)成为第一个战胜人类围棋世界冠军的人工智能程序,标志着深度强化学习在复杂决策任务上的能力跃升。2018年,Bengio、Hinton与LeCun因对深度学习贡献获得图灵奖(Turing Award)。

2.2.2 ChatGPT通过了图灵测试,ANN实现从学术模型向通用底座的跨越

2017年,在认知科学注意力机制的启发下,Google Brain团队发表了《Attention Is All You Need》(Vaswani等,2017)。其重大创新包括:1)完全抛弃RNN/CNN,只依靠注意力机制进行序列建模,提出全新的Transformer架构;2)引入自注意力和多头注意力,让每个位置都能直接关注所有位置,从而显著增强长距离依赖建模能力;3)通过位置编码,在无循环结构的情况下用正弦/余弦编码注入位置信息;4)建立编码器—解码器标准结构,奠定后来翻译、生成与大模型的基础骨架;5)支持高度并行计算,解决RNN必须逐步计算的问题,大幅提升训练速度,支撑起后来的大模型发展。该文也是AI历史上最有影响力的5篇论文之一,概括而言,就是“用自注意力取代循环,开启了整个大模型时代”。

2020年,OpenAI研发的生成式预训练Transformer 3(GPT-3)是大语言模型(LLM)发展的里程碑。该研究奠基性论文于2020年发表(Brown等,2020),首次验证了缩放定律(scaling laws)在超大规模语言模型中的上下文学习能力,开创了“预训练+提示”的全新范式,直接推动了生成式AI的爆发式发展,标志着ANN从“学术模型”向“通用底座”的跨越。这篇论文是AI历史上最有影响力的5篇论文之一。2022年,以ChatGPT为代表的大语言模型在开放域对话、知识问答与指令跟随等任务中显示出强大的语言生成与对齐能力,所涌现的智能通过了数亿计的对话式图灵测试,以生成式AI为代表推动“信息AI”进入快速迭代阶段,成为AI发展史上的里程碑事件。

然而GPT能力仍受到幻觉、可解释性不足、跨分布外推不稳和锯齿型智能等问题的制约,不能直接用到需要强物理约束与稳定、连续输出的闭环交互具身系统。

2.2.3 NBIC聚合发展,AI研究成果获诺贝尔物理学奖和化学奖,ANN成为主流。

2024年诺贝尔物理学奖授予John J. Hopfield与Geoffrey E. Hinton,奖励其“奠定了利用人工神经网络实现机器学习的基础性发现与发明”。Hopfield网络为第4代ANN提供了物理基础,反向传播算法(BP)解决了深度学习的核心训练问题;这一跨学科融合成果开创了物理学方法在信息处理和机器学习

中的应用,建立了兼具数学严谨性与生物启发性的计算框架。诺贝尔物理学委员会主席埃伦·穆恩斯评价:“霍普菲尔德和辛顿利用物理学方法寻找信息的特征,构建了为当今强大的机器学习技术奠定基础的神经网络框架”,这标志着物理启发的神经网络模型和深度学习获得了学术上的最高认可,AI研究路线进一步收敛到以物理主义为基础的ANN研究,成为当今AI研究的主流。

2024年诺贝尔化学奖一半授予大卫·贝克(David Baker),表彰其在计算蛋白质设计方面的贡献;另一半奖金由德米斯·哈萨比斯(Demis Hassabis)和约翰·江珀(John M. Jumper)共同获得,表彰其在蛋白质结构预测方面的突破性成就。其获奖理由是开发AI模型解决了50年科学难题,即从氨基酸序列预测蛋白质复杂三维结构,体现了计算方法与机器学习在结构生物学中的关键作用,AI推开了认知的大门,拓展人类探索分子生物尺度的边界,率先进入“生物AI”,也标志着开启了AI for sciences时代。

3 当今ANN前沿与下一代ANN发展方向的探索(21世纪20年代—中叶)

3.1 AI发展方向和路线的争论和分歧

4G-ANN是基于规模扩展缩放定律(scaling laws)的大语言模型及其结构。尽管其通过了对话式图灵测试,但仍属于“黑箱”测试,缺乏可解释性,局限于离身大脑智能的涌现,而具身机器人还缺少大脑神经网络的支撑,只能按预设程序完成规定动作,因此ANN在离身智能和具身智能之间出现了巨大鸿沟。根本原因在于,大模型既缺乏对现实世界物理规律的理解,又呈现出锯齿型智能涌现这两个重大局限。围绕下一代神经网络的发展方向和路线,也因此出现了很多争论和分歧。分歧大体分为两派:一派主张继续“数据拟合”、“大力出奇迹”,认为基于规模扩展定律的大模型仍有巨大发展空间,继续增加模型参数和算力是通往通用智能的道路;另一派则主张“结构重组”和“认知重构”,探索新一代类脑ANN模型和架构,让ANN走进现实世界,按物理定律提取现实世界事物背后的因果,追求极致的时间与能量效率。工程上,后者首先瞄准具身智能机器人技术的突破,代表方向有世界模型

(LeCun)、空间智能(李飞飞)和认知物理学(李德毅)等。以下各节将分析、比较并评述当前至21世纪中叶下一代ANN的研究前沿与走向。

3.2 物理—信息融合的 ANN — 4G-ANN 的物理信息增强版

如前述, 霍斯菲尔德提出的 Hopfield 神经网络模型为 ANN 注入了物理理论根基。近年, “物理 AI”(physical AI) 开始流行, 主要有两个研究方向:

1) 实体物理 AI, 即具身智能机器方向。物理人工智能(physical AI, PAI) 是创建能够执行智能生物体相关任务的实体系统, 常见于机器人、自动驾驶等场景, 强调机体、控制、形态、感知和动作的协同进化。最早由 Miriyev 和 Kovač(2020) 提出, 并首次正式定义了 PAI 这一新兴研究方向, 为解决数字“大脑”与物理“身体”协同脱节、导致机器人等实体系统难以灵活适配复杂多变物理环境并完成高精度交互任务的瓶颈, 提供了新的理论框架与研究范式。产业界则由英伟达推动“物理 AI”, 2025 年 1 月 CES 展会上, 英伟达 CEO 黄仁勋提出, 其定位是让 AI 从数字空间走向物理世界, 使自主系统能够感知、理解、推理并执行物理世界中的复杂任务。

2) 物理信息 AI, 即科学计算方向。物理信息 AI 的 ANN 核心是物理信息神经网络(physics-informed neural network, PINN)。该方向最早由 Raissi、Perdikaris、Karniadakis 在 2017 至 2019 年间系统性提出并命名(Raissi 等, 2019)。2017—2025 年, PINN 技术经历了快速发展。尤其在 2024—2025 年, 出现了 PirateNets 架构、自适应权重平衡方法以及大规模工业应用案例等多项重要技术突破, 这些进展显著提升了 PINN 的性能和实用性。PINN 由此成为科学机器学习(scientific machine learning, SciML) 中“数据驱动 + 物理约束”范式的代表性方法之一。作为 SciML 领域的重要突破, 它代表了深度学习与传统计算物理相结合的前沿方向, 也常被视为 AI for science 的核心技术支柱, 是 4G-ANN 的物理增强版。

3.3 动态可变的时间常数机制 — ANN 基础架构的创新

在 RNN 中, 时间常数(time constant) 不是动态可变的, 而是由权重和激活函数固定决定的。2020 年首次提出液态神经网络(liquid neural networks, LNN)(Hasani 等, 2021), 其核心创新在于动态可变的时间常数机制液态时间常数网络(liquid time-

constant network, LTC), 使模型能够自适应调节神经状态更新速率, 高效处理复杂时序数据。LTC 是液态神经网络(LNN) 的核心基础架构, 本质上是用常微分方程(ordinary differential equation, ODE) 建模神经元动态, 以自适应时间常数解决 RNN 梯度消失问题, 并与拉格朗日/哈密顿力学的数学框架高度契合, 奠定了 LNN 连续时间学习的基础。它是一种受生物神经系统启发的连续时间循环神经网络模型。受线虫神经回路启发, 研究又提出了神经回路策略(neural circuit policy, NCP) 架构, 将 LTC 与层次化神经元结构结合, 实现高可解释、低参数量, 并用 19 个神经元完成自动驾驶安全关键任务控制的科学验证(Lechner 等, 2020)。这标志着生命科学与物理科学研究范式的深度融合, 也表明 AI 从数据拟合迈向自主发现科学理论的“生物 AI”具有可行性。

3.4 世界模型和联合嵌入预测架构 — 面向 5G-ANN 的模型与架构设计

针对离身智能的生成式 AI 路线的局限, 例如像素级细节生成易拟合噪声、容易模糊且难以学到因果, LeCun(2022) 提出回到世界模型(world models) (鲍泓 等, 2025) 来看人的认知。其核心目标是让 AI 具备“类人认知”能力, 即理解外部世界的动态规律, 并预测未来情景, 实现“感知—预测—决策”的闭环, 打破当前 AI(如大语言模型)“懂符号、不懂物理世界”的认知瓶颈。世界模型不重细节, 而是学习抽象因果结构(如物理规则、物体关系、运动趋势), 以完成现实世界中的复杂任务; 其特点是高效、低算力和强泛化, 适配具身智能与机器人。其核心理念认为, 智能本质上是“预测行动后果 + 规划”, AI 需要拥有内部世界模拟器, 像人类婴儿一样通过感知与互动学习世界规律。

为实现世界模型, LeCun 团队采用联合嵌入预测架构(joint embedding predictive architecture, JEPA) (Assran 等, 2023), 核心采用的是视觉 Transformer(vision Transformer, ViT) 模型作为编码器主干网络。核心组件与网络: 上下文编码器 / 目标编码器: 用于提取图像 / 视频的抽象表示; 预测器: 轻量级 Transformer, 用于在表示空间做预测。通用场景: 也可兼容 CNN(如 LeNet、ResNet), 但主力是时空基础版 ViT(V-JEPA), 其他有图像基础版(I-JEPA)、认知进阶版(H-JEPA) 等。

JEPA 是通用自监督学习元架构, 不是单一网

络。其底层逻辑是:1)拒绝像素级生成,专注抽象世界表征学习;2)自监督为主,数据效率远超监督/生成式范式;3)建模物理世界的因果与时空规律,而非统计拟合;4)Transformer 为基座,兼容多模态下游任务。

JEPA 最大的技术挑战是表征崩溃:一是网络将所有输入映射到几乎相同的嵌入向量,丢失所有语义区分能力;二是嵌入被压缩到低维子空间,无法充分表达输入数据的复杂度。当前解决方案需根据具体应用场景,用针对性启发式方法(如调整掩码比例、优化 EMA 速率、引入对比损失),但缺乏系统性理论指导。另外,改变架构的方法,如 Meta 与纽约大学团队提出的“原生多模态 MoE 架构”是解决表征崩溃问题的尝试,该架构混合训练纯文本、图文对、原始视频和动作视频,搭配高维视觉表征,适配稀疏架构不崩塌。

3.5 认知螺旋模型与智痕元胞网络——“认知 AI”的理论与底层架构

在认知物理学理论中提出了认知机器的经验、推理、创造以及发现构成的“认知螺旋”模型,相应有 4 种认知模式,用这 4 个相对独立的认知模式完成认知的形式化,建立“认知螺旋模型”,实现物理空间和认知空间的映射和转化,提出面向 5G-ANN(认知神经网络)的架构,实现在认知层面的真正融合。

智痕元胞网络(smart trace cellular network, STCN)(李德毅等,2024c)是一种用于模拟人脑记忆与认知机制的思维软构体,是认知机器体系架构的核心组成部分,旨在解决传统计算机“孤立计算、忽视记忆”和“孤立思维、忽视具身”的局限性。智痕元胞是神经元细胞的镜像或指代。本质是模拟神经元细胞在记忆形成过程中留下的印迹、痕迹和残迹,智痕可深可浅,连接有强有弱,体现神经元和突触的可塑性。功能是作为记忆的基本单元,承载特定的概念、经验或知识片段。智痕元胞网络是由大量智痕元胞通过动态连接形成的复杂网络结构,记忆智能是其整体具象,而非单个元胞的功能体现。特性是随着知识增长不断演化,具有二次扩张和持久重塑能力,形成多元认知基础。定位是认知机器中记忆系统的核心实现方式,连接感知、交互与计算部件。

智痕元胞网络特性与结构有以下核心特征:1)复杂网络属性:具有小世界特征和无标度特性,连接度服从幂律分布;2)拓扑结构:包含重要路由节

点、平凡路由节点、边缘节点和骑墙节点(连接多个集群),呈现抱团特征、层次结构、自相似性和长尾分布;3)动态重构:通过学习不断微重构,修剪权重、改变连接,实现记忆调整与控制;4)记忆表达:用不同侧面、不同尺度上的整体拓扑状态表达记忆,提取方式为再认知和回忆;5)激活机制:长期记忆中多数元胞休眠,通过注意力(当前待解决问题)激活局部网络,带有不确定性。

STCN 具有低数据依赖、可解释认知、动态结构演化、类脑记忆与自主推理等特征,突破了传统静态网络的局限,更接近人类认知模式,符合 5G-ANN 的核心特征,为类脑认知智能提供了重要理论与基础架构支撑。

3.6 具身认知物理神经网络(E-CoPNN)——面向 5G-ANN 的轻量化架构

作者团队在“驾驶脑认知技术与平台”等基金项目支持下,经过多年的研究和实践,提出了一种具身认知物理神经网络(E-CoPNN)架构,依据大脑神经、小脑神经和身体周围神经的具身协同机制,数据与认知先验(语义概念)双驱动,增加人类概念注意力机制、选择性机制、快慢双系统以及内插和外推相结合,提高模型的可解释性、泛化性、自纠错、持续学习和精准推理能力,拓扑结构可自适应动态组织,按类脑认知规律设计的多层嵌套式学习,改进提高连续学习、持续学习和强化学习,提高学习和推理的效率,是一种面向 5G-ANN 的认知神经网络的紧凑型架构设计,如图 3 所示。

图 3 中,3 层嵌套循环反馈反映类脑具身认知过程,是遍及具身分布式神经系统,其中各功能区和作用是:大脑思维机制:借助云大脑通过“蒸馏+剪枝”,实现模型的极致轻量化与高性能平衡进行;小脑协调机制:神经网络和路由+注意力的负载均衡器;脊髓神经用套接(socket)机制:实现核心交换机与骨干通路的双工连接,如快慢系统的切换。周围神经的感觉机制:选择性机制+I/O 滤波器,实现多模态、稀疏多任务自适应变换。

1)数据形态与流程。数据场中的数据形态主要分为两种:欧氏数据和非欧数据,欧氏数据包括图像像素、特征图和坐标点等,其特点是规则、对称、均匀、平移不变,但人类视觉对纯欧氏空间的像素和坐标变化不敏感;非欧数据包括结构、关系、拓扑、语义关联以及目标间依赖,其特点是物体的形状和它们

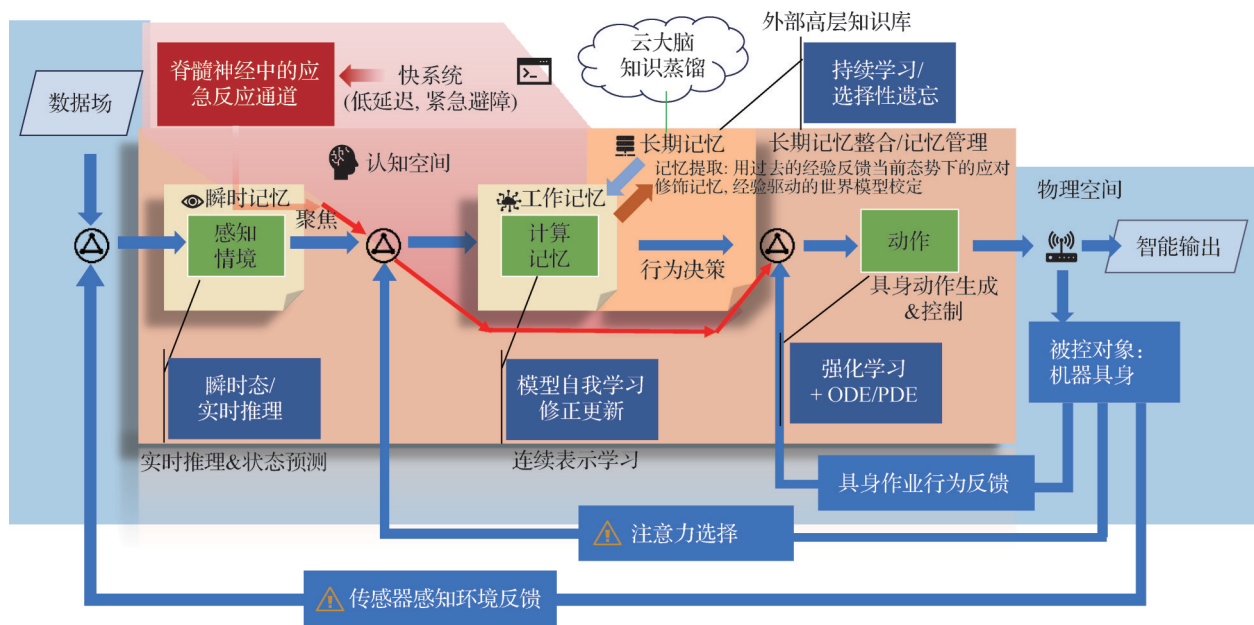


图3 具身认知物理神经网络 E-CoPNN 架构

Fig. 3 Architecture of the embodied cognitive physical neural network E-CoPNN

之间的关系(如车—行人—信号灯/标识牌)、场景拓扑结构(如人体和物体的骨架结构)、交互以及依赖,就是说人类视觉对非欧数据在语义层级、注意力上依存。需要用 socket 机制,实现语义级的双向交互。数据流全程闭环:数据输入→渠道适配→网关校验→决策规划→动作执行→结果返回→记忆持久化,实现稳定、可靠和可追溯的自动化执行。

2)选择性机制。选择性机制是注意力的“过滤器”;Transformer架构中的多头注意力依靠全局数据的并行运算机制让大算力涌现,但能效比极低,而人脑神经的选择性机制是进化形成的“筛选+抑制”机制,即筛选要关注的信息,抑制无关信息,选择性机制+瞬时记忆提高了对敏感信息的反应精准性和速度,也节省神经网的能耗。因此,在具身认知轻量化的网络中,选择性机制是神经网络高效运行的保障机制。本架构中神经回路中承担选择性机制的是滤波器、智能路由、网关和交换机,在系统内部和外部交互和反馈的数据和指令须经过滤波、路由和网关,保证安全与状态一致。

3)人类概念注意力机制。属于层次注意力,数据的层次分为像素、词元和概念3层。例如交通场景中,交通警示标识、行人横穿等危险状况,人类驾驶员只关注概念层的语义信息。在类人识别非欧数据方法上用单样本(one-shot)视觉概念学习,人类可从1个示例标识学会新概念并泛化,而传统 ML/DL

需数千/数万样本,可采用类人概念注意力机制和贝叶斯程序学习(Bayesian program learning, BPL),在给定少量示例时,用贝叶斯后验推断找到最可能解释观测数据的程序(即概念)(Lake等,2015),类人的层级认知先验:内置跨字符/跨字母表的通用先验(如标识和行人中笔画和动作顺序、空间关系),实现知识迁移。例如,在自动驾驶领域(Zheng等,2021),可以从“眼动+注意力”升级成“人类注意力=非欧结构感知→用图神经网络(GNN)/图结构建模”,以此实现类人自动驾驶。

4)可交互。多层嵌套学习,最里层(动作)有具身行为的动力学反馈回路:周围神经—本体感觉,用 ODE/PDE、强化学习约束;中间层(决策)有人类视觉听觉注意力的反馈回路,实现大脑、小脑神经高级认知“感知—思维—行动”慢系统过程(用连续时间递归 LTC 机制),工作记忆+推理+知识蒸馏;最外层(感知)有传感器的感知环境的大循环反馈回路:其中有瞬时记忆+快速推理小循环双向通道,用事件驱动 LIF 机制配合类脑芯片实现脊髓神经系统的快速反应通路,以提高实时性,实现危险场景的“感知—行动”的快速处理,称为快系统。使机器可在认知空间内部和与物理空间外部进行交互。

5)自成长。3种记忆类型:有瞬时记忆、工作记忆和长期记忆的转换,有感知和决策层的双推理,有搜索相关事实和知识的引擎,有行为的决策,更有对

记忆的修改和对记忆的快速提取,通过纠正错误的负反馈回路对齐使命目标,使机器不断积累知识能够自成长。

6)会学习。采用连续学习、持续学习、强化学习和渐进式学习机制,解决目前连续学习中瞬时记忆阶段推理缺失的增量更新问题(Parisi等,2019)、持续学习中的“灾难性遗忘”问题(French,1999),以及强化学习中传统算法“泛化能力差”等局限,实现深度神经网络与强化学习的有效融合,验证深度强化学习在复杂任务中实现人类级控制的可行性(Mnih等,2015),并已在自动驾驶中得到应用(潘峰和鲍泓,2021),用洋葱模型的认知核结构(鲍泓等,2025)模拟幼儿成长过程的渐进式学习,渐进式学习模拟人类认知发育的循序渐进特性,是实现类人智能的重要路径。

7)自纠错。E-CoPNN用3层嵌套学习和动态局部反向传播(Local BP)架构,区别于传统“训练—部署”分离和全局反向传播传播(Global BP)的静态模式,解决传统ANN架构中误差累计、性能差和“灾难性遗忘”的瓶颈问题。这是一种嵌套递归自我纠错与改进机制,Xia等人(2017)提出双阶段递归修正架构,为ANN递归自我纠错的可行性提供了可落地的技术范式。本文核心是通过多层递归反馈循环,让网络自主检测各层输出误差、定位错误根源,并动态调整自身参数(权重、结构等),实现性能的持续优化。

E-CoPNN具有动态拓扑(3层嵌套、两套系统)、记忆认知(3种记忆类型)、神经符号融合(欧氏与非欧数据融合,词元与概念结合)、跨形态基础模型(离身与具身结合,自动驾驶与具身智能统一)等特征,具有自适应结构、小样本泛化、可解释、低功耗和具身认知等5G-ANN典型特征。该架构支持具身认知一元理论下的跨离身和具身的基础模型,已有研究团队进行了量化验证,如郝孝帅(HAO X S, et al. 2025)及其团队研发的跨具身视觉语言基础模型MiMo-Embodied是首个统一具身与自动驾驶的开源模型,有三个模块:ViT视觉编码器+MLP投影器+7B大语言模型。输入:支持单图、多图、视频;输出:感知→推理→规划→执行全链路。用一套模型、一套权重,并融入大语言模型,在自动驾驶12项基准与具身智能17项基准两类场景均达到SOTA,验证统一模型的双向正向迁移能力,并完全开源。与E-

CoPNN设计思想一致,实现跨具身基座模型的自动驾驶与机器人统一表征。推动5G-ANN从数据拟合走向结构重构与具身认知一体化。

4 展望与总结(21世纪20—50年代)

4.1 展望

21世纪20—50年代是人类认知组计划/提出的分阶段实施路线图的第3阶段(2020—2050年),这一时期也将是4G-ANN向5G-ANN过渡发展的阶段。此阶段的ANN发展主要体现在以下几个领域。

1)20世纪具身智能与离身智能相互独立的研究范式跃迁到21世纪的具身认知基础层面与高阶层面统一研究范式。离身与具身互补发展,离身智能领域要补“身体”,具身智能领域要补“大脑”,充分发展“小脑”和“脊髓”神经的调控功能。脑成像、脑机接口技术的突破,支撑对大脑、小脑和身体神经系统的协同机理和统一机制的研究和应用;全脑介观神经联接图谱国际大科学计划到21世纪中叶,分3阶段在细胞类型分辨率层面系统绘制跨物种(果蝇、小鼠、灵长类动物)介观神经连接图谱,继果蝇全脑连接组图谱测绘完成,最近一项以小鼠、猕猴等模式生物的视觉皮层生理数据为验证,提出的视觉皮层的紧凑型深度神经网络模型(Cowley等,2026)研究,揭示视觉皮层“高效信息处理”的核心机制,推翻传统观念,证明极小、可解释的模型同样能达到极高精度,验证了“结构即智能”的生物学机制,打破“唯大模型论”,为类脑智能及认知神经网络(5G-ANN)的紧凑型架构设计提供重要参考,其理论意义是重塑了认知与框架,也为进一步揭示具身全脑高级功能的神经机制提供了数据基础,骆清铭指出:构建人脑图谱的终极目标在于揭示其宏大而精密的网络结构与工作原理。

2)网络底层逻辑的数学基础不断拓展。近期研究表明,大脑几何学以微分拓扑、高维统计几何与流形学习为数学基础,揭示了生物智能依赖结构编码而非数据拟合的核心机制(Wakhloo等,2026)。这一理论为理解智能的泛化、鲁棒性与高效表征提供了底层范式,为新一代“结构即智能”的结构重构研究奠定了关键的数学支撑。

3)不同粒度和尺度的网络结合优势互补,ANN感知和处理数据的粒度从像素级、词元级到概念级。

神经网络尺度上,微观尺度:结合量子神经网、分子生物神经网等优势,使计算能力更高;介观尺度:类脑科学和全脑介观神经联接图谱研究为探索智能本质提供了新路径,对推动新一代人工智能、新型计算架构与脑机接口技术发展具有重要理论价值与应用前景;宏观尺度:认知为导向聚合“物理—生物—信息—认知”四大科技,实现对认知本质的研究的目标。到21世纪中叶(2050年前后),通过四大技术聚合,信息AI、物理AI和生物AI融合向结构重构的认知AI模式跃迁,实现人类身体、认知、感官、社交与创造能力的全面增强,使人类摆脱生理与智能的自然限制,进入能力自主提升、社会高效协同以及生命质量极大改善的文明新阶段。实现21世纪科技愿景、延展人类认知能力,迎接新的认知的革命。

4.2 结语

在具身知觉一元论的视野下,具身智能与离身智能并非相互对立的两种范式,而是同一智能结构的基础层面与高阶层面。具身知觉一元论不仅完成了对传统身心二元论哲学的超越,也为21世纪认知科学和人工智能的具身转向奠定了坚实的哲学基础;基础理论实现了从20世纪生物物理学到21世纪认知物理学的跃进,具身认知神经网络将填补ANN发展中离身智能与具身智能之间的鸿沟,推动研究范式从数据拟合转向结构重构,为5G-ANN、脑科学、通用AI等领域提供全新的研究范式,并支撑4G-ANN向5G-ANN的跃迁。

人工智能与其他领域一样,也是有认知边界的,人工智能的边界是科学约束与哲学差异的叠加;科学上受数学、物理、生物、计算、数据与工程的硬核限制;哲学上无法跨越意识与主体性、具身性与意义、创造与伦理、认知与元认知的鸿沟。认清这些边界,能避免人工智能“万能论”、“威胁论”等,更理性地发展与应用人工智能,让技术服务于人类价值与福祉。

致谢:本文的形成与撰写受益于与李德毅院士及认知物理学研究团队成员多年来的研讨和交流,在此表示感谢。

参考文献(References)

Ackley D H, Hinton G E and Sejnowski T J. 1985. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(1): 147-169

- Assran M, Duval Q, Misra I, Bojanowski P, Vincent P, Rabbat M, et al. 2023. Self-supervised learning from images with a joint-embedding predictive architecture [EB/OL]. [2026-02-28]. <https://arxiv.org/pdf/2301.08243.pdf>
- Bao H, Zheng Y and Liang T J. 2025. Research progress and trends on models and structures of cognitive machines. *Journal of Image and Graphics*, 30(4): 895-921 (鲍泓, 郑颖, 梁天骄. 2025. 认知机器的模型与结构研究进展. *中国图象图形学报*, 30(4): 895-921) [DOI: 10.11834/jig.240108]
- Boyd L A, Vidoni E D and Daly J J. 2007. Answering the call: the influence of neuroimaging and electrophysiological evidence on rehabilitation. *Physical Therapy* 87 (6): 684-703 [DOI: 10.2522/ptj.20060164]
- Brown T B, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. 2020. Language models are few-shot learners//*Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates, Inc.: #159
- Cowley B R, Stan P L, Pillow J W, and Smith M A. 2026. Compact deep neural network models of the visual cortex. *Nature*, 652: 947-954 [DOI: 10.1038/s41586-026-10150-1]
- Dorkenwald S, Matsliah A, Sterling A R, Schlegel P, Yu S C, Mckellar C E, et al. 2024. Neuronal wiring diagram of an adult brain. *Nature*, 634(8032): 124-138 [DOI: 10.1038/s41586-024-07558-y]
- French R M. 1999. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3 (4): 128-135 [DOI: 10.1016/S1364-6613(99)01294-2]
- Fukushima K. 1980. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4): 193-202 [DOI: 10.1007/BF00344251]
- Fukushima K. 1981. Cognitron: a self-organizing multilayered neural network. *NHK Technical Monograph*, (30): #25P
- Hasani R, Lechner M, Amini A, Rus D and Grosu R. 2021. Liquid time-constant networks//*Proceedings of 2021 AAAI Conference on Artificial Intelligence*. [s.l.]: AAAI: #16936 [DOI: 10.1609/aaai.v35i9.16936]
- Hopfield J J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8): 2554-2558 [DOI: 10.1073/pnas.79.8.2554]
- Hopfield J J. 1984. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 81(10): 3088-3092 [DOI: 10.1073/pnas.81.10.3088]
- Kaplan J, McCandlish S, Henighan T, Brown T B, Chess B, Child R, et al. 2020. Scaling laws for neural language models [EB/OL]. [2026-02-28]. <https://arxiv.org/pdf/2001.08361.pdf>
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//*Proceedings of the*

- 26th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA: Curran Associates, Inc.: 1097-1105
- Lake B M, Salakhutdinov R and Tenenbaum J B. 2015. Human-level concept learning through probabilistic program induction. *Science*, 350(6266): 1332-1338 [DOI: 10.1126/science.aab3050]
- Lechner M, Hasani R, Amini A, Henzinger T A, Rus D and Grosu R. 2020. Neural circuit policies enabling auditable autonomy. *Nature Machine Intelligence*, 2(10): 642-652 [DOI: 10.1038/s42256-020-00237-3]
- LeCun Y. 2022. A Path Towards Autonomous Machine Intelligence, Position Paper, Version 0.9.2
- Li D Y. 2023. Cognitive physics—the enlightenment by Schrödinger, Turing, and Wiener and beyond. *Intelligent Computing*, 2: #0009
- Li D Y, Gan W Y and Liu L Y. 2003. Artificial intelligence and cognitive physics//Proceedings of the 10th National Academic Annual Conference of the Chinese Association for Artificial Intelligence. Guangzhou, China: Chinese Society for Artificial Intelligence (李德毅, 涂文燕, 刘璐莹. 2003. 人工智能与认知物理学//中国人工智能学会第10届全国学术年会论文集. 广州: 中国人工智能学会)
- Li D Y, Liu Y C, Bao H and Jiang S. 2025. Understanding Albert Einstein The “Four Elements Theory” is the first principle that governs human and machine cognition. *CAAI Transactions on Intelligent Systems*, 20(4): 1046-1052 (李德毅, 刘玉超, 鲍泓, 蒋升. 2025. 读懂爱因斯坦“四要素说”是统揽人和机器认知的第一性原理. *智能系统学报*, 20(4): 1046-1052) [DOI: 10.11992/tis.202506028]
- Li D Y, Liu Y C and Yin J L. 2024a. How cognitive machines make create. *China Basic Science*, 26(6): 1-11 (李德毅, 刘玉超, 殷嘉伦. 2024a. 认知机器如何创造. *中国基础科学*, 26(6): 1-11) [DOI: 10.3969/j.issn.1009-2412.2024.06.001]
- Li D Y, Yin J L, Zhang T L, Han W and Bao H. 2023. Four most basic elements in machine cognition. *China Basic Science*, 25(3): 1-10, 22 (李德毅, 殷嘉伦, 张天雷, 韩威, 鲍泓. 2023. 机器认知四要素说. *中国基础科学*, 25(3): 1-10, 22) [DOI: 10.3969/j.issn.1009-2412.2023.03.001]
- Li D Y, Zhang T L, Han W, Hai D, Bao H and Gao H B. 2024b. Structure and activation of cognitive machines. *CAAI Transactions on Intelligent Systems*, 19(6): 1604-1613 (李德毅, 张天雷, 韩威, 海丹, 鲍泓, 高洪波. 2024b. 认知机器的结构和激活. *智能系统学报*, 19(6): 1604-1613) [DOI: 10.11992/tis.202409024]
- Li D Y, Zheng S Y, Huang L W and Liu Y C. 2024c. Cognition formalization. *China Basic Science*, 26(2): 1-14, 28 (李德毅, 郑思仪, 黄立威, 刘玉超. 2024c. 认知的形式化. *中国基础科学*, 26(2): 1-14, 28) [DOI: 10.3969/j.issn.1009-2412.2024.02.001]
- Maass W. 1997. Networks of spiking neurons: the third generation of neural network models. *Neural Networks*, 10(9): 1659-1671 [DOI: 10.1016/s0893-6080(9700011-7)]
- McCulloch W S and Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4): 115-133 [DOI: 10.1007/BF02478259]
- Merleau-Ponty M. 2001. *Phénoménologie de la Perception*. Jiang Z H, trans. Beijing: The Commercial Press (莫里斯·梅洛—庞蒂. 2001. 知觉现象学. 姜志辉, 译. 北京: 商务印书馆)
- Miriyev A and Kovač M. 2020. Skills for physical artificial intelligence. *Nature Machine Intelligence*, 2(11): 658-660 [DOI: 10.1038/s42256-020-00258-y]
- Mnih V, Kavukcuoglu K, Silver D, Rusu A A, Veness J, Bellemare M G, et al. 2015. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529-533 [DOI: 10.1038/nature14236]
- Nelson P. 2003. *Biological Physics: Energy, Information, Life*. New York: W. H. Freeman
- Pan F and Bao H. 2021. Research progress of automatic driving control technology based on reinforcement learning. *Journal of Image and Graphics*, 26(1): 28-35 (潘峰, 鲍泓. 2021. 强化学习的自动驾驶控制技术研究进展. *中国图象图形学报*, 26(1): 28-35) [DOI: 10.11834/jig.200428]
- Parisi G I, Kemker R, Part J L, Kanan C and Wermter S. 2019. Continual lifelong learning with neural networks: a review. *Neural Networks*, 113: 54-71 [DOI: 10.1016/j.neunet.2019.01.012]
- Raissi M, Perdikaris P and Karniadakis G E. 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378: 686-707 [DOI: 10.1016/j.jcp.2018.10.045]
- Roco M C and Bainbridge W S. 2003. *Converging Technologies for Improving Human Performance*. Dordrecht, the Netherlands: Springer [DOI: 10.1007/978-94-017-0359-8]
- Rosenblatt F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6): 386-408 [DOI: 10.1037/h0042519]
- Rumelhart D E, Hinton G E and Williams R J. 1986. Learning representations by back-propagating errors. *Nature*, 323(6088): 533-536 [DOI: 10.1038/323533a0]
- Samsonovich A V, Shumsky S A, Karpov V E, Kotov A A and Kolonin A G. 2022. Key advanced research initiative: a manifesto for the new-generation artificial intelligence. *Procedia Computer Science*, 213: 824-831
- Searle J R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3): 417-424 [DOI: 10.1017/S0140525X00005756]
- Shiu P K, Sterne G R, Spiller N, Franconville R, Sandoval A and Zhou J. 2024. A Drosophila computational brain model reveals sensorimotor processing. *Nature*, 634: 210-219 [DOI: 10.1038/s41586-024-07763-9]
- Turing A M. 1950. Computing machinery and intelligence. *Mind*, LIX(236): 433-460 [DOI: 10.1093/mind/LIX.236.433]
- Turing A M. 1969. *Intelligent Machinery*//Meltzer B and Michie D, eds.

- Machine Intelligence. Edinburgh, UK: Edinburgh University: 3-23
- Varela F J, Thompson E and Rosch E. 1991. The Embodied Mind: Cognitive Science and Human Experience. Cambridge, USA: MIT Press
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates, Inc.: 6000-6010
- von Neumann J. 1951. The general and logical theory of automata//Jeffress L A, ed. Cerebral Mechanisms in Behavior: The Hixon Symposium. New York, USA: John Wiley and Sons, Inc.: 1-41
- Wakhloo A J, Slatton W and Chung S Y. 2026. Neural population geometry and optimal coding of tasks with shared latent structure. Nature Neuroscience, 29 (3) : 682-692 [DOI: 10.1038/s41593-025-02183-y]
- Xia Y C, Tian F, Wu L J, Lin J X, Qin T, Yu N H, et al. 2017. Deliberation networks: sequence generation beyond one-pass decoding//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 1782-1792
- Yang Z N. 1991. Physics in the 20th century. Yan Z X, trans. Science and Technology Review, (2): 12-16 (杨振宁. 1991. 二十世纪的物理学. 严志雄, 译. 科技导报, (2): 12-16)
- Zheng Y, Bao H, Meng C C and Ma N. 2021. A method of Traffic police detection based on attention mechanism in natural scene. Neuro-computing, 458: 592-601 [DOI: 10.1016/j.neucom.2019.12.144]
- Zhong Q, Li A, Jin R, Zhang D, Li X, Jia X, et al. 2021. High-definition imaging using line-illumination modulation microscopy. Nature Methods, 18 (3) : 309-315 [DOI: 10.1038/s41592-021-01074-x]

作者简介

鲍泓,男,教授,主要研究方向为视觉信息的认知计算和智能驾驶。E-mail:baohong@bnu.edu.cn

梁天骄,通信作者,男,博士研究生,主要研究方向为机器视觉和智能驾驶。E-mail:liangtianjiao23@mails.ucas.ac.cn

郑颖,女,博士,讲师,主要研究方向为图神经网络和智能驾驶。E-mail:jqrzhengyin@bnu.edu.cn