

对未来 20 年智能体分化和演进的 若干思考

——从算力与智能的角度

刘宇航

中国科学院计算技术研究所

引言

人工智能(AI)的发展进程在当前处于一个重要的历史节点。到 2025 年,人工智能这一概念从 1956 年达特茅斯会议被提出已历经近 70 年。现阶段 AI 大模型是以高算力、大数据为基础的。规模定律(scaling law)也被称为大模型的尺度定律,揭示了大语言模型的模型性能与其参数量、训练数据集大小,以及训练资源之间存在着一种可预测的关系,即投入资源越多,模型规模越大,最终的模型效果也就可能越好。

当前 AI 一方面在大模型等技术方向取得显著进展,一方面面临一些亟待解决的科学问题。这些问题包括:1)基础设施能量消耗高、算力(computing utility, CU)耗费大,规模定律尽管取得一定成效,但难以持续。2)智能水平(intelligence level, IL)的本质尚未充分揭示, AI 远远没有达到人脑那样高的能量效率, AI 的基础理论尚为薄弱。3)包括半导体工艺、开源处理器芯片、新型材料、量子计算、超导计算、神经形态计算、存算一体技术在内的各种传统和新兴技术受到关注,但在潜在增益和现实性等方面表现各异。4)我国面临芯片高端制造工艺被限制的问题,尚需要一段时间实现自主自立。需要认真审视当前流行的“高端制造工艺等于高算力,而高算力等于高智能”观点。我国产业界一方面面临“算力焦虑”问题,一方面面临算力利用不充分不均衡的问题。我国 AI 发展在未来 20 年

可能和应该遵循何种路径,是一个值得思考和研究的问题。

费林分类法在 AI 时代的作用有限。指令与数据是计算机系统的 2 个基本概念,1967 年费林(Flynn)根据指令流和数据流的并行性高低将计算机系统分为 4 类:单指令流单数据流(single instruction single data, SISD)、单指令流多数据流(single instruction multiple data, SIMD)、多指令流单数据流(multiple instruction single data, MISD)、多指令流多数据流(multiple instruction multiple data, MIMD)^[1]。这种分类法简洁、完备,对于学术界理解和设计不同种类的计算机系统的微体系结构产生了重要的积极影响,但对 AI 时代的计算机系统分类作用有限。在 AI 时代,研究者和实践者关注的焦点不再是指令和并行度的高低,而是算力水平和智能水平的高低。

就像指令与数据是计算机系统的 2 个基本要素一样,计算与智能是智能体(Agent)的 2 个基本要素。智能是否等同于计算?高智能是否一定需要高算力?智能的本质是什么?各种新兴技术对未来计算机系统或智能体形态具有怎样的影响?诸如此类,这些重要问题的答案尚未被完全揭示。在 AI 已有较大发展但规模定律难以延续的时代,为深化对这些问题的研究,并预判未来的智能体类型的分布,本文从计算与智能角度构建智能体的分类学(taxonomy of Agents),并思考未来 20 年 4 种可能分类形态的演化,提出构建智能体演化动力学(evolutionary dynamics of Agents)的设想。

与费林分类法不同,本文讨论的“智能高低”与“算

DOI: 10.11991/cccc.202603010

通信作者:刘宇航, E-mail: liuyuhang@ict.ac.cn

力高低”的分类是从算力和智能的角度出发。比较这2种分类法,可以发现:1)分类目的不同。费林分类法描述计算机硬件架构中的并行处理能力,帮助理解不同类型的计算机如何同时执行指令和处理数据;算力与智能分类法旨在评估智能体的算力水平和智能水平,为不同应用场景选择合适的智能体提供指导。2)适用范围不同。费林分类法侧重于硬件层面的分类,不直接涉及软件或算法的表现;算力与智能分类法适用于涉及智能体的应用场景,不仅考虑硬件性能,还结合了软件、算法以及用户体验等因素。

本文就“算力—智能”的关系提出以下观点:CU是计算机系统在实际资源约束下可被用户感受的计算效用;IL是智能体面向目标的达成能力。智能并不等同于计算,但可由计算实现;两者以智算效率 $\eta = \frac{n_{IL}}{n_{CU}}$ 耦合,其中, n_{CU} 是算力水平的归一化度量(normalized computing utility), n_{IL} 是智能水平的归一化度量(normalized intelligence level)。在此立场下,本文的具体贡献是:1)提出可操作的 n_{IL} 与 n_{CU} 度量与聚合方法;2)以 (n_{CU}, n_{IL}) 四象限建立智能体分类;3)以收益驱动的动力系统描述四类智能体份额随时间的演化,并给出场景化条件与可复现实验设定。该框架使得本文在提出问题的同时给出可检验的答案与方法路径。

算力和智能的水平度量

度量算力和智能的水平是理解每个智能体类别特性的基础。

算力水平度量

度量算力的水平需要考量各种类型关于计算速度的指标。包括但不限于:1)每秒浮点运算次数,这是衡量计算机性能最常用的指标之一,尤其是在科学计算、图形渲染等领域;2)每秒整数运算次数,对于某些类型的应用,如加密算法或数据库查询,整数运算性能可能更为重要。这些指标既有合理性,亦有局限性,学术界正在研究统一的度量指标^[2-4]。

定义1(算力水平) 设计计算机系统的资源—性能向量为

$$c = (\text{吞吐}\tau, \text{时延}\lambda, \text{功率}P, \text{能耗/样本}\varepsilon, \text{存储}M, \text{带宽}B, \text{可用性}A)$$

按任务域权重 w 归一化得到算力水平

$$n_{CU} = g(c; w) \in [0, 1]$$

式中 g 为单调聚合如加权调和/几何平均,体现在给定成本、能耗、带宽与可用性约束下的可用计算效用。

智能水平度量

度量智能水平需要考量的指标^[5]包括但不限于:

1)任务成功率,即智能体完成特定任务的成功率,例如图像识别的准确率、自然语言处理的正确解析率等。通过测试数据集评估模型的表现,并计算平均精度、召回率等指标。2)学习效率,即系统学习新知识的速度和效果。这包括从给定的数据集中提取有效信息的能力以及泛化到未见过的数据上的能力,需要观察训练时间、收敛速度、过拟合程度等。3)自适应性和灵活性,即系统应对变化环境或未知情况的能力,设计复杂的、动态变化的任务场景以评估系统的反应能力和创新解决问题的能力。4)复杂性处理能力,即系统处理复杂问题或多重任务的能力。这涉及理解上下文、推理、规划等方面,使用包含多层次逻辑推理、长期依赖关系等问题的数据集进行测试。

定义2(智能水平) 取指标簇 $M = \{\text{任务成功率, 样本效率, 泛化迁移, 自适应/在线学习, 鲁棒安全, 推理—规划深度}\}$ 。归一化后以加权几何平均聚合:

$$n_{IL} = \left(\prod_{m \in M} m^{\alpha_m} \right)^{1/\sum \alpha_m} \in [0, 1]$$

式中 α_m 为对应指标 m 的权重,反映在特定评估体系中不同能力维度的重要性差异。

定义3(高/低阈值) 在指定任务域与资源约束下,采用分位阈值:记样本集的 n_{CU} 第50分位(即中位数)为 θ_c^{mid} 、 n_{IL} 第50分位为 θ_I^{mid} ;样本满足相应阈值,即判定为高/低。阈值与权重在评测过程中公开透明固定,避免事后调参。

定义4(智算效率) 定义 $\eta = \frac{n_{IL}}{n_{CU}}$ 。当比较不同智能体实现时, η 直接衡量单位算力带来的智能水平提升。

智能体分类学

在顶层设计AI技术发展路径时,常常需要探析算力与智能的高低水平的组合。如图1所示,本文从算

力与智能的角度, 基于 (n_{CU}, n_{IL}) 二元表述, 将智能体划分为4类: 低算力高智能 (low computing-utility high intelligence, LCHI)、低算力低智能 (low computing-utility low intelligence, LCLI)、高算力低智能 (high computing-utility low intelligence, HCLI)、高算力高智能 (high computing-utility high-intelligence, HCHI)。该分类旨在服务于任务域内的智能体方案选择与路线图设计, 而非仅以硬件并行度区分(与费林分类法的目的与适用范围不同)。

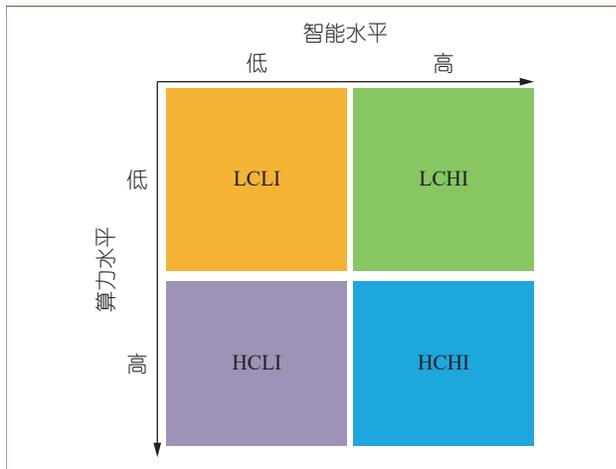


图1 从算力和智能2个维度对智能体的分类

低算力高智能 (LCHI) 智能体

定义(判定规则) 对于给定的智能体, 若 $n_{CU} \leq \theta_C^{mid}$ 且 $n_{IL} \geq \theta_I^{mid}$, 则将其归类为 LCHI 智能体。

特点 1)采用高效能算法: 采用轻量化模型、剪枝和量化等, 在压缩原有模型大小的同时, 不显著损失性能。2)采用优化的架构设计: 如专用集成电路(ASIC)和现场可编程门阵列(FPGA)等, 实现高效能计算。

应用场景 1)边缘计算: 边缘计算设备上的轻量化 AI 模型, 尽管硬件资源有限, 但在特定任务上有较高的准确性和自适应性, 如自动驾驶和智能家居设备中的语音识别、图像分析等功能直接在本地完成, 减少对云端计算资源的依赖, 提高响应速度和服务质量。2)移动设备: 智能手机、平板电脑等便携式设备中嵌入 AI 功能, 实现实时翻译、拍照增强等。3)工业自动化: 小型机器人和自动化生产线通过内置传感器和控制器独立完成复杂任务, 降低整体系统的能耗和成本。

优势 LCHI 智能体在能效、隐私、低时延等边缘/移动/离线受限环境中优势尤为显著。1)节能高效: 较

低的功耗有助于延长电池寿命, 减少能源消耗。2)保护隐私: 数据处理发生在本地, 减少了敏感信息上传至云端的风险。3)快速响应: 缩短了网络延迟, 提高了用户体验。

低算力低智能 (LCLI) 智能体

定义(判定规则) 对于给定的智能体, 若 $n_{CU} < \theta_C^{mid}$ 且 $n_{IL} < \theta_I^{mid}$, 则将其归类为 LCLI 智能体。

特点 1)规则简单或逻辑预设: 主要基于预先设定的规则或有限的学习能力来执行特定任务, 缺乏复杂的推理和适应性。2)功能硬件固定: 通常由专用芯片或微控制器实现, 针对特定应用进行了优化。3)数据处理能力有限: 能够处理相对简单的输入输出关系, 但无法应对复杂的决策过程。

应用场景 1)简单的嵌入式系统: 如家电产品的定时器、温度控制器等, 这些设备只需要执行基本的条件判断和控制操作, 既不需要大量计算也不涉及复杂智能。2)传感器网络: 用于环境监测、健康追踪的小型传感器节点, 仅需要传输简单的测量值。3)消费电子产品: 某些低成本电子玩具、健身手环等, 提供有限的交互功能。

优势 1)成本低廉: 由于使用较为简单的硬件和软件, 制造成本较低, 易于大规模推广。2)稳定性高: 固定的逻辑结构使得系统运行更加稳定可靠, 不易受到外界干扰影响。

发展趋势 1)集成度提升: 将更多的功能集成到单个芯片上, 进一步降低成本和体积。2)模块化设计: 开发标准化的模块组件, 方便用户根据需要灵活组合。

高算力低智能 (HCLI) 智能体

定义(判定规则) 对于给定的智能体, 若 $n_{CU} \geq \theta_C^{mid}$ 且 $n_{IL} \leq \theta_I^{mid}$, 则将其归类为 HCLI 智能体。

特点 1)大量计算资源: 利用高性能中央处理器(CPU)、图形处理器(GPU)集群进行大规模数据处理, 但主要用于执行重复性的计算任务而非复杂的智能行为。2)以规则为导向: 尽管拥有强大的计算能力, 但其决策过程仍然依赖于明确编写的程序代码或固定算法, 不具备自主学习和适应环境变化的能力。3)采用批处理作业: 常见于科学计算、金融建模等领域, 用于解决需要大量数值运算的问题。

应用场景 1)科学仿真: 气候预测、分子动力学模

拟等需要大量计算资源的任务,但最终结果是由确定性算法得出,不涉及智能决策。2) **数据分析**: 大型数据库查询、统计分析等工作,虽然可能涉及复杂的数学运算,但本质上是遵循既定逻辑的计算过程。3) **图形渲染**: 电影特效制作、视频游戏开发中的实时渲染,需要大量计算资源来生成高质量视觉效果,但并不包含智能元素。4) **视频转码**: 高性能服务器用于批量处理视频转码任务,虽然具备强大的算力,但任务本身相对简单,不需要复杂的决策过程。

优势 1) **精确度高**: 对于已知问题,可以提供非常精确的答案或解决方案。2) **可预测性强**: 因为是基于固定的规则和算法,所以每次运行的结果都具有较高的可预测性和一致性。

发展趋势 1) **优化分布式计算**: 改进算法和架构,使计算任务能够在多个节点间更有效地分配。2) **普及云计算**: 借助云服务平台提供的弹性资源,降低企业自建数据中心的成本和技术门槛。3) **倡导绿色计算**: 研发更节能的硬件和软件,减少数据中心碳足迹。

高算力高智能 (HCHI) 智能体

定义(判定规则) 对于给定的智能体,若 $n_{cu} \geq \theta_C^{mid}$ 且 $n_{il} \geq \theta_I^{mid}$, 则将其归类为 HCHI 智能体。

特点 1) **智能技术先进**: 包括深度学习、强化学习在内的多种先进技术,能够处理复杂的感知、认知、规划等任务。2) **算力基础设施强大**: 依赖于超级计算机、云计算平台提供的海量计算资源,支持大规模模型训练和复杂场景下的实时推理。3) **应用领域广泛**: 从自然语言处理、图像识别到自动驾驶、医疗诊断等多个领域都有广泛应用。

应用场景 1) **自动驾驶汽车**: 集成多种传感器数据,实时感知周围环境并做出驾驶决策。2) **智能客服系统**: 通过自然语言理解与生成技术,为用户提供个性化服务和支持。3) **医疗影像分析**: 辅助医生进行疾病诊断,提高准确率和效率。4) **超级计算机上运行的大规模深度学习模型**: 具有极高的浮点算力和优秀的任务成功率,在图像识别、语音识别等领域表现出色。

优势 1) **灵活性强**: 可以根据不同情况进行自适应调整,具备较强的泛化能力和迁移学习能力。2) **创新潜力大**: 不断推动新技术的研发和应用,为各行各业带来革命性的变革。

发展趋势 1) **通用 AI 探索**: 追求构建能够广泛适

用的智能体。2) **多模态融合深化**: 整合文本、图像、音频等多种信息源,提供更加全面的服务体验。

若干智能体实例的分类

基于上述分类法,作为示例,本文分析 2 种重要的智能体的类别。

车载智能体的类别分析

车载智能体可以根据其算力和智能水平归类为 4 类中的某一类,具体取决于该智能体的硬件性能及其支持的应用程序功能。

高算力高智能车载智能体

特点 1) **处理能力强大**: 配备高性能 CPU、GPU 或专用 AI 加速器,如 NVIDIA Drive PX 系列; 2) **智能化功能先进**: 支持复杂的自动驾驶算法、环境感知、路径规划、决策制定等; 3) **多传感器融合**: 能够处理来自摄像头、雷达、超声波传感器等多种传感器的数据; 4) **实时数据处理**: 具备快速响应时间,确保驾驶安全性和系统可靠性。

应用场景 1) 自动驾驶汽车(Level 3 及以上); 2) 高级驾驶辅助系统(advanced driver assistance systems, ADAS),如自动紧急制动、自适应巡航控制、车道保持辅助等; 3) 智能交通管理,如车联网(V2X 通信)。

高算力低智能车载智能体

特点 1) **处理能力较强**: 具有较高的计算性能,但主要用于执行特定任务而非复杂决策; 2) **智能化功能有限**: 可以进行基本的数据处理和简单的自动化操作,但缺乏高级推理和学习能力。

应用场景 1) 中级驾驶辅助系统(如 Level 2 ADAS); 2) 车载娱乐系统和信息系统; 3) 数据记录仪(黑匣子)。

低算力高智能车载智能体

特点 1) **算力水平较低**: 虽然硬件资源相对有限,但通过优化算法和轻量化模型实现较高智能化水平; 2) **智能化功能先进**: 能够在本地执行一些复杂的任务,如语音识别、手势控制、场景理解等。

应用场景 1) 基础级别的驾驶辅助功能(如 Level 1 ADAS); 2) 车内个人助手(如语音导航、音乐播放控制); 3) 特定任务的边缘计算节点。

低算力低智能车载智能体

特点 1)有限的处理能力:仅能满足最基本的任务需求,通常不具备强大的计算资源;2)基础功能:主要用于执行预编程的固定功能,很少涉及动态决策或学习。

应用场景 1)简单的车载娱乐设备(如 CD/DVD 播放器);2)基本仪表盘显示器;3)传统的车载控制系统(如空调控制、窗户升降)。

无人机智能体的类别分析

无人机智能体的类型取决于无人机的应用场景、性能需求和技术实现。

高算力高智能无人机智能体

特点 1)强大的处理能力:配备高性能 CPU、GPU 或专用 AI 加速器;2)先进的智能化功能:支持复杂的感知与规划算法,如深度学习模型、即时定位与地图构建(simultaneous localization and mapping, SLAM)、路径规划等;3)多传感器融合:能够处理来自摄像头、雷达、激光雷达(light detection and ranging, LiDAR)、惯性测量单元(inertial measurement unit, IMU)等多种传感器的数据;4)实时数据处理:具备快速响应时间,确保飞行安全性和任务可靠性。

应用场景 1)高级自主飞行:如完全自主导航、避障、跟随目标等。2)复杂环境下的任务执行:如搜索救援、农业监测、基础设施检查等。3)实时数据分析:如现场视频流分析、地形测绘、3D 建模等。

实例 1)NVIDIA 专为自主机器、边缘 AI 和嵌入式计算打造的高性能计算平台 NVIDIA Jetson 系列(Jetson Xavier NX、Jetson AGX Orin);2)边缘 AI 视觉处理器 Intel Movidius Myriad X VPU;3)高集成度基于智能手机系统芯片的无人机/机器人高通 Snapdragon Flight。

高算力低智能无人机智能体

特点 1)较强的处理能力:具有较高的计算性能,但主要用于执行特定任务而非复杂决策。2)有限的智能化功能:可以进行基本的数据处理和简单的自动化操作,如预设航线导航、图像采集等。

应用场景 1)中级自主飞行:如基于 GPS 的自动返航、预编程的飞行路径。2)固定任务执行:如定期巡逻、航拍任务等。

实例 1)微型单板计算机 Raspberry Pi 4 Model B;

2)面向工业与高性能嵌入式的推理平台 BeagleBone AI。

低算力高智能无人机智能体

特点 1)中等或较低的处理能力:虽然硬件资源相对有限,但通过优化算法和轻量化模型实现较高智能化水平。2)先进的智能化功能:能够在本地执行一些复杂的任务,如物体识别、语音控制等。

应用场景 1)基础级别的自主飞行:如简单的避障、自动降落等。2)特定任务的边缘计算节点:如小型农业无人机上的作物健康监测。

实例 1)STM32 系列微控制器(带机器学习扩展库);2)ESP32(带外部协处理器)。

低算力低智能无人机智能体

特点 1)仅能满足最基本的任务需求,通常不具备强大的计算资源。2)智能化功能基础:主要用于执行预编程的固定功能,如遥控飞行、简单图像采集等。

应用场景 1)玩具级无人机:如室内娱乐用的小型无人机。2)基本任务执行:如简单的航拍、手动操控飞行。

实例 1)Arduino 系列(如 Arduino Mega);2)微型飞行控制器(如 Naze32)。

智能体分类的演化动力学

智能体分类的演化动力学研究 4 类智能体的市场占比分布随着时间变化的演化过程,特点是量化、动态、综合,而战略科学家囿于有限的数据和简单的模型,往往只能进行定性、静态、单一的思考,因此前者是后者的有效支撑和补充。

演化动力学在其他领域已有一些进展。例如,中国科学院院士、武汉大学教授夏军带领团队研发了长江模拟器^[6],为长江的综合治理和认识过去、现在、未来的长江提供科技支撑。但据调研,目前在 AI 领域,还没有从算力和智能的角度对智能体的发展轨迹和趋势进行系统分析的演化动力学研究。

本文建议研究“智能体分类演化动力学”,并开发相应的模拟器,对未来不同尺度的时间范围进行模拟,量化预测发展趋势,发挥“模拟—评估—决策支持”功能。在模拟器中可以考量半导体工艺、开源芯片(如 RISC-V 开源处理器)、材料(如碳纳米管)、量子计算(虽潜力较大,但何时能成熟可用尚有较大不确定性)、

神经形态计算、存算一体技术等因素对智能体分类分布的影响,以指导 AI 发展路径的顶层设计。

2025—2035 年这 4 种智能体的市场占比的演变

预测未来 10 年内这 4 种智能体的市场占比变化,需要考虑技术进步、市场需求等多方面的因素。

在技术进步与成本下降方面,具有 2 个特点。1)高算力硬件变得更加普及:随着半导体技术的进步(如更先进的制程工艺),高性能计算资源的成本将逐渐降低,使得更多的设备能够配备强大的处理器和加速器。这意味着高算力的设备可能会更加普遍。2)AI 模型优化与轻量化:研究者们正在不断探索如何让智能体模型变得更小、更快、更节能,这将促进低算力高智能设备的发展。例如,微型机器学习(TinyML)和其他边缘 AI 技术可以使更多资源受限的设备具备智能化功能。

在市场需求与应用场景扩展方面,具有 2 个特点。1)物联网和边缘计算:随着智能家居、智慧城市、工业 4.0 等领域的发展,低算力高智能计算机系统的需求预计会大幅增长。这类设备能够在本地进行数据处理,减少延迟并提高隐私保护。2)云服务与数据中心:高算力低智能计算机系统在数据中心内的应用将继续保持重要地位,尤其是在视频转码、大规模数据分析等方面。同时,云端 AI 服务也会推动高算力高智能系统的使用。

在成本效益与能效比方面,具有 2 个特点。1)能源效率的重要性增加:全球对可持续发展的关注不断提高,能效比成为选择智能体的重要考量因素之一。因此,低算力高智能设备由于其较低的能耗,在移动设备、便携式电子产品中将占据更大的市场份额。2)性价比驱动市场选择:尽管高算力高智能设备提供了较高的性能,但它们的成本也相对较高。对于某些不需要极致性能的应用场景,用户可能会倾向于选择性价比更高的低算力高智能解决方案。

基于上述分析,可以推测 2025—2035 年这 4 种智能体的市场占比可能会发生如下变化:1)低算力高智能设备显著增加。随着 AI 技术的进步和边缘计算的发展,这类设备将在物联网、移动设备等领域得到广泛应用。2)低算力低智能设备逐渐减少。虽然一些简单任务仍然可以通过这类设备完成,但由于智能化需求

的增长,部分市场可能会被低算力高智能设备所取代。3)高算力低智能设备维持一定市场占比。在数据中心、科学计算等领域,高算力低智能设备将继续扮演重要角色,但在其他领域可能面临来自高算力高智能设备的竞争。4)高算力高智能设备稳步增长。随着 AI 应用的深入和高性能计算资源成本的降低,这类系统将在科研、医疗、金融等行业获得更广泛的应用。

2035—2045 年这 4 种智能体的市场占比的演变

预测 2035—2045 年这 4 种智能体的市场占比变化,需要考虑更长远的技术发展,以下是对这些因素的深入分析及对市场占比变化的推测。

在技术进步与突破方面,具有 3 个特点。1)量子计算与神经形态计算:如果量子计算机和神经形态计算机在未来 20 年实现商业化应用,它们可能会彻底改变计算的范式。特别是神经形态计算,旨在模仿人脑的工作方式,可能极大地提升高智能计算机系统的性能。2)AI 模型与硬件协同进化:随着智能体算法的进步和专用硬件,如张量处理器(TPU)、神经网络处理器(NPU)的发展,低算力高智能智能体将变得更加普遍和高效。可能会有更多轻量化但功能强大的 AI 模型应用于各种场景。3)新材料与新架构:新型半导体材料(如碳纳米管)和非冯·诺依曼架构(如存算一体芯片)有望进一步提高计算效率,降低能耗,使得低算力高智能设备在资源受限环境中更加实用。

在市场需求与应用场景扩展方面,具有 3 个特点。1)全面智能化:到 2043 年,几乎所有类型的设备都将具备一定程度的智能化。从日常消费品到工业设备,智能系统将成为标配。这意味着低算力高智能和高算力高智能类型智能体的市场占比会显著增加。2)边缘计算与分布式智能:随着 5G/6G 网络的普及和物联网的成熟,边缘计算和分布式智能将变得至关重要。低算力高智能类型智能体将在本地处理数据,提供即时响应,减少对云端的依赖。3)个性化与定制化服务:消费者对个性化体验的需求将持续增长,推动企业开发更多高智能产品和服务,如虚拟助手、健康监测器等。

在成本效益与能效比方面,具有 2 个特点。1)绿色计算与可持续发展:全球对环境问题的关注将促使智能体向更节能的方向发展。因此,低算力高智能设

备由于其较低的能耗,在移动设备、便携式电子产品中将继续占据重要地位。2)高性能计算的成本下降:高算力高智能系统的成本可能会大幅降低,使其不再局限于科研机构或大型企业,而是进入更多的商业和个人应用领域。

基于上述分析,可以推测 2035—2045 年这 4 种智能体的市场占比可能发生如下变化: 1)低算力高智能设备大幅增加。这类设备将在智能家居、可穿戴设备、医疗保健等领域广泛部署,成为日常生活不可或缺的一部分。随着技术的进步,即使是小型设备也能执行复杂的任务。2)低算力低智能设备显著减少。虽然一些简单任务仍然可以通过这类设备完成,但由于智能化需求的增长和技术进步,大部分市场可能会被低算力高智能设备所取代。3)高算力低智能设备逐渐减少。随着高算力高智能系统的成本下降和应用范围扩大,高算力低智能设备在数据中心之外的应用可能会减少。然而,在特定领域(如大规模数据处理、视频转码)中,它们仍会保持一定市场占比。4)高算力高智能设备大幅增加。这类系统不仅会在科研、医疗、金融等行业获得广泛应用,还可能进入更多的消费级市场,如自动驾驶汽车、智能家居中枢等。随着技术进步,普通用户也能够负担得起高性能的智能设备。

基于复制子动力学演化模型的数值模拟预测

在前两节定性分析的基础上,为了以动态视角定量地刻画群体策略的演化规律,我们建立智能体的复制子动力学(replicator dynamics)演化模型,编写相应程序进行数值模拟预测。复制子动力学是演化博弈论的核心动态模型,其精髓可概括为:群体中不同策略的比例变化受“适应性差异”驱动——当某个策略的收益高于群体平均水平时,其占比将增长;反之则衰减。这一“优胜劣汰”过程使系统自发向更稳定状态演进,并自然淘汰劣势策略。该模型揭示了即使个体不具备完全理性,群体行为仍可通过竞争与复制机制收敛至均衡,从而为生物进化、社会习俗及技术传播等宏观现象提供统一的演化分析框架。

以 2025 年 10 月为起点,记 4 类智能体在时间 t 的份额向量 $\mathbf{x}(t) = [x_{LCHI}, x_{LCLI}, x_{HCLI}, x_{HCHI}]$, $\sum x_i = 1$; 设收益 $\pi_i(t)$ 由能耗/电价、硬件成本、带宽/隐私合规、时延与可靠性、技术冲击(工艺/架构/算法/材料)等驱动;离

散化为按情景切换的马尔可夫转移;收益函数由能价、带宽价、隐私/合规惩罚、时延惩罚、可靠性奖励、技术冲击构成。

考虑以下 3 种场景。场景 A: 能价不断上升、隐私/时延权重不断下降、“边缘优先”技术冲击; 场景 B: 能价不断下降、可靠性不断上升、“云优先”技术冲击; 场景 C: 带宽价不断上升、隐私/时延中高、“云-端均衡”技术冲击。

如图 2 所示,在能价、隐私与时延权重均较高且技术冲击偏向端侧的场景下,LCHI 智能体的份额随时间单调上升,在 $t \approx 75$ (月)时跨越 0.5,并在 $t \approx 150$ (月)后逼近稳态; LCLI 智能体早期小幅上行后转而下降,HCLI/HCHI 智能体自始即衰减。最终份额为: LCHI 智能体 99.998%、LCLI 智能体 0.0017%、HCLI 智能体接近 0%、HCHI 智能体接近 0%,显示以能效/隐私/低时延为导向的端侧高效率方案在该场景下具备演化优势与稳定占优。

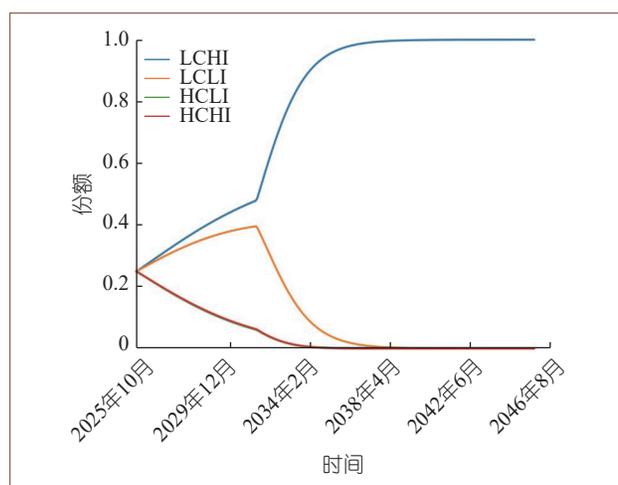


图2 “边缘优先、高能价、强隐私”场景下的智能体份额演化

如图 3 所示,尽管该场景对数据中心侧的可追溯性、可再现批处理更友好(技术冲击偏云侧、可靠性奖励更高、能价更低),仿真中 LCHI 智能体仍凭借效率进步维持长期优势: LCLI 智能体形成一期“驼峰”后回落,HCHI/HCLI 难以反超。最终份额为: LCHI 智能体 98.22%、LCLI 智能体 1.73%、HCHI 智能体 0.05%、HCLI 智能体接近 0%。这说明若端侧/小模型的技术进步(或“效率红利”)持续,即使能价走低且可追溯性重要,LCHI 智能体仍可能主导;若希望体现 HCLI 智能体的稳定存在,须进一步提高“可靠性/审计”收益权

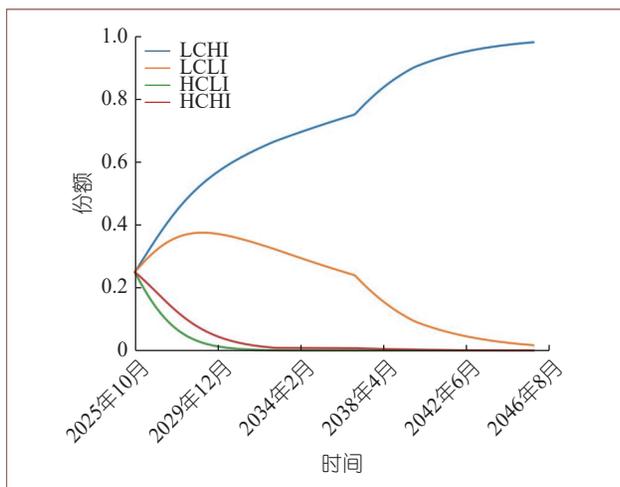


图3 “低能价、大吞吐批处理、高可靠性”场景下的智能体份额演化

重、放缓 LCHI 智能体技术冲击。

如图4所示，在云-端协同且带宽受限场景下，LCLI 智能体早期占比上升(以较低数据搬运成本适配带宽约束)，但随着端侧效率与能力提升，LCHI 智能体在 $t \approx 110 \sim 120$ (月)反超并进入稳态主导。最终份额为：LCHI 智能体 98.57%、LCLI 智能体 1.43%、HCLI 智能体接近 0%、HCHI 智能体接近 0%。结果表明：带宽受限+端侧预处理的组合促使高效率端侧智能体长期受益，而大规模中心化方案在此场景受限更明显。

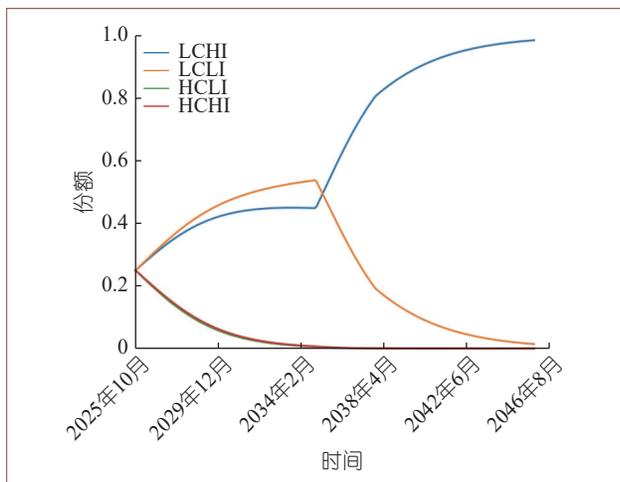


图4 云-端协同且带宽受限场景下的智能体份额演化

发展建议

以分类学和演化动力学为抓手，加强 AI 发展路径的顶层设计。顶层设计是以实现中国式现代化为目

标，利用可用的有限资源，使 AI 对经济社会发展的支撑作用最大化。立足我国国情和实际需求，充分考量半导体工艺、开源芯片、新型材料、量子计算、超导计算、神经形态计算、存算一体技术等各种传统和新兴技术对智能体分类分布的影响，以指导 AI 发展路径的顶层设计。当前的 AI 尽管取得了一定的进展，但未必是最终的解决方案，多条路径都有成功的可能，但都存在一定的试错成本。在资源允许的范围内，应宽容和鼓励百家争鸣、百舸争流，避免孤注一掷、错失机遇。

对各国近现代科技特别是信息技术发展路径进行正反两方面的总结。美国信息高速公路成功的原因、欧洲蓝脑计划的得失、苏联计算机发展失败的教训、日本第五代机计划失败的教训、我国光刻机技术初期尚可但之后中断研究的教训，等等，需要系统地研究和总结，以为镜鉴。钱学森对我国智能计算机的研制曾给予深入的思考^[7]，有很多历久弥新的宝贵观点，需要当代 AI 研究者特别是路径顶层设计者认真参考。

培育 AI 发展的成熟生态和创新文化。不断提升我国开源软硬件的成熟度，避免同质化竞争和重复“造轮子”，加强共性技术的开源共享，降低创新的门槛，把我国有限的科研资源用到更能发挥效用的地方^[8]。以近期在大模型领域取得显著进展的 DeepSeek 公司为例，其成功离不开开源机制带来的高起点创新基础，离不开诸如在多头潜注意力(multi-head latent attention, MLA)的核心技术，离不开多版本的快速迭代，离不开我国培养的数十位经过严格学术训练的年轻人才，离不开领军人才的技术判断力和知人善任的领导力，这些要素相互作用形成创新合力。高显示度的突破性成果往往只是冰山在水面之上的极小部分，没有冰山之下的创新所需的生态和文化，突破性成果会成为无源之水、无本之木，因此需要坚持问题导向以克服“四唯”^[9]，加强集成以克服分散，倡导开源共享以克服重复，坚持长期主义以克服急功近利，这些都应纳入创新文化的体系。

结束语

AI 是人类的非凡创造，是实现中国式现代化的重要支撑。未来 20 年 AI 的发展路径还存在较大的不确定性。本文提出构建智能体的分类学及演化动力学，

以辅助我国 AI 发展路径的顶层设计,抛砖引玉,期望有识之士共同思考。



刘宇航

CCF 高级会员,《计算》编委。中国科学院计算技术研究所副研究员。主要研究方向为计算机体系结构、高性能计算、存储系统、智能并发系统。
liuyuhang@ict.ac.cn

参考文献

[1] M J Flynn. Very High-Speed Computing Systems[J]. *Proceedings of the IEEE*, 1966, 54(12): 1901-1909.

[2] 孙凝晖,张云泉,刘宇航. 算力 [J]. 中国计算机学会通讯, 2022, 18(12): 106-109.
 [3] 刘宇航,张菲. 计算概念谱系: 算势、算力、算术、算法、算礼 [J]. 中国科学院院刊, 2022, 37(10): 1500-1510.
 [4] 王磊,孙凝晖. BOPs: 一种算力度量指标 [J]. 中国计算机学会通讯, 2024, 20(1): 44-49.
 [5] 李昂生. 人工智能原理: 从计算到谋算的模型、原理与方法 [M]. 北京: 科学出版社, 2024: 6.
 [6] 夏军,林忠辉,占车生,等. 长江流域水生态调度与长江模拟器研发 [J]. 中国科学院院刊, 2023, 38(12): 1767-1780.
 [7] 卢明森. 钱学森思维科学思想 [M]. 北京: 科学出版社, 2012: 164, 169, 241.
 [8] 孙凝晖. 对信息技术新体系的思考 [J]. 中国科学院院刊, 2022, 37(1): 8-14.
 [9] 刘宇航,张云泉. 构建与方法学对称的问题学: 以计算机学科为例 [J]. 中国科学院院刊, 2024, 39(7): 1264-1275.

Some Thoughts on the Taxonomy and Evolutionary Dynamics of Agents from the Perspective of Computational Utility and Intelligence Level

Yuhang Liu

Institutes of Computing Technology, Chinese Academy of Sciences

Abstract: Flynn's taxonomy has significantly advanced our understanding and design of computer system architectures. However, its relevance to classifying computer systems in the era of artificial intelligence (AI) is limited. Currently, AI stands at a critical historical juncture: on one hand, it has achieved considerable progress; on the other hand, it faces challenges such as the unsustainable continuation of scaling laws. The future development path of AI remains highly uncertain. To address these issues, this article proposes constructing a taxonomy and evolutionary dynamics for computer systems from the perspective of computational utility and intelligence. The taxonomy aims to provide a systematic review of agents, while the evolutionary dynamics seeks to offer forward-looking insights into the future evolution of system distributions derived from this classification. By integrating both taxonomy and evolutionary dynamics, this approach aims to optimize and strategically plan the development pathway of AI in China.

Keywords: artificial intelligence; pathway design; computational utility; intelligence level; taxonomy of agents; evolutionary dynamics of agents

摘要: 费林分类法促进了对计算机系统结构的理解和设计,但对人工智能时代的计算机系统分类作用有限。当前人工智能发展进程处于一个重要的历史节点,一方面已经取得一定进展,一方面遇到规模定律难以持续等问题,未来人工智能发展路径还有很大的不确定性。为此,本文提出从算力与智能的角度构建智能体的分类学及演化动力学的设想。通过分类学对智能体进行审视,通过演化动力学对未来分类结果的分布的演变进行前瞻预见,以优化设计和前瞻规划我国人工智能发展路径。

关键词: 人工智能; 路径设计; 算力; 智能水平; 智能体分类学; 智能体演化动力学

中图分类号: TP32; K86

中文引用格式: 刘宇航. 对未来 20 年智能体分化和演进的若干思考——从算力与智能的角度 [J]. 计算, 2026, 2(3): 69-77.

英文引用格式: Yuhang Liu. Some Thoughts on the Taxonomy and Evolutionary Dynamics of Agents from the Perspective of Computational Utility and Intelligence Level[J]. *Computing Magazine of the CCF*, 2026, 2(3): 69-77.

(本文责任编辑: 郭 斌)