

Xiansheng Cai,^{1,*} Sihan Hu,^{2,3,*} Tao Wang,^{4,5,†} Yuan

蔡先生,^{1,*} 思涵 Hu,^{2,3,*} 陶王,^{4,5,†} 元

Huang,^{6,†} Pan Zhang,^{1,7,3} Youjin Deng,^{2,3, (3)} and Kun
Chen^{1,**}

黄,^{6,†} 潘张,^{1,7,3} 尤金邓,^{2,3, (3)} 和 昆陈^{1,**}

¹ Institute of Theoretical Physics, Chinese Academy of
Sciences, Beijing 100190, China

¹ 中国科学院理论物理研究所, 北京 100190, 中国

² Department of Modern Physics, University of Science and
Technology of China, Hefei, Anhui 230026, China

² 中国科学技术大学现代物理系, 安徽合肥 230026, 中国

³ Hefei National Laboratory, University of Science and
Technology of China, Hefei 230088, China

³ 中国科学技术大学合肥国家实验室, 合肥 230088, 中国

⁴ Department of Physics, University of Massachusetts,
Amherst, MA 01003, USA

⁴ 美国马萨诸塞大学阿默斯特分校物理系, 阿默斯特, 马萨
诸塞州 01003, 美国

⁵ Institute of Physics, Chinese Academy of Sciences, Beijing
100190, China

⁵ 中国科学院物理研究所, 北京 100190, 中国

⁶ DP Technology, Beijing 100080, China

中国北京市海淀区, DP 科技, 邮编 100080

⁷ School of Fundamental Physics and Mathematical
Sciences, Hangzhou Institute for Advanced Study, UCAS,
Hangzhou 310024, China

中国科学院大学杭州高等研究院基础物理与数学科学学院,
杭州 310024, 中国

(Dated: June 11, 2025)

(日期: 2025 年 6 月 11 日)

Abstract 摘要

Fundamental physics often confronts complex symbolic problems with few guiding exemplars or established principles. While artificial intelligence (AI) offers promise, its typical need for vast datasets to learn from hinders its use in these information-scarce frontiers. We introduce learning at criticality (LaC), a reinforcement learning (RL) scheme that tunes Large Language Models (LLMs) to a sharp learning transition, addressing this information scarcity. At this transition, LLMs achieve peak generalization from minimal data, exemplified by 7-digit base-7 addition—a test of nontrivial arithmetic reasoning. To elucidate this peak, we analyze a minimal concept-network model (CoNet) designed to capture the essence of how LLMs might link tokens. Trained on a single exemplar, this model also undergoes a sharp learning transition. This transition exhibits hallmarks of a second-order phase transition, notably power-law distributed solution path lengths. At this critical point, the system maximizes a “critical thinking pattern” crucial for generalization, enabled by the underlying scale-free exploration. This suggests LLMs reach peak performance by operating at criticality, where such explorative dynamics enable the extraction of underlying operational rules. We demonstrate LaC in quantum field theory: an 8B-parameter LLM, tuned to its critical point by LaC using a few exemplars of symbolic Matsubara sums, solves unseen, higher-order problems, significantly outperforming far larger models. LaC thus leverages critical phenomena, a physical principle, to empower AI for complex, data-sparse

challenges in fundamental physics.

基础物理学常常面临复杂的符号问题，且缺乏指导性示例或已建立的原理。尽管人工智能（AI）提供了希望，但其通常需要海量数据集进行学习，这阻碍了其在信息稀缺的前沿领域的应用。我们引入临界学习（LaC），这是一种强化学习（RL）方案，通过将大型语言模型（LLMs）调整至尖锐的学习转变点来解决这一信息稀缺问题。在这一转变点，LLMs 通过极少量数据实现峰值泛化，例如 7 位数的 7 进制加法——这是对非平凡算术推理的测试。为阐明这一峰值，我们分析了一个最小概念网络模型（CoNet），该模型旨在捕捉 LLMs 可能链接标记的本质。仅通过单个示例训练，该模型也经历了尖锐的学习转变。这一转变呈现出二阶相变的特征，尤其是解决路径长度呈幂律分布。在这个临界点，系统最大化了对泛化至关重要的“批判性思维模式”，这是由潜在的无标度探索所实现的。这表明 LLMs 通过在临界点运行来达到峰值性能，在此处，这种探索性动态使得提取底层运作规则成为可能。我们在量子场论中展示了 LaC：一个经过 LaC 调整的 80 亿参数的 LLM，仅使用几个符号 Matsubara 和的示例，就能解决未见过高阶问题，显著优于更大的模型。因此，LaC 利用临界现象这一物理原理，为人工智能应对基础物理学中复杂且数据稀缺的挑战赋能。

Introduction-Artificial intelligence (AI) has accelerated scientific discovery, yet its primary successes are in data-rich domains leveraging pattern recognition, a capability closely aligned with intuitive “System 1” thinking [1-5]. A distinct class of frontier scientific problems, particularly in theoretical physics, presents a different challenge: they often necessitate deriving complex analytical solutions through extended abstract, “System 2” reasoning, yet the very nature of these frontiers means training data for AI models is inherently limited [6], seemingly placing them beyond the reach of conventional AI reliant on vast statistical correlations.

引言——人工智能（AI）已经加速了科学发现，但其主要成功往往在于数据丰富的领域，依靠模式识别，这种能力与直觉性的“系统 1”思维密切相关。一类独特的前沿科学问题，特别是在理论物理学中，提出了一个不同的挑战：它们通常需要通过延伸的抽象“系统 2”推理来推导复杂的分析解，然而这些前沿领域的本质意味着 AI 模型的训练数据本质上是有限的，这似乎将这些问题置于依赖大规模统计关联的传统 AI 之外。

This chasm between current AI strengths and the needs of theoretical physics is starkly evident. In Quantum Electrodynamics (QED), the electron’s anomalous magnetic moment (a_e) acts as a critical test of the Standard Model. While numerical evaluations of a_e coefficients are known to high precision (e.g., up to the fifth loop), their complete analytical derivation-essential for deep theoretical insight-is achieved only to the third loop after decades of effort [10-14]. Similarly, in many-electron systems in condensed matter, understanding phenomena like high-temperature superconductivity relies on analytically mastering the Fermi-surface complexities in Feynman diagrams. Here too, numerical meth-

当前人工智能能力与理论物理学需求之间的差距显而易见。在量子电动力学（QED）中，电子的反常磁矩（ a_e ）作为标准模型的关键测试。尽管已经能够高精度地进行 a_e 系数的数值评估（例如，达到第五圈），但经过数十年的努力，完整的分析推导（对于深入的理论洞察至关重要）仅达到第三圈。同样地，在凝聚态物理的多电子系统中，理解高温超导现象依赖于分析性地掌握费曼图中费米面的复杂性。在这里，数值方法也

ods like Diagrammatic Monte Carlo (DiagMC) provide crucial estimates for higher-order terms [15-18], but the analytical solution of even low-order diagrams remains a challenging task. For fields demanding generalizable symbolic reasoning from a few solved instances, AI models traditionally associated with “System 1” appear ill-equipped.

像示意蒙特卡罗（DiagMC）这样的方法为高阶项提供关键估计[15-18]，但即使是低阶图表的解析解仍然是一个具有挑战性的任务。对于需要从少数已解决实例中进行可泛化符号推理的领域，传统上与“系统 1”相关的人工智能模型似乎并不胜任。

However, recent advancements offer a new perspective. Large Language Models (LLMs), particularly when augmented with Reinforcement Learning (RL), are beginning to exhibit capabilities that transcend simple pattern matching [19-26]. RL enables LLMs to actively explore problem spaces, learn from feedback on their generated reasoning pathways, and refine strategies-processes that foster more coherent, goal-oriented, and multi-step thought, intriguingly reminiscent of “System 2” cognition. This opens a promising, albeit challenging, avenue for AI to assist in domains demanding true analytical depth.

然而，近期的进展提供了一个新的视角。大语言模型（LLMs），特别是在强化学习（RL）的增强下，开始展现出超越简单模式匹配的能力 [19-26]。强化学习使 LLMs 能够主动探索问题空间，从其生成的推理路径的反馈中学习，并完善策略，这些过程促进了更连贯、面向目标和多步骤的思考，令人兴奋地让人想起“系统 2”认知。这为人工智能在需要真正分析深度的领域提供了一个充满希望但具有挑战性的途径。

While LLMs augmented by RL show promise for “System 2”-like reasoning, deploying them effectively in frontier science requires navigating inherent challenges. Firstly, the highly specialized nature of these problems means they are often statistical outliers to an LLM’s general pre-training [6]-[9]. This misalignment can lead

虽然经过强化学习增强的 LLMs 对于“系统 2”式推理颇具前景，但在前沿科学中有效部署它们需要应对固有的挑战。首先，这些问题的高度专业性意味着它们对于 LLM 的通用预训练来说通常是统计异常值[6]-[9]。这种不匹配可能

to unreliable outputs or “hallucinations” [27,29] when precise symbolic manipulation is critical, diminishing the utility of off-the-shelf models. A natural corrective is targeted RL fine-tuning, tailoring the LLM to the specific nuances of the cutting-

edge problem. Yet, this essential fine-tuning step itself encounters a profound obstacle: the very frontier nature that necessitates such specialization also implies an extreme scarcity of existing data suitable for this RL process. This creates a central dilemma: if the indispensable RL fine-tuning must operate with exceptionally few examples, can these advanced AI models genuinely acquire the algorithmic understanding and robust generalization capabilities characteristic of “System 2” reasoning [30, 31] Addressing this question is crucial for determining AI’s true potential in advancing fundamental science.

导致不可靠的输出或“幻觉”[27,29]，当需要精确的符号操作时，降低了现成模型的实用性。一个自然的纠正方法是针对性的强化学习微调，使 LLM 适应特定前沿问题的细微差别。然而，这一至关重要的微调步骤本身遇到了一个深刻的障碍：正是那些需要这种专业化的前沿性质也意味着用于此强化学习过程的现有数据极其稀缺。这造成了一个核心困境：如果不可或缺的强化学习微调必须在极少数示例的情况下进行，这些先进的人工智能模型是否能真正获得“系统 2”推理所具有的算法理解和强大的泛化能力[30, 31]。解决这个问题对于确定人工智能在推进基础科学中的真正潜力至关重要。

In this Letter, we resolve this central dilemma by introducing “learning at criticality” (LaC), a learning scheme inspired by critical phenomena in physics. LaC precisely guides an LLM [24] via RL to a narrow ‘critical’ training phase. At this threshold, analogous to a phase transition, the LLM attains strong problem-solving capabilities and optimal generalization. This targeted training to criticality imparts to the LLM genuine, generalizable algorithmic understanding from minimal data—even a single exemplar—circumventing the data-dependency that has historically limited conventional AI in numerous scientific domains. Consequently, LaC enables AI to address complex, abstract theoretical problems in fundamental science previously considered intractable due to data scarcity.

在这封信中，我们通过引入“临界学习”（LaC），解决了这个核心难题，这是一种受物理学临界现象启发的学习方案。LaC 精确地通过强化学习（RL）引导一个 LLM 进入狭窄的“临界”训练阶段。在这个阈值处，类似于相变，LLM 获得了强大的问题解决能力和最佳泛化能力。这种针对性的临界训练使 LLM 能够从最少量的数据中获得真正的、可泛化的算法理解，甚至仅凭单一示例，绕过了历史上限制传统 AI 在众多科学领域发展的数据依赖性。因此，LaC 使 AI 能够解决那些由于数据稀缺而曾被认为难以处理的复杂、抽象的基础科学理论问题。

We first establish LaC by training an LLM on a single instance of 7 -digit, base-7 addition, observing a peak in its generalization capability at a precise training stage (Fig. 1, left panel). To understand the physics underlying this peak, we propose the concept-network model (CoNet). This minimal model abstracts LLM reasoning, where cohesive sequences of autoregressively generated tokens form “concepts”. The LLM’s problem-solving is viewed as a stochastic traversal within an implicit network of these concepts, with RL training (e.g., a variation of Group Relative Policy Optimization (GRPO) [20–23]) adjusting inter-concept transition probabilities. Our CoNet embodies this as a Markovian walk on a random graph of such “concepts”, where transitions are learned via reinforcement to find paths from question to answer. Remarkably, this simplified CoNet reproduces a sharp learning transition exhibiting hallmarks of a second-order phase transition: problem-solving accuracy (Fig 2, upper panel) increases sigmoidally, while reasoning path length variance diverges, signaling critical fluctuations. Crucially, near this critical point, path lengths become power-law distributed ($L^{-\gamma}$, Fig. 3), indicating an emergent “critical thinking pattern” of diverse strategic exploration.

我们首先通过在单个 7 位、基数为 7 的加法实例上训练 LLM 来建立 LaC，观察到其泛化能力在特定训练阶段达到峰值（图 1，左侧面板）。为了理解这一峰值背后的物理机制，我们提出了概念网络模型（CoNet）。这个最小化模型抽象了 LLM 的推理过程，其中自回归生成的连贯标记序列形成“概念”。LLM 的问题求解被视为在这些概念的隐式网络中进行随机遍历，通过强化学习训练（例如，群体相对策略优化（GRPO）的变体）调整概念间的转换概率。我们的 CoNet 将此体现为在这些“概念”的随机图上进行马尔可夫游走，通过强化学习找到从问题到答案的路径。值得注意的是，这个简化的 CoNet 重现了一个显示出二阶相变特征的 sharply 学习转变：问题求解准确性（图 2，上层面板）呈 S 型增长，而推理路径长度方差发散，标志着关键波动。关键的是，在这个临界点附近，路径长度呈幂律分布（ $L^{-\gamma}$ ，图 3），表明出现了一种多样性策略探索的“关键思考模式”。

Finally, we apply LaC to the symbolic Matsubara fre-

最后，我们将拉氏算法应用于符号玻色-马塔拉频率

quency summation in Feynman diagrams - a challenging problem in finite-temperature many-body quantum field theory (QFT) [32–35]. Remarkably, fine-tuning an 8 billion parameter LLM [25] to its critical point using only low-order diagrams enables it to learn the symbolic procedure and solve unseen, more complex diagrams (Fig. 1, right panel), outperforming models with nearly two orders of magnitude more parameters (Tab. II). Our work establishes LaC as a data-efficient strategy for AI-driven discovery in theoretical physics and suggests that emergent reasoning in AI can be understood as a critical phenomenon.

费曼图中的频率求和 - 这是有限温度多体量子场论（QFT）中的一个具有挑战性的问题 [32–35]。值得注意的是，仅使用低阶图通过精细调整一个 80 亿参数的 LLM [25] 到其临界点，使其能够学习符号程序并解决未见过的更复杂图（图 1，右侧面板），其性能超过了参数数量几乎多出两个数量级的模型（表 II）。我们的工作将 LaC 确立为理论物理中人工智能驱动发现的高效数据策略，并暗示人工智能中的涌现推理可以被理解作为一种临界现象。

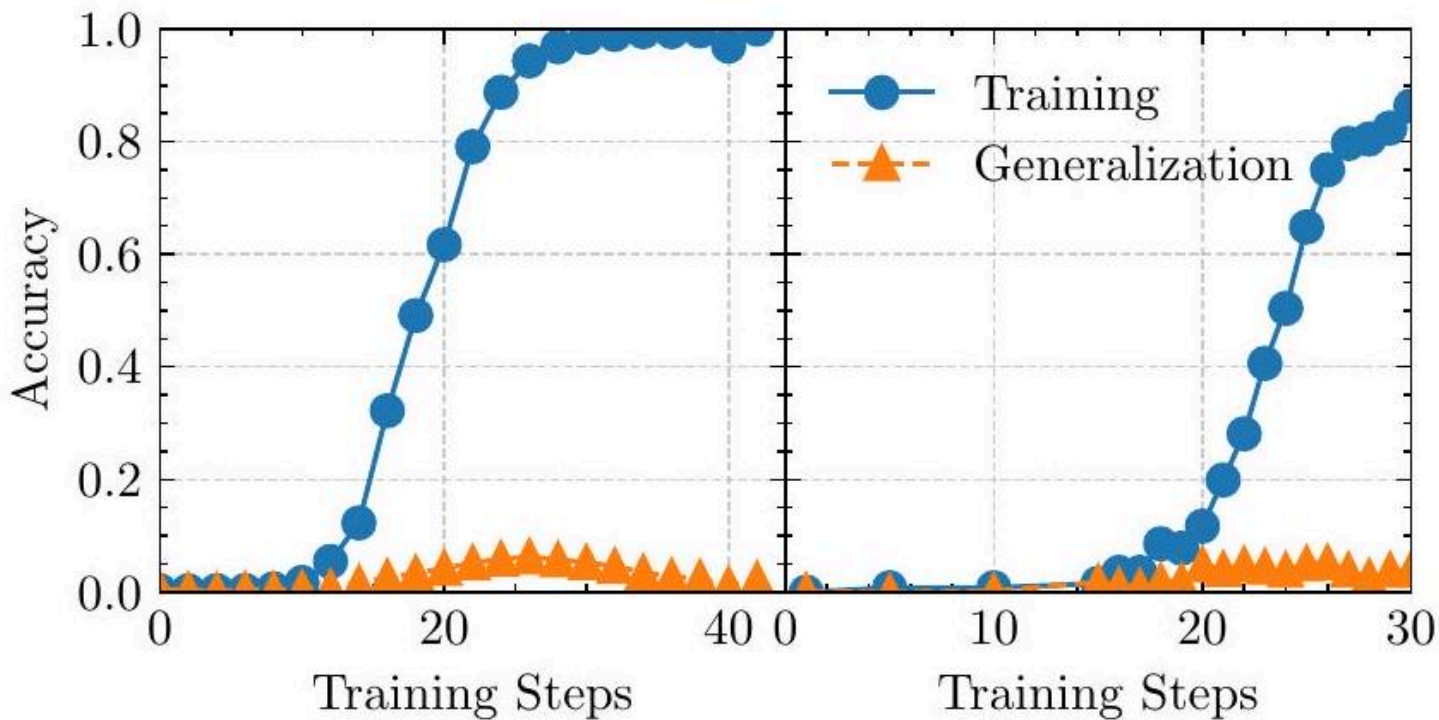


FIG. 1. Critical learning from a single training example. (Left) Training a Qwen2.5-7B model on one 7 -digit base-7 addition. Training accuracy (blue circles) shows a sharp transition. Generalization to unseen additions (orange triangles) peaks precisely at this critical point before overfitting. (Right) Similar phenomenon for a Qwen3-8B model trained on Matsubara frequency summation (2-loop sunrise self-energy diagram). Generalization to other unseen 2-loop diagrams is maximized at the critical learning point.

图 1. 从单个训练示例中的关键学习。(左侧) 在一个 7 位 7 进制加法中训练 Qwen2.5-7B 模型。训练准确率（蓝色圆圈）显示出 Sharp 过渡。对未见过的加法的泛化性（橙色三角形）在临界点之前达到峰值，之后开始过拟合。(右侧) 对于在马苏巴拉频率求和（2 环日出自能图）上训练的 Qwen3-8B 模型，出现类似现象。对其他未见过的 2 环图的泛化性在关键学习点达到最大。

Learning at Criticality: A Phenomenon-Before presenting a theoretical model, we empirically demonstrate that effective learning from sparse data, while achievable, critically depends on the training regime. We tasked a Qwen2.5-7B model [24, initially unable to perform 7digit base-7 addition, with learning this procedure from a single problem instance. This specific task was chosen to probe algorithmic reasoning. Its multi-step, rule-based nature (requiring sequential digit-by-digit processing and explicit base-7 carrying rules) serves as a strong proxy for deliberative “System 2” cognition. Furthermore, its presumed rarity in pre-training corpora, particularly when contrasted with arithmetic in common bases like 2 or 10, significantly minimizes the likelihood of the model recalling a memorized solution rather than learning the underlying algorithm.

临界学习：一个现象-在提出理论模型之前，我们通过实证证明，从稀疏数据中有效学习是可行的，但关键在于训练方案。我们让一个 Qwen2.5-7B 模型[24]，最初无法执行 7 位 7 进制加法，从单个问题实例中学习这个过程。选择这个特定任务是为了探测算法推理。其多步骤、基于规则的特性（需要逐位顺序处理和明确的 7 进制进位规则）可以很好地代表深思熟虑的“系统 2”认知。此外，与 2 进制或 10 进制算术相比，它在预训练语料中的罕见性，显著降低了模型记忆已有解决方案而非学习底层算法的可能性。

We trained the model on this one example using Direct-Advantage Policy Optimization (DAPO) [23, which is a variant of the GRPO algorithm [20, 21. It operates by having the LLM generate multiple potential reasoning paths for a given problem. These paths are evaluated, and their scores, relative to the group average, guide policy updates to the LLM’s parameters. This

我们使用直接优势策略优化（DAPO）[23]方法，这是 GRPO 算法[20, 21]的一个变体，对这个单一示例进行模型训练。其运作方式是让 LLM 为给定问题生成多个潜在推理路径。这些路径被评估，并根据组平均值计算其得分，以指导对 LLM 参数的策略更新。

process strengthens transitions on paths yielding aboveaverage rewards and weakens those on paths performing below average, thereby preferentially guiding the policy towards effective reasoning strategies.

这个过程通过加强高于平均奖励路径的转换并削弱低于平均奖励路径的转换，从而有针对性地引导策略朝向有效的推理策略。

As shown in Fig. 1 (left panel), the training accuracy on the single problem instance (blue circles) displays a sharp, sigmoidal transition, indicating a sudden acquisition of the solution. This transition is not perfectly singular, likely due to multiple valid reasoning paths (e.g., different ways to handle carrying operations), but the qualitative feature is clear. More importantly, when we tested the model’s performance on 128 new, unseen 7digit base-7 addition problems, its generalization accuracy (orange triangles) peaked precisely in the vicinity of this learning transition. Although this initial peak generalization rate was approximately 7% - a modest absolute value, yet significant given it stems from a single training example - we further validated that by carefully continuing the training within this critical regime, the model progressively enhanced its capability to solve unseen instances 36 correctly. As training continued past this point (overfitting), the model’s general ability declined, even as its performance on the single training example remained perfect. This finding is the cornerstone of our “learning at criticality” proposal: it demonstrates the remarkable capacity of an LLM to acquire genuinely generalizable, algorithmic understanding from even a single exemplar, provided the training navigates and sustains it within a transient, optimal learning state.

如图 1（左侧面板）所示，单个问题实例的训练准确率（蓝色圆圈）呈现出急剧的 S 型过渡，表明解决方案突然被获得。这种过渡并非完全单一，可能是由于存在多种有效的推理路径（例如，处理进位操作的不同方式），但定性特征却很明确。更重要的是，当我们在 128 个新的、未见过的 7 位数以 7 为基的加法问题上测试模型性能时，其泛化准确率（橙色三角形）恰好在这个学习过渡点附近达到峰值。尽管这个初始泛化率大约为 7% ——这是一个适中的绝对值，但考虑到它源于单个训练样例，仍然意义重大——我们进一步验证，通过在这个关键阶段谨慎地继续训练，模型逐步增强了正确解决没见过实例的能力。当训练超过这个点（过拟合）时，模型的整体能力下降，即使其在单个训练样例上的性能仍保持完美。这一发现是我们“临界学习”提议的基石：它展示了 LLM 从单个示例中获得真正可泛化的算法理解的卓越能力，只要训练将其引导并维持在一个短暂的、最优的学习状态中。

CoNet Model for Learning Transitions-LLMs generate text autoregressively, predicting subsequent tokens based on prior context. When an LLM predicts a sequence of tokens where each next token has nearly 100% certainty, we consider these tokens to form a cohesive unit, which we abstract as a “concept”. LLM reasoning can then be viewed as a stochastic traversal - akin to a random walk-within an underlying network of these concepts. GRPO training step effectively acts as an external “tuning parameter” that modifies the transition probabilities within the LLM’s implicit concept network, optimizing pathways from the “question concepts” to the “answer concepts”.

CoNet 模型用于学习转换-LLMs 通过自回归方式生成文本，基于先前的上下文预测后续词符。当 LLM 预测一个词符序列，其中每个下一个词符的确定性几乎为 100% 时，我们将这些词符视为形成一个内聚的单元，我们将其抽象为一个“概念”。然后，LLM 推理可以被视为在这些概念的潜在网络中进行随机遍历 - 类似于随机游走。GRPO 训练步骤有效地充当外部“调优参数”，修改 LLM 隐式概念网络中的转换概率，优化从“问题概念”到“答案概念”的路径。

To model this process and understand the physics of LaC, we propose the CoNet. In this minimal model, the LLM’s abstract concept space is represented as a Kregular random graph with N nodes (concepts). A reasoning task is modeled as finding a path from a source node Q (question) to a target node A (answer), constrained to a maximum path length of $L_{\max} = 200$. The LLM’s token generation is simplified to a Markovian walk on this graph, where the transition probability from concept i to a neighboring concept j is

为了模拟这一过程并理解 LaC 的物理机制，我们提出了 CoNet。在这个最小模型中，LLM 的抽象概念空间表示为一个具有 N 个节点（概念）的 K 正则随机图。推理任务被建模为从源节点 Q（问题）到目标节点 A（答案）找到一条路径，路径长度受 $L_{\max} = 200$ 的最大长度限制。LLM 的词符生成简化为在该图上的马尔可夫游走，从概念 i 到相邻概念 j 的转换概率是

$$\pi_{\theta}(j | i) = \frac{\theta_{ij}}{\sum_{k \in \text{neighbors}(i)} \theta_{ik}}$$

where $\theta_{ij} \in [0, 1)$ are learnable parameters representing

其中 $\theta_{ij} \in [0, 1)$ 是可学习的参数，表示

transition strengths. For a given Q-A pair, $M = 10^4$ reasoning paths (indexed by m) are sampled. Each path receives a reward, and its advantage A_m (relative to the average reward) guides the update of θ_{ij} via a GRPOvariant rule [20, 22, 23] $\Delta\theta_{ij} \propto \sum_m A_m \nabla_{\theta_{ij}} \log \pi_{\theta}(j | i)$ This reinforces transitions on above-average paths.

转换强度。对于给定的 Q-A 对，对 $M = 10^4$ 个推理路径（由 m 索引）进行采样。每条路径都会获得一个奖励，其优势 A_m （相对于平均奖励）通过 GRPO 变体规则[20, 22, 23]引导 θ_{ij} 的更新 $\Delta\theta_{ij} \propto \sum_m A_m \nabla_{\theta_{ij}} \log \pi_{\theta}(j | i)$ 。这加强了高于平均水平的路径上的转换。

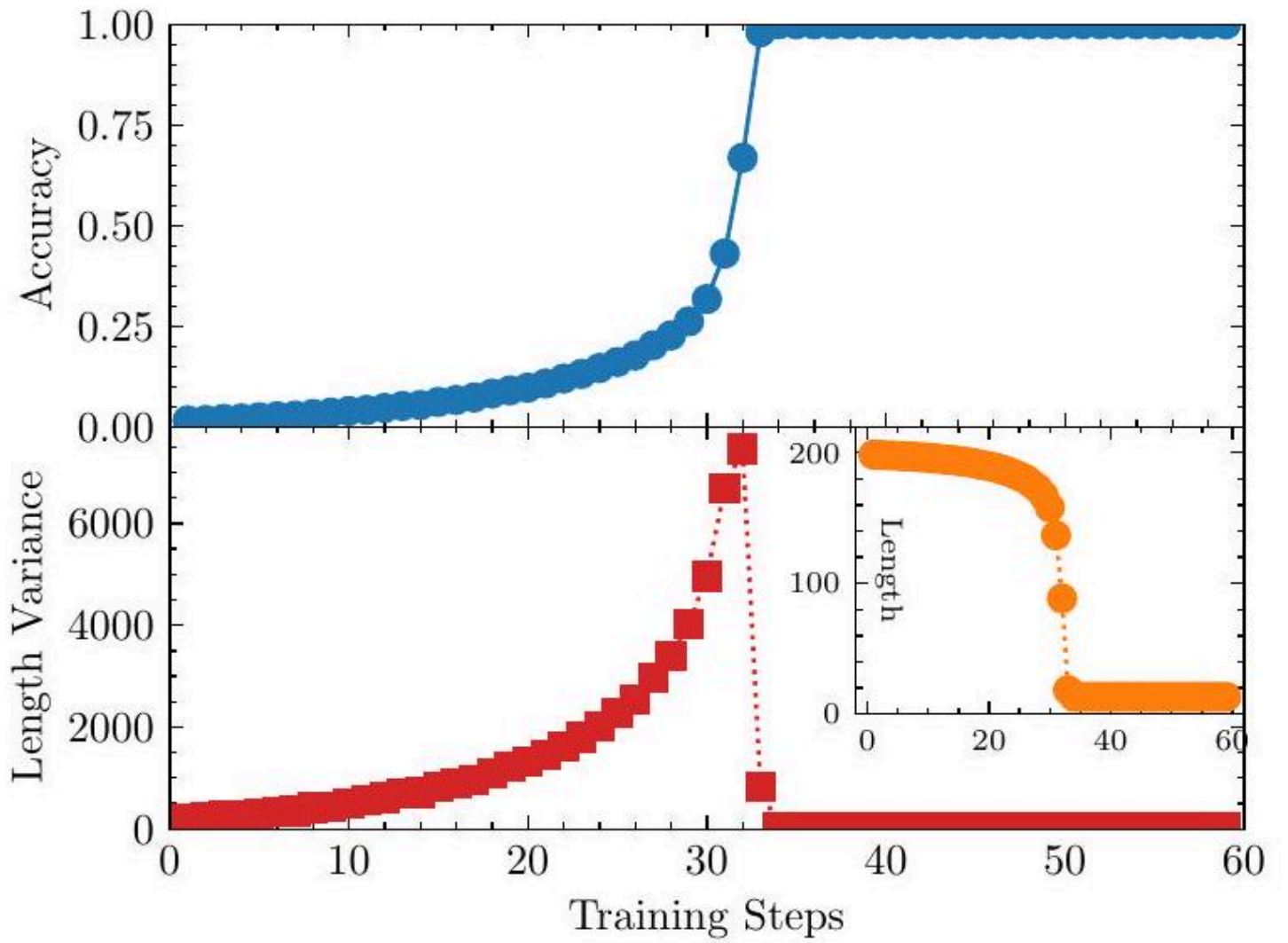


FIG. 2. Training dynamics of the minimal concept-network model (CoNet). The figure shows the accuracy, average response length, and the response length's variance of the minimal model on the training problem, plotted against training steps. The accuracy (blue) increases and the average response length (orange) decreases in a sigmoidal manner. Concurrently, the response length's variance (red) exhibits a lambda shape discontinuity at the learning transition, mirroring the behavior of specific heat at the lambda point marking the normal-to-superfluid helium phase transition.

图 2. 最小概念网络模型 (CoNet) 的训练动态。图中显示了最小模型在训练问题上的准确率、平均响应长度和响应长度的方差, 并绘制了训练步骤的变化。准确率 (蓝色) 以 S 型方式增加, 平均响应长度 (橙色) 减少。同时, 响应长度的方差 (红色) 在学习转变点呈现 λ 形状的不连续性, 这反映了标志着氦从普通态到超流态相变的 λ 点处的比热行为。

Simulations of CoNet ($N = 8000$ nodes, $K = 5$) reveal a sharp learning transition (Fig. 2). The accuracy (fraction of successful paths, serving as an order parameter) displays a steep, sigmoidal increase. Concurrently, the variance of the reasoning path lengths exhibits a pronounced peak at the transition, analogous to diverging susceptibility at a critical point in physical systems. These features are hallmarks of a continuous phase transition and mirror the empirical learning peak.

CoNet 的模拟 ($N = 8000$ 个节点, $K = 5$) 揭示了一个明显的学习转变 (图 2)。准确率 (成功路径的比例, 作为秩序参数) 显示出陡峭的 S 型增长。同时, 推理路径长度的方差在转变点表现出显著的峰值, 类似于物理系统临界点处发散的敏感性。这些特征是连续相变的标志, 并反映了经验学习峰值。

The microscopic origin of this transition is revealed in the hybrid nature of the reasoning path distribution $P(L)$ at criticality. As shown in Fig. 3 (around step 30), the distribution is effectively a superposition of two distinct search strategies. A peak at short lengths signifies local exploration around an emergent optimal path, a behavior that consolidates into a purely exponential decay ($P(L) \sim e^{-\alpha L}$) post-transition (step 36). Coexisting with this is a pronounced power-law tail, $P(L) \sim L^{-\gamma}$ ($\gamma \approx 0.16$), a canonical signature of scale-free, critical phenomena. The prominence of this exploratory, powerlaw mode is maximized at the transition, directly causing the large path-length variance (susceptibility). This “critical

thinking pattern” represents a diverse, longrange exploratory search, and its coexistence with effi-

这一转变的微观起源体现在临界点处推理路径分布的混合性质 $P(L)$ 。如图 3 所示 (约 30 步), 该分布实际上是两种不同搜索策略的叠加。短长度处的峰值表明围绕新生最优路径进行局部探索, 这一行为在转变后巩固为纯指数衰减 ($P(L) \sim e^{-\alpha L}$) (36 步)。与此共存的是显著的幂律尾部, $P(L) \sim L^{-\gamma}$ ($\gamma \approx 0.16$), 这是标志性的无标度临界现象。这种探索性的幂律模式在转变点达到最大, 直接导致路径长度的大方差 (敏感性)。这种"批判性思维模式"代表了多样化的、长距离的探索性搜索, 其与高效的