Learning-at-Criticality in Large Language Models for Quantum Field Theory and Beyond

临界学习：大语言模型在量子场论及其他领域的应用

Xiansheng Cai, [1,*] Sihan Hu, [2,3,*] Tao Wang, [4,5,†] Yuan

蔡贤生， [1,*] 司涵 Hu, [2,3,*] 陶望， [4,5,†] 袁

Huang, [6,†] Pan Zhang, [1,7,3] Youjin Deng, [2,3,], [(3)] and Kun Chen [1,**]

黄， [6,†] 张潘， [1,7,3] 邓友金， [2,3,] ， [(3)] 和陈坤 [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

[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

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

[4] 马萨诸塞大学阿默斯特分校物理系，美国马萨诸塞州 01003

[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

[6] DP Technology, Beijing 100080, China

[6] 北京深势科技有限公司，中国北京 100080

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

[7] 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，中国杭州 310024

(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.

基础物理学常面临符号复杂且范例稀少或原理未明的问题。尽管人工智能（AI）展现出潜力，但其通常依赖海量数据学习的特点阻碍了在这些信息匮乏领域的应用。我们提出临界学习（LaC）——一种通过强化学习（RL）将大型语言模型（LLMs）调整至敏锐学习转变的方案，以应对信息稀缺性。在此转变点，LLMs 能以极少数据实现最佳泛化能力，例如 7 进制 7 位数加法这一非平凡算术推理测试。为阐明这一峰值现象，我们分析了一个最小化概念网络模型（CoNet），其设计旨在捕捉 LLMs 关联标记的核心机制。该模型仅用单一范例训练后，同样会出现陡峭的学习转变。这种转变呈现出二阶相变的特征，尤其是解路径长度呈现幂律分布。在此临界点，系统通过底层的无标度探索，最大化对泛化至关重要的"临界思维模式"。 这表明 LLMs 通过在临界状态下运行达到峰值性能，这种探索性动态使其能够提取潜在的操作规则。我们在量子场论中展示了 LaC 方法：一个 80 亿参数的 LLM，通过 LaC 使用少量符号化松原求和的示例调整至临界点后，能够解决未见过的更高阶问题，其表现显著优于规模大得多的模型。因此，LaC 利用临界现象这一物理原理，使 AI 能够应对基础物理学中数据稀疏的复杂挑战。

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)已加速了科学发现，但其主要成功集中在利用模式识别的数据丰富领域[1-5]，这种能力与直觉性的"系统 1"思维高度契合。而理论物理等前沿科学领域存在一类截然不同的问题：它们往往需要通过延展的抽象"系统 2"推理来推导复杂解析解，但这些前沿问题的本质决定了 AI 模型的训练数据天然受限[6]，这使得依赖海量统计关联的传统 AI 似乎难以触及这类问题。

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 manyelectron 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-

当前人工智能的优势与理论物理学需求之间的鸿沟极为明显。在量子电动力学（QED）中，电子反常磁矩（ a_e ）是检验标准模型的关键指标。虽然 a_e 系数的数值计算已达到高精度（例如直至第五圈图），但要获得完整的解析推导——这对深入理论认知至关重要——经过数十年努力仅实现到第三圈图[10-14]。类似地，在凝聚态多电子体系中，理解高温超导等现象需要从解析层面掌握费米面在费曼图中的复杂表现。即便在此领域，

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 illequipped.

图解蒙特卡洛（DiagMC）等数值方法为高阶项提供了关键估算[15-18]，但即便低阶图解的解析求解仍是艰巨挑战。对于需要从少量已解案例中归纳符号推理的领域，传统"系统 1"类型的人工智能模型显得力不从心。

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 strate-gies-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，近期研究进展提供了新的视角。大型语言模型（LLMs）——尤其是结合强化学习（RL）增强的版本——正开始展现出超越简单模式匹配的能力[19-26]。RL 使 LLMs 能够主动探索问题空间，从其生成的推理路径反馈中学习，并优化那些促进更连贯、目标导向和多步骤思考的策略流程，这种特性与"系统 2"认知模式有着耐人寻味的相似性。这为 AI 在需要真正分析深度的领域提供了一条充满前景（尽管充满挑战）的发展路径。

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

尽管经过 RL 增强的 LLMs 在"类系统 2"推理方面展现出潜力，但要将其有效部署在前沿科学领域仍需克服固有挑战。首先，这些高度专业化的问题往往构成 LLM 通用预训练数据中的统计离群值[6-9]，这种错位可能导致

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 cuttingedge 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，当精确的符号操作至关重要时，这会降低现成模型的实用性。一个自然的纠正方法是针对性的强化学习微调，使 LLM 适应前沿问题的特定细微差别。然而，这一关键的微调步骤本身遇到了一个深刻的障碍：正是这种需要专业化的前沿性质，也意味着适合这一强化学习过程的现有数据极其稀缺。这就形成了一个核心困境：如果不可或缺的强化学习微调必须在极少数例子上运行，这些先进的人工智能模型能否真正获得“系统 2”推理 30,31 所特有的算法理解和稳健泛化能力？解决这个问题对于确定人工智能在推动基础科学方面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）解决了这一核心困境——该学习方案受物理学中临界现象的启发。LaC 通过强化学习精确引导 LLM 24 进入一个狭窄的"临界"训练阶段。在这个类似于相变的阈值上，LLM 获得了强大的问题解决能力和最优泛化性能。这种针对临界状态的定向训练，使 LLM 能够从极少数据（甚至单个样本）中获得真正可泛化的算法理解，从而规避了传统 AI 在众多科学领域长期受限于数据依赖性的问题。因此，LaC 使 AI 能够解决基础科学中那些因数据稀缺而被认为难以处理的复杂抽象理论问题。

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.

我们首先通过在一个 7 位七进制加法实例上训练 LLM 来建立 LaC，观察到其泛化能力在某个精确训练阶段达到峰值（图 1 左面板）。为理解这一峰值背后的物理机制，我们提出概念网络模型（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 的问题解决被视为在这些概念隐式网络中的随机遍历，通过 RL 训练（例如 GRPO 2023 的变体）调整概念间转移概率。我们的 CoNet 将其建模为这种"概念"随机图上的马尔可夫游走，其中通过强化学习来寻找从问题到答案的路径转移。值得注意的是，这个简化的 CoNet 再现了具有二阶相变特征的陡峭学习转变：问题解决准确度（图 2 上面板）呈 S 型增长，而推理路径长度方差发散，标志着临界涨落。关键的是，在这个临界点附近，路径长度呈现幂律分布（ $L^{-\gamma}$ ，图。 3），表明了一种新兴的"批判性思维模式"，即多样化的战略探索。

Finally, we apply LaC to the symbolic Matsubara fre-

最后，我们将 LaC 应用于符号松原频

quency summation in Feynman diagrams - a challenging problem in finite-temperature many-body quantum field theory (QFT) 32-35. Remarkably, fine-tuning an 8billion 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，仅使用低阶图将 80 亿参数的 LLM[25]微调至临界点后，它就能学会符号化流程并解决未见过的更复杂图形（图 1 右面板），其表现优于参数规模大近两个数量级的模型（表 II）。我们的工作确立了 LaC 作为理论物理中 AI 驱动发现的数据高效策略，并表明 AI 中涌现的推理能力可被理解为一种临界现象。

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. 从单个训练样本中实现关键学习。(左)在七进制 7 位数加法任务上训练 Qwen2.5-7B 模型。训练准确率(蓝色圆点)呈现急剧转变。对未见加法运算的泛化能力(橙色三角)恰在此临界点达到峰值后出现过拟合。(右)Qwen3-8B 模型在松原频率求和任务(2 环日出自能图)训练中呈现类似现象。对其他未见 2 环图式的泛化能力在关键学习点达到最大化。

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：现象观察——在提出理论模型前，我们通过实证研究表明，从稀疏数据中进行有效学习虽然可行，但关键取决于训练机制。我们让一个最初无法完成 7 进制 7 位数加法运算的 Qwen2.5-7B 模型[24]，仅通过单个问题实例来学习该运算流程。选择这一特定任务旨在探究算法推理能力：其多步骤、基于规则的本质（需要逐位顺序处理并遵循明确的 7 进制进位规则）能有效模拟审慎的"系统 2"认知过程。此外，考虑到该任务在预训练语料中（尤其是与常见进制如 2 进制或 10 进制算术相比）的罕见性，能极大降低模型直接调用记忆答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]在这个示例上训练模型，该方法是 GRPO 算法[20,21]的一个变体。其运作方式是让 LLM 为给定问题生成多条潜在推理路径。这些路径会被评估，其相对于群体平均分的得分将指导对 LLM 参数的策略更新。这一

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.

过程会强化产生高于平均回报路径的转移概率，同时削弱表现低于平均水平的路径，从而优先引导策略向有效推理策略方向发展。

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（左面板）所示，单个问题实例的训练准确率（蓝色圆圈）呈现出陡峭的 S 型跃迁，表明解决方案的突然获取。这种跃迁并非完全单一，可能源于多种有效推理路径（例如处理进位运算的不同方式），但其定性特征十分清晰。更重要的是，当我们在 128 道新的、未见过的 7 位数七进制加法题上测试模型性能时，其泛化准确率（橙色三角）恰好在学习跃迁附近达到峰值。尽管初始峰值泛化率约为 7% ——绝对值虽小却意义重大，因其仅源于单个训练样本——我们进一步验证发现，通过在此关键阶段谨慎延续训练，模型逐步提升了对未见实例的正确求解能力。当训练越过该临界点（过拟合）后，即使模型在单一训练样本上的表现保持完美，其泛化能力仍持续下降。 这一发现是我们"临界点学习"理论的核心基石：它证明了 LLM 具备非凡的能力，即使仅通过单个示例也能获得真正可泛化的算法性理解——前提是训练过程能引导并维持模型处于这种短暂存在的最优学习状态。

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”.

学习跃迁的共网模型-LLMs 通过自回归方式生成文本，基于先前的上下文预测后续标记。当 LLM 预测的标记序列中每个后续标记都具有接近 100%的确定性时，我们认为这些标记构成了一个连贯单元，并将其抽象为"概念"。因此，LLM 的推理过程可视为在这些基础概念网络中的随机游走——类似于随机漫步。GRPO 训练步骤实质上充当了外部"调谐参数"，通过修改 LLM 隐含概念网络中的转移概率，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

为模拟这一过程并理解 LaC 的物理机制，我们提出了 CoNet 模型。在这个最小化模型中，LLM 的抽象概念空间被表示为具有 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）寻找路径的过程，且路径长度限制在 $L_{\max}=200$ 以内。LLM 的标记生成过程被简化为该图上的马尔可夫游走，其中从概念 i 到相邻概念 j 的转移概率由下式给出：

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

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

其中 $\theta_{ij} \in[0,1)$ 表示可学习参数，用于表征

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] $\Delta \theta_{ij} \propto \sum_m A_m \nabla_{\theta_{ij}} \log \pi_\theta(j \mid i)$ This reinforces transitions on above-average paths.

转移强度。对于给定的 Q-A 配对，会采样 $M=10^4$ 条推理路径（以 m 索引）。每条路径获得一个奖励值，其优势函数 A_m（相对于平均奖励）通过 GRPO 变体规则[20,22,23] $\Delta \theta_{ij} \propto \sum_m A_m \nabla_{\theta_{ij}} \log \pi_\theta(j \mid i)$ 指导 θ_{ij} 的更新，从而强化那些高于平均水平的路径转移。

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 lambdashape 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。该图展示了最小模型在训练问题上的准确率、平均响应长度及响应长度方差随训练步数的变化曲线。准确率(蓝色)以 S 型曲线上升，平均响应长度(橙色)以 S 型曲线下降。与此同时，响应长度方差(红色)在学习转变点呈现λ形突变，这与氦在正常态-超流态相变λ点处比热的突变行为相呼应。

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 的模拟实验($N=8000$ 节点， $K=5$)显示出陡峭的学习转变曲线(图 2)。准确率(作为序参量的成功路径占比)呈现陡峭的 S 型增长。与此同时，推理路径长度的方差在转变点出现显著峰值，这与物理系统临界点处发散的敏感性类似。这些特征标志着连续相变的存在，并与实证学习峰值现象相吻合。

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-

这一转变的微观机制体现在临界状态下推理路径分布 $P(L)$ 的混合特性中。如图 3 所示（约第 30 步附近），该分布实质上是两种不同搜索策略的叠加。短路径处的峰值表明系统在涌现的最优路径周围进行局部探索，这种行为在相变后（第 36 步）会固化为纯粹的指数衰减（ $P(L) \sim e^{-\alpha L}$ ）。与之共存的是一个显著的长尾幂律分布 $P(L) \sim L^{-\gamma}$ （$\gamma \approx 0.16$），这是无标度临界现象的典型特征。这种探索性的幂律模式在相变点达到最大强度，直接导致路径长度方差（敏感性）的显著增大。这种"临界思维模式"代表了一种多样化、长程的探索性搜索，其与高效...

FIG. 3. Distinct reasoning dynamics: critical powerlaw search transitions to post-convergence exponential exploration. Toy model reasoning response length distributions $P(L)$ across training epochs. During the critical learning transition (e.g., step 30, orange), long exploratory responses exhibit characteristic power-law decay $P(L) \sim L^{-\gamma}$ with $\gamma \approx 0.16$ (dashed fit, left panel; also evident for early-stage odd paths, right panel), a signature of scale-invariant critical search. Post-transition, as the policy converges (e.g., to a 7 -step optimal odd path, step 34, dark grey), local perturbations around this path display exponential decay $P(L) \sim e^{-\alpha L}$ (dotted fit, right panel). These distinct scaling regimes characterize the evolution from broad, critical exploration to refined exploitation, crucial for learning.

图 3. 差异化的推理动态：临界幂律搜索向收敛后指数探索的转变。玩具模型推理响应长度分布随训练周期的变化 $P(L)$ 。在关键学习转变阶段（如第 30 步，橙色），长探索性响应呈现典型的幂律衰减特征 $P(L) \sim L^{-\gamma}$ （左图虚线拟合；右图中早期奇数路径亦可见），这是标度不变临界搜索的标志。转变后，随着策略收敛（如收敛至 7 步最优奇数路径，第 34 步，深灰色），该路径周围的局部扰动呈现指数衰减 $P(L) \sim e^{-\alpha L}$ （右图点线拟合）。这两种不同的标度机制刻画了从广泛临界探索到精细利用的演化过程，这对学习至关重要。

cient exploitation is the hallmark of the learning transition, enabling the discovery of generalizable strategies before the system collapses into a single, non-exploratory state.

高效利用是学习转变的标志，它使系统在坍缩至单一非探索状态前，能够发现可泛化的策略。

AI for QFT - To demonstrate LaC's potential for addressing data-scarce 6-9, symbolically complex problems in theoretical physics, we apply it to a challenging representative task: the symbolic evaluation of Matsubara frequency sums 34, 35. This multi-step, algorithmically rich procedure is foundational to finitetemperature QFT [32, 33] and methods like DiagMC for many-electron problems [15-18, 37-43. An example is the 2 -loop sunrise self-energy diagram (Eq 2), whose symbolic solution involves contour integration and complex analysis.

人工智能应用于 QFT ——为展示 LaC 在解决理论物理中数据稀缺（6-9）、符号复杂度高的问题上的潜力，我们将其应用于一个具有挑战性的代表性任务：松原频率求和的符号计算（34,35）。这一多步骤、算法丰富的流程是有限温度量子场论[32,33]的基础，也是处理多电子问题的 DiagMC 等方法[15-18,37-43]的核心。以双圈日出图自能（公式 2）为例，其符号解涉及围道积分和复变分析。

$$
\begin{aligned}
\nu_0+\nu_{\nu_0}^{\nu_1}-\nu_1 &= \sum_{\nu_0,\nu_1} \frac{1}{\nu_0+\varepsilon_0} \frac{1}{\nu_1+\varepsilon_1} \frac{1}{\nu_0+\nu_2-\nu_1+\varepsilon_2} \\
&= \oint_C \frac{dz'}{2\pi i} \frac{f\left(z'\right)}{z'+\varepsilon_1} \oint_C \frac{dz}{2\pi i} \frac{f(z)}{z+\varepsilon_0} \frac{1}{z+\nu_2-z'+\varepsilon_2}
\end{aligned}
$$

where ν are fermionic Matsubara frequencies, ϵ are energy parameters, and $f(z)$ is the Fermi-Dirac distribution. This task is an ideal LaC testbed due to its structured complexity, availability of low-order diagrams for sparse training, and verifiable analytical solutions, allowing us to probe for emergent algorithmic understanding.

其中 ν 表示费米子松原频率， ϵ 为能量参数， $f(z)$ 是费米-狄拉克分布函数。该任务因其结构化复杂性、可用于稀疏训练的低阶图例可得性、以及可验证的解析解，成为 LaC 的理想测试平台，使我们能够探索算法理解的涌现特性。

The well-defined, symbolic nature of this procedure makes it an ideal testbed for LaC. Specifically: (i) low-

这一流程具有明确的符号化特征，使其成为 LaC 的理想测试平台。具体表现为：（i）低阶

order diagrams offer sparse training instances; (ii) complexity scales with diagram order, naturally testing generalization; and (iii) exact analytical solutions permit unambiguous verification. Our objective is to investigate if an LLM, trained via staged LaC on minimal examples, can acquire the algorithmic reasoning inherent in this task, highlighting LaC's potential for data-efficient learning of complex physics procedures.

阶数图提供了稀疏的训练实例；(ii) 复杂度随图阶数增加而自然测试泛化能力；(iii) 精确解析解允许明确验证。我们的目标是研究通过分阶段 LaC 在最小示例上训练的 LLM，是否能掌握该任务固有的算法推理能力，从而突显 LaC 在数据高效学习复杂物理流程方面的潜力。

Our training employed a staged LaC strategy with a Qwen3-8B model [25] using GRPO. Phase 1: LaC training on only tree-level and 1 -loop Matsubara sums. This not only enabled solutions for these simpler diagrams but also induced initial generalization to higher-order diagrams decomposable into these learned components. Notably, while the base model had zero success on the challenging sunrise self-energy diagram (Eq. 2) - a 2-loop graph with nested frequencies-this first LaC phase elevated its performance to a non-zero baseline. Phase 2: This crucial improvement enabled a second LaC phase focused specifically on the sunrise diagram. As shown in Fig. 1 (right panel), the model again exhibited a sharp learning transition for this complex task. Critically, at this new critical point, the ability gained from mastering the sunrise diagram is generalized to other 2-loop topologies (e.g., polarization, vertex functions), where performance was previously negligible, confirming the LaC hypothesis. Phase 3: Finally, starting from this enhanced critical state, we incorporated all 2 -loop diagrams into the LaC training until a critical point was reached for this comprehensive set.

我们的训练采用了分阶段 LaC 策略，使用 Qwen3-8B 模型[25]结合 GRPO 方法。第一阶段：仅针对树状图和单圈松原求和的 LaC 训练。这不仅使模型能够求解这些简单图表，还促使模型初步泛化到可分解为这些已学习组件的高阶图表。值得注意的是，虽然基础模型在具有嵌套频率的双圈日出自能图（公式 2）这一挑战性问题上完全无法求解，但首个 LaC 阶段将其性能提升至非零基准水平。第二阶段：这一关键改进使得我们可以开展专门针对日出图的第二阶段 LaC 训练。如图 1 右面板所示，模型对这一复杂任务再次展现出陡峭的学习跃迁。至关重要的是，在这个新的临界点上，通过掌握日出图所获得的能力被泛化到其他双圈拓扑结构（如极化函数、顶点函数），而这些结构此前几乎无法求解，这验证了 LaC 假设。第三阶段：最终从这个增强的临界状态出发，我们将所有双圈图表纳入 LaC 训练，直至针对这一完整集合达到新的临界点。

Accuracy(%) Problem Model 准确率(%) 问题模型	1-loop 单循环	2-loop 双循环	3-loop 三循环	4-loop 4 循环
Qwen3-8B	45.0	18.4	0.9	0.2
Qwen3-8B-1-loop Qwen3-8B-1 循环	90.2	25.5	3.3	0.8
Qwen3-8B-2-loop Qwen3-8B-2 循环	97.5	56.9	9.5	1.7
Qwen3-32B	47.8	25.5	1.6	0.6
DeepSeek-R1-0120(671B)	12.5	10.0	1.1	0.4

TABLE I. Critical learning transition enables generalization for Matsubara sums. Sharp increase in the success rate for Qwen3-8B fine-tuned on problems up to 2-loop complexity. This critical learning phase unlocks robust generalization to unseen 3-loop overlapped sums, markedly outperforming significantly larger base models (Qwen3-32B, DeepSeek-R1 671B) on this task, indicating data-efficient acquisition of algorithmic reasoning.

表一. 临界学习跃迁实现松原求和的泛化能力。经 2-loop 复杂度问题微调的 Qwen3-8B 模型成功率急剧上升，这一关键学习阶段使其对未见过的 3-loop 重叠求和展现出强大泛化能力，显著超越参数量大得多的基础模型（Qwen3-32B、DeepSeek-R1 671B），表明其算法推理能力的数据高效习得特性。

The results are shown in Table I The base Qwen38 B model's success is modest for 1 -loops (45.0%) and 2-loops (18.4%), and minimal for 3 -loops (0.9%) and 4 -loops (0.2%). The latter non-zero values reflect occasional solutions to simple, decomposable higher-order diagrams, not general complex reasoning. In contrast, the fully fine-tuned model (Qwen3-8B-2-loop, post-Phase 3) greatly improved accuracy on its 2 -loop training set. Crucially, it then achieved strong accuracy on complex

如表一所示，基础 Qwen3-8B 模型在 1-loop（ 45.0% ）和 2-loop（18.4%）任务上表现平平，在 3-loop（ 0.9% ）和 4-loop（ 0.2% ）任务上成功率极低。后者的非零值仅反映对简单可分解高阶图的偶然求解，而非通用复杂推理能力。相比之下，完整微调模型（Qwen3-8B-2-loop，第三阶段后）在 2-loop 训练集上准确率大幅提升，更关键的是，该模型随后在复杂

3-loop problems, despite no explicit training on them, and generalized to 4-loop cases. This LaC-trained model far surpasses its base and larger models (DeepSeek-R1, Qwen3-32B), underscoring LaC's efficiency. This success demonstrates that navigating critical points enabled the LLM to develop an emergent, algorithmic grasp of Matsubara summation.

尽管没有针对 3 循环问题的明确训练，该模型仍能解决此类问题，并推广至 4 循环情形。经过 LaC 训练的模型性能远超其基础版本及更大规模的模型（DeepSeek-R1、Qwen3-32B），彰显了 LaC 方法的高效性。这一成功表明，通过对关键点的探索，LLM 能够形成对松原求和的涌现式算法理解。

Conclusion-We addressed the challenge of “System 2” AI reasoning in data-scarce science by introducing “learning at criticality” (LaC), where LLM fine-tuning via RL induces a critical phase transition towards generalizable algorithmic understanding from minimal data.

结论——我们通过引入"临界点学习"（LaC）方法，解决了数据稀缺科学领域中"系统 2"AI 推理的难题。该方法通过强化学习对 LLM 进行微调，使其在极少量数据下实现向可泛化算法理解的关键相变。

This LaC principle was first shown with 7 -digit base7 addition: generalization from a single example peaked at the sharp learning transition. Our minimal physical model CoNet indicates that it is a continuous phase transition, revealing power-law scaling in reasoning paths-a signature of emergent “critical thinking”. Critically, applying LaC to symbolic Matsubara frequency summation, an 8-billion parameter LLM, trained on a few low-order diagrams, learn the procedure and solves unseen, more complex diagrams, substantially outperforming models nearly two orders of magnitude larger.

这一 LaC 原理最初通过 7 位七进制加法任务得以验证：仅凭单个示例的泛化能力在陡峭的学习转折点达到峰值。我们构建的最小物理模型 CoNet 表明这是一个连续相变过程，揭示了推理路径中的幂律标度特征——这正是涌现式"临界思维"的标志性表现。关键突破在于，将 LaC 应用于符号化的松原频率求和任务时，一个仅通过少量低阶图训练、参数量达 80 亿的 LLM 不仅掌握了计算流程，还能求解从未见过的更复杂图表，其表现显著优于规模近乎大两个数量级的模型。

Critically, this LaC-induced phase transition occurs in an implicit concept network linked to “System 2” reasoning, differentiating it from grokking phenomena which occur during pre-training or supervised fine-tuning and are associated with “System 1” pattern matching 44, 45 . LaC thus provides a novel mechanism for cultivating advanced AI capabilities.

值得注意的是，LaC 诱导的相变发生在与"系统 2"推理相关的隐式概念网络中，这使其区别于预训练或有监督微调阶段出现的顿悟现象（后者与"系统 1"模式匹配相关[44,45]）。LaC 由此为培养高级 AI 能力提供了一种全新机制。

This work provides both a physical framework for understanding emergent AI reasoning and a practical, dataefficient strategy for applying it to data-scarce domains like theoretical physics. The implications of LaC extend deeply into domains like QFT, where navigating immense symbolic complexity with limited examples of optimal strategies is paramount 46, 47. Beyond theoretical physics, LaC provides a pathway towards AI systems as powerful scientific collaborators, equipped to tackle complex symbolic manipulations and accelerate discovery across diverse frontiers of knowledge where data is sparse but the demand for deep reasoning is high.

这项工作不仅为理解人工智能涌现的推理能力提供了物理框架，还提出了一种实用且数据高效的应用策略，可适用于理论物理等数据稀缺领域。LaC 的深远影响延伸至量子场论等领域——在这些领域中，如何通过有限的最优策略示例来驾驭庞大的符号复杂性至关重要[46,47]。除理论物理外，LaC 为构建强大的人工智能科学协作系统开辟了道路，使其能够处理复杂的符号运算，并在数据匮乏但亟需深度推理的多元知识前沿领域加速科学发现。

Acknowledgements - The authors thank Haijun Zhou, Lei Wang, Linfeng Zhang, and Pengcheng Hou for inspiring discussions. K. C. and X. C. are supported by the National Key Research and Development Program of China under Grant No. 2024 YFA1408604 and the National Natural Science Foundation of China under Grants No. 12047503 and No. 12447103. P.Z. is supported by Project 12325501 of the National Natural Science Foundation of China. Y.D. and S.H. are supported by the National Natural Science Foundation of China under Grants No. 12275263, the Innovation Program for Quantum Science and Technology under Grant No. 2021ZD0301900, and the Natural Science Foundation of Fujian Province of China under Grant No. 2023J02032.

致谢 - 作者感谢周海军、王磊、张林峰和侯鹏程富有启发性的讨论。K.C.和 X.C.的研究工作得到中国国家重点研发计划（项目编号 2024YFA1408604）及国家自然科学基金（项目编号 12047503、12447103）的资助。P.Z.的研究工作得到国家自然科学基金项目（编号 12325501）的资助。Y.D.和 S.H.的研究工作得到国家自然科学基金（项目编号 12275263）、科技创新 2030-量子调控与量子信息专项（项目编号 2021ZD0301900）以及福建省自然科学基金（项目编号 2023J02032）的资助。

These authors contributed equally to this paper.

这些作者对本文的贡献同等重要。

† taowang@umass.edu
‡ huangyuan@dp.tech
§ panzhang@itp.ac.cn

yjdeng@ustc.edu.cn
** chenkun@itp.ac.cn
[1] K. E. Stanovich and R. F. West, Behavioral and Brain Sciences 23, 645 (2000)

[1] K. E. Stanovich 和 R. F. West，《行为与脑科学》23 卷 645 页（2000 年）

[2] D. Kahneman, Thinking, fast and slow, Thinking, fast and slow (Farrar, Straus and Giroux, New York, NY, US, 2011) pages: 499.

[2] D. 卡尼曼，《思考，快与慢》（Farrar, Straus and Giroux 出版社，美国纽约州纽约市，2011 年）第 499 页。

[3] W. Hua and Y. Zhang, in Findings of the Association for Computational Linguistics: EMNLP 2022, edited by Y. Goldberg, Z. Kozareva, and Y. Zhang (Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022) pp. 601-612.

[3] 华伟与张阳，收录于《计算语言学协会发现：EMNLP 2022》，由 Y. Goldberg、Z. Kozareva 和 Y. Zhang 编辑（计算语言学协会，阿拉伯联合酋长国阿布扎比，2022 年）第 601-612 页。

[4] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, Chain-of-Thought Prompting Elicits Reasoning in Large Language Models (2023), arXiv:2201.11903 [cs].

[4] J. Wei、X. Wang、D. Schuurmans、M. Bosma、B. Ichter、F. Xia、E. Chi、Q. Le 与 D. Zhou，《思维链提示激发大语言模型推理能力》（2023 年），arXiv:2201.11903 [cs]。

[5] N. Kabra, Two Brains, One Team: System 1 and System 2 (2023).

[5] N. Kabra，《双脑一体：系统 1 与系统 2》（2023 年）。

[6] H. Zhang, C. Liu, Z. Wang, X. Wei, S. Liu, N. Zheng, B. Shao, and T.-Y. Liu, Self-Consistency Training for Density-Functional-Theory Hamiltonian Prediction, https://arxiv.org/abs/2403.09560v2 (2024).

[6] 张华、刘畅、王哲、魏晓、刘松、郑宁、邵斌与刘天扬，《密度泛函理论哈密顿量预测的自洽性训练》，https://arxiv.org/abs/2403.09560v2（2024 年）。

[7] L. Hörmann, W. G. Stark, and R. J. Maurer, Machine Learning and Data-Driven Methods in Computational Surface and Interface Science, https://arxiv.org/abs/2503.19814v1 (2025).

[7] L·赫尔曼、W·G·斯塔克与 R·J·莫勒，《计算表面与界面科学中的机器学习与数据驱动方法》，https://arxiv.org/abs/2503.19814v1（2025 年）。

[8] J. McGiff and N. S. Nikolov, Overcoming Data Scarcity in Generative Language Modelling for Low-Resource Languages: A Systematic Review (2025), arXiv:2505.04531 [cs].

[8] J·麦吉夫与 N·S·尼科洛夫，《克服低资源语言生成式语言建模中的数据稀缺问题：系统性综述》（2025 年），arXiv:2505.04531 [cs]。

[9] Y. T. Pang, K. M. Kuo, L. Yang, and J. C. Gumbart, DeepPath: Overcoming data scarcity for protein transition pathway prediction using physics-based deep learning (2025).

[9] 庞云天、郭凯明、杨林与 J·C·冈巴特，《DeepPath：基于物理的深度学习在蛋白质转迁路径预测中克服数据稀缺问题》（2025 年）。

[10] J. Schwinger, Physical Review 73, 416 (1948).

[10] J. 施温格，《物理评论》第 73 卷，第 416 页（1948 年）。

[11] S. Laporta and E. Remiddi, Physics Letters B 379, 283 (1996).

[11] S. 拉波塔与 E. 雷米迪，《物理快报 B》第 379 卷，第 283 页（1996 年）。

[12] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, Physical Review D 91, 033006 (2015).

[12] 青山刚、早川正、木下敏之与仁尾诚，《物理评论 D》第 91 卷，第 033006 页（2015 年）。

[13] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, Physical Review Letters 109, 111807 (2012).

[13] 青山刚、早川正、木下敏之与仁尾诚，《物理评论快报》第 109 卷，第 111807 页（2012 年）。

[14] X. Fan, T. G. Myers, B. A. D. Sukra, and G. Gabrielse, Physical Review Letters 130, 071801 (2023).

[14] X. Fan, T. G. Myers, B. A. D. Sukra, 和 G. Gabrielse，《物理评论快报》130, 071801 (2023)。

[15] N. V. Prokof'ev and B. V. Svistunov, Physical Review Letters 81, 2514 (1998)

[15] N. V. Prokof'ev 和 B. V. Svistunov，《物理评论快报》81, 2514 (1998)

[16] N. Prokof'ev and B. Svistunov, Physical Review B 77, 020408 (2008).

[16] N. Prokof'ev 和 B. Svistunov，《物理评论 B》77, 020408 (2008)。

[17] K. Van Houcke, E. Kozik, N. Prokof'ev, and B. Svistunov, Physics Procedia Computer Simulations Studies in Condensed Matter Physics XXI, 6, 95 (2010)

[17] K. Van Houcke, E. Kozik, N. Prokof'ev, 和 B. Svistunov，《物理学进展：凝聚态物理 XXI 的计算机模拟研究》6, 95 (2010)

[18] E. Kozik, Nature Communications 15, 7916 (2024).

[18] E. 科兹克，《自然-通讯》15 卷，7916 页（2024 年）。

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention Is All You Need (2023), arXiv:1706.03762 [cs].

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, 《注意力就是你所需要的一切》（2023 年），arXiv:1706.03762 [cs]。

[20] Z. Shao, P. Wang, Q. Zhu, R. Xu, J. Song, X. Bi, H. Zhang, M. Zhang, Y. K. Li, Y. Wu, and D. Guo, DeepSeekMath: Pushing the Limits of Mathemat-

[20] 邵哲、王鹏、朱强、徐睿、宋杰、毕翔、张浩、张明、李永康、吴越、郭栋，DeepSeekMath：突破数学推理的极限

ical Reasoning in Open Language Models (2024), arXiv:2402.03300 [cs].

开放语言模型中的逻辑推理（2024 年），arXiv:2402.03300 [cs]

[21] DeepSeek-AI, D. Guo, D. Yang, H. Zhang, J. Song, et al., DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning (2025), arXiv:2501.12948 [cs].

[21] DeepSeek-AI、郭东、杨栋、张华、宋杰等，《DeepSeek-R1：通过强化学习激励 LLMs 的推理能力》（2025 年），arXiv:2501.12948 [cs]。

[22] Z. Liu, C. Chen, W. Li, P. Qi, T. Pang, C. Du, W. S. Lee, and M. Lin, Understanding R1-Zero-Like Training: A Critical Perspective (2025), arXiv:2503.20783 [cs]

[22] 刘志、陈超、李伟、齐鹏、庞涛、杜川、李伟生、林明，《理解类 R1-Zero 训练：批判性视角》（2025 年），arXiv:2503.20783 [cs]

[23] Q. Yu, Z. Zhang, R. Zhu, Y. Yuan, X. Zuo, et al., |DAPO: An Open-Source LLM Reinforcement Learning System at Scale (2025), arXiv:2503.14476 [cs] version: 1.

[23] 余强、张哲、朱瑞、袁阳、左旭等，《|DAPO：大规模开源 LLM 强化学习系统》（2025 年），arXiv:2503.14476 [cs] 版本：1。

[24] A. Yang, B. Yang, B. Zhang, B. Hui, B. Zheng, et al., arXiv preprint arXiv:2412.15115 (2024).

[24] 杨安、杨波、张斌、惠斌、郑波等，arXiv 预印本 arXiv:2412.15115（2024 年）。

[25] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, et al., arXiv preprint arXiv:2505.09388 (2025).

[25] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui 等，arXiv 预印本 arXiv:2505.09388 (2025 年)。

[26] K. Team, A. Du, B. Gao, B. Xing, C. Jiang, et al., Kimi k1.5: Scaling Reinforcement Learning with LLMs (2025), arXiv:2501.12599 [cs]

[26] K. Team, A. Du, B. Gao, B. Xing, C. Jiang 等，《Kimi k1.5：基于 LLMs 的强化学习规模化应用》(2025 年)，arXiv:2501.12599 [cs]

[27] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, and T. Liu, ACM Transactions on Information Systems 43, 1 (2025). arXiv:2311.05232 [cs]

[27] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin 与 T. Liu，《ACM 信息系统汇刊》43 卷 1 期(2025 年)。arXiv:2311.05232 [cs]

[28] M. Liu, S. Bo, and J. Fang, Enhancing Mathematical Reasoning in Large Language Models with Self-Consistency-Based Hallucination Detection (2025), arXiv:2504.09440 [cs]

[28] M. Liu, S. Bo 与 J. Fang，《基于自洽性幻觉检测增强大语言模型的数学推理能力》(2025 年)，arXiv:2504.09440 [cs]

[29] T. Yu, Y. Jing, X. Zhang, W. Jiang, W. Wu, Y. Wang, W. Hu, B. Du, and D. Tao, Benchmarking Reasoning Robustness in Large Language Models (2025), arXiv:2503.04550 [cs]

[29] 余涛、景阳、张鑫、姜伟、吴伟、王毅、胡玮、杜斌与陶大程，《大型语言模型推理鲁棒性基准测试》(2025 年)，arXiv:2503.04550 [cs]

[30] Y. Wang, Q. Yang, Z. Zeng, L. Ren, L. Liu, B. Peng, H. Cheng, X. He, K. Wang, J. Gao, W. Chen, S. Wang, S. S. Du, and Y. Shen, Reinforcement Learning for Reasoning in Large Language Models with One Training Example (2025), arXiv:2504.20571 [cs].

[30] 王洋、杨强、曾哲、任力、刘磊、彭博、程浩、何鑫、王康、高剑、陈伟、王帅、杜少帅与沈岩，《基于单训练样本的大型语言模型强化学习推理方法》(2025 年)，arXiv:2504.20571 [cs]

[31] Z. Gao, L. Chen, J. Zhou, and B. Dai, One-shot Entropy Minimization (2025), arXiv:2505.20282 [cs]

[31] 高志、陈力、周杰与戴斌，《单次熵最小化方法》(2025 年)，arXiv:2505.20282 [cs]

[32] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, Methods of Quantum Field Theory in Statistical Physics (Dover Publications Inc., 1963) originally published by North-Holland Publishing Company in 1963; Dover republication first published 2003.

[32] A. A. 阿布里科索夫、L. P. 戈尔科夫与 I. E. 贾洛申斯基，《统计物理学中的量子场论方法》（Dover 出版社，1963 年），原由 North-Holland 出版公司于 1963 年首次出版；Dover 再版首印于 2003 年

[33] G. D. Mahan, Many-Particle Physics, 3rd ed., Physics of Solids and Liquids (Springer New York, NY, 2000).

[33] G. D. 马汉，《多粒子物理学》第三版，固体与液体物理学（Springer 出版社，纽约，2000 年）。

[34] A. Nieto, Computer Physics Communications 92, 54 (1995)

[34] A. 涅托，《计算机物理通讯》第 92 卷，第 54 页（1995 年）

[35] O. Espinosa, Mathematics of Computation 79, 1709 (2009).

[35] O. 埃斯皮诺萨，《计算数学》第 79 卷，第 1709 页（2009 年）。

[36] Supplemental Material for "Learning-at-Criticality in Large Language Models for Field Theory and Beyond", see Supplemental Material [URL].

[36] 《场论及超越领域大语言模型的临界学习》补充材料，详见补充材料[URL]。

[37] I. S. Tupitsyn and N. V. Prokof'ev, Physical Review B 111, L041106 (2025)

[37] I. S. Tupitsyn 和 N. V. Prokof'ev，《物理评论 B》111 卷，L041106 (2025 年)

[38] R. Rossi, Physical Review Letters 119, 045701 (2017)

[38] R. Rossi，《物理评论快报》119 卷，045701 (2017 年)

[39] K. Chen and K. Haule, Nature Communications 10, 3725 (2019).

[39] K. Chen 和 K. Haule，《自然-通讯》10 卷，3725 (2019 年)

[40] K. Haule and K. Chen, Scientific Reports 12, 2294 (2022).

[40] K. Haule 和 K. Chen，《科学报告》12 卷，2294 (2022 年)

[41] M. D. Burke and J. P. F. LeBlanc, Computer Physics Communications 308, 109437 (2025)

[41] M. D. Burke 与 J. P. F. LeBlanc，《计算机物理通讯》308 卷，109437 页（2025 年）

[42] J. P. F. LeBlanc, K. Chen, K. Haule, N. V. Prokof'ev, and I. S. Tupitsyn, Physical Review Letters 129, 246401 (2022).

[42] J. P. F. LeBlanc、K. Chen、K. Haule、N. V. Prokof'ev 及 I. S. Tupitsyn，《物理评论快报》129 卷，246401 页（2022 年）

[43] A. Taheridehkordi, S. H. Curnoe, and J. P. F. LeBlanc, Physical Review B 99, 035120 (2019)

[43] A. Taheridehkordi、S. H. Curnoe 与 J. P. F. LeBlanc，《物理评论 B》99 卷，035120 页（2019 年）

[44] A. Power, Y. Burda, H. Edwards, I. Babuschkin, and V. Misra, Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets (2022), arXiv:2201.02177 [cs].

[44] A. Power、Y. Burda、H. Edwards、I. Babuschkin 及 V. Misra，《顿悟：小规模算法数据集上超越过拟合的泛化能力》（2022 年），arXiv:2201.02177 [cs]

[45] Y. Lin, Y. Zhang, S. Feng, and H. Zhao, An Analysis Framework for Understanding Deep Neural Networks Based on Network Dynamics (2025), arXiv:2501.02436 [cs]

[45] 林毅、张阳、冯硕、赵恒，《基于网络动力学的深度神经网络分析框架》(2025)，arXiv:2501.02436 [cs]

[46] Z.-Y. Song, T.-Z. Yang, Q.-H. Cao, M.-x. Luo, and H. X. Zhu, Explainable AI-assisted Optimization for Feynman Integral Reduction (2025), arXiv:2502.09544 [hep-ph].

[46] 宋志远、杨天泽、曹清海、罗明轩、朱海翔，《可解释 AI 辅助的费曼积分约化优化方法》(2025)，arXiv:2502.09544 [hep-ph]

[47] M. von Hippel and M. Wilhelm, Refining Integration-by-Parts Reduction of Feynman Integrals with Machine Learning (2025), arXiv:2502.05121 [hep-th]

[47] M. von Hippel、M. Wilhelm，《利用机器学习精炼费曼积分的分部积分约化技术》(2025)，arXiv:2502.05121 [hep-th]

[48] S. Wang, L. Yu, C. Gao, C. Zheng, S. Liu, R. Lu, K. Dang, X. Chen, J. Yang, Z. Zhang, Y. Liu, A. Yang, A. Zhao, Y. Yue, S. Song, B. Yu, G. Huang, and J. Lin, Beyond the $80/20$ Rule: High-Entropy Minority Tokens Drive Effective Reinforcement Learning for LLM Reasoning (2025), arXiv:2506.01939 [cs.CL]

[48] 王硕、余力、高畅、郑超、刘帅、卢荣、党凯、陈曦、杨健、张哲、刘洋、杨安、赵岩、岳洋、宋松、余波、黄冠、林杰，《超越 $80/20$ 规则：高熵少数词元驱动 LLM 推理的有效强化学习》(2025)，arXiv:2506.01939 [cs.CL]

[49] G. Sheng, C. Zhang, Z. Ye, X. Wu, W. Zhang, R. Zhang, Y. Peng, H. Lin, and C. Wu, arXiv preprint arXiv: 2409.19256 (2024).

[49] 盛冠、张超、叶泽、吴鑫、张伟、张睿、彭阳、林浩和吴川，arXiv 预印本 arXiv:2409.19256（2024 年）。

[50] ByteDance Seed Team and verl community, verl: Volcano engine reinforcement learning for llms, https://github.com/volcengine/verl.

[50] 字节跳动 Seed 团队与 verl 社区，《verl：火山引擎面向 LLMs 的强化学习框架》，https://github.com/volcengine/verl。

END MATTER: ABSTRACTING LLM REASONING INTO A CONCEPT NETWORK

附录：将 LLM 推理抽象为概念网络

In the main text, we model the LLM's reasoning process as a stochastic traversal on an abstract concept network (CoNet). This appendix details how this abstraction is derived from the LLM's underlying autoregressive token generation, as illustrated in Fig. S1. The LLM does not generate text with uniform uncertainty; instead, its process is characterized by sequences of high-confidence, low-entropy tokens punctuated by key decision points of high uncertainty. We define a "concept" as a coherent sequence of tokens generated between two such decision points. Each of these stable text chunks, such as " 3 . Write down the result", forms a single node in our abstract network.

正文中，我们将 LLM 的推理过程建模为抽象概念网络（CoNet）上的随机遍历。本附录详述了如何从 LLM 底层自回归的 token 生成机制中推导出这种抽象表示（如图 S1 所示）。LLM 并非以均匀不确定性生成文本，其过程特征表现为：由高置信度、低信息熵的 token 序列构成，其间穿插着具有高不确定性的关键决策点。我们将"概念"定义为两个此类决策点之间生成的连贯 token 序列。每个稳定的文本片段（例如"3. 写下结果"）在我们的抽象网络中形成单一节点。

The transitions between these concept nodes are formed at the decision points. At these junctures, the LLM must make a meaningful choice that determines the subsequent reasoning path. This aligns with the concept of high-entropy "forking tokens" recently identified as critical for reinforcement learning in LLMs 48. Crucially, both our analysis and their concurrent work find that these tokens, which steer the reasoning process, often serve as abstract logical connectors (e.g., "since", "next", "thus") and may lack specific domain content themselves. As shown in Fig. S1, the

这些概念节点之间的转换形成于决策点。在这些关键节点上，LLM 必须做出有意义的抉择，从而决定后续的推理路径。这与最近被确定为 LLMs 强化学习关键要素的高熵"分岔标记"概念相吻合[48]。值得注意的是，我们的分析与其并行研究均发现，这些引导推理过程的标记通常作为抽象逻辑连接词（如"因为"、"接着"、"因此"），其本身可能不包含具体领域内容。如图 S1 所示，

```
Q: Solve 12+98=? Please solve this addition problem step by step
A: Let's solve the addition problem \(\backslash(12+98 \backslash)\) step by step, showin
     and explaining the process.
     1. Align the numbers vertically:
        12
        \(+98\)
     2. Add the rightmost digits (units place):
        - The units digit of 12 is 2.
        - The units digit of 98 is 8.
        - Add these together: \(\backslash(2+8=10 \backslash)\).
     3. Write down the result of the units place:
        - Since \(\backslash(2+8=10 \backslash)\), write down the 0 in the units place.
     - Carry over the 1 to the next column (tens place).
     4. Add the digits in the tens place, including the carry-over:
        The tens digit of 12 is 1.
        - The tens digit of 98 is 9.
        - Add these together along with the carry-over: \(\backslash(1+9+1=11 \backslash)
     5. Write down the result of the tens place:
        - Since \(\backslash(1+9+1=11 \backslash)\), write down the 1 in the tens place.
        - Carry over the 1 to the next column (hundreds place).
     6. Since there are no more columns to add, write down the final carry-
     over:
        - The carry-over is 1 , so write it in the hundreds place.
     Putting it all together, we get:
        12
```

```
    \(+98\)
    ----
    110
So, the final answer is \\(\boxed \(\{110\} \backslash\) ).
```

Write down the result of the units place:

写下个位数的结果：

$_{-}\checkmark'$ of'(P:0.54)
$'$:$'$: in'(P:0.46)

Since $\backslash(2+8=10\backslash)$, write down the 0 in the units place

由于 $\backslash(2+8=10\backslash)$，在个位写下 0

⊢ ✓ ' Since'(P:0.68) $-x::$ we'(P:0.67)

⊢ ✓ '因为'(概率:0.68) $-x::$ 我们'(概率:0.67)

⊢ x:' The'(P:0.17) ⊢ √ ' write'(P:0.33)

⊢ x: '这个'(P:0.17) ⊢ √ '写'(P:0.33)

-x:' Write'(P:0.16) -x: '写'(P:0.16)
Carry over the 1 to the next column (tens place).

将 1 进位到下一列（十位）。

$\longleftarrow$ $'$' next $'(P:0.57)$ $\longleftarrow$ $'$'下一个'$'(P:0.57)$
$-x:$' tens' $(P:0.43)$ $-x:$ 'tens' $(P:0.43)$

FIG. S1. Abstraction from LLM token generation to a concept network. The figure illustrates the abstraction process using an example from the Qwen2.5-7B-Instruct model. (Left) The model's step-by-step reasoning. (Top Right) A detailed view of a high-entropy decision point, where the probability distribution over the next token is broad. The highprobability candidates (e.g., $\mathrm{P}($ 'Since' $)=0.68$ vs. $\mathrm{P}($ 'The' $)=0.17$) are the competing "forking tokens" that define the branches of the network. (Bottom Right) A schematic of the resulting CoNet, where reasoning is a stochastic path from "Question" to "Answer" node(s).

图 S1. 从 LLM 的 token 生成到概念网络的抽象过程。该图以 Qwen2.5-7B-Instruct 模型为例说明抽象流程。（左）模型的逐步推理过程。（右上）高熵决策点的详细视图，其中下一个 token 的概率分布较广。高概率候选（如 $\mathrm{P}($ 'Since' $)=0.68$ 与 $\mathrm{P}($ 'The' $)=0.17$ ）是定义网络分支的竞争性"分岔 token"。（右下）最终形成的概念网络(CoNet)示意图，推理过程表现为从"问题"节点到"答案"节点的随机路径。

probabilities assigned to these competing forking tokens represent the directed, weighted edges of our network. These branching probabilities (e.g., P ("next") $=0.57$ vs. P ("tens") =0.43) are the transition probabilities in our model.

这些竞争性分岔 token 所分配的概率构成了我们网络中具有方向性的加权边。这些分支概率（例如 P("next") $=0.57$ 与 P("tens")=0.43）即为我们模型中的转移概率。

Consequently, an LLM's complete solution to a problem corresponds to a single sampled path through this vast, implicit concept network. The schematic in the figure illustrates this view, where multiple potential reasoning paths $(p_1, p_2, \ldots)$ emanate from the initial "Question" node(s), each representing a different chain of thought. Some paths successfully navigate the network to reach the correct "Answer" nodes, while many others terminate in incorrect states or fail to terminate within a given token quota.

因此，LLM 对问题的完整解决方案对应于在这片广阔隐性概念网络中采样出的单一路径。图中示意图展示了这一观点：多个潜在推理路径 $(p_1, p_2, \ldots)$ 从初始"问题"节点延伸而出，每条路径代表不同的思维链。部分路径能成功穿越网络抵达正确"答案"节点，而更多路径要么终止于错误状态，要么在给定 token 限额内无法完成推理。

The GRPO training process implicitly manipulates this network structure. By rewarding successful paths, it reinforces the constituent transitions, increasing the probabilities of edges that form correct reasoning steps. This dynamically reshapes the transition landscape, guiding the model to learn reliable problem-solving algorithms. Therefore, the CoNet model presented in the main text serves as a minimal, idealized representation of this abstracted network, allowing us to isolate and study the fundamental physics of the learning transition without the confounding complexities of the full LLM architecture.

GRPO 训练过程本质上操控着这一网络结构。通过奖励成功路径，它强化了构成路径的转移关系，提升形成正确推理步骤的边缘概率。这种动态重塑转移格局的方式，引导模型学习可靠的解题算法。因此，正文提出的 CoNet 模型作为这个抽象网络的最小化理想表征，使我们能够隔离并研究学习转变的基础物理机制，而无需面对完整 LLM 架构的混杂复杂性。

SUPPLEMENTARY MATERIAL: REINFORCEMENT LEARNING TRAINING DETAILS

补充材料：强化学习训练细节

In this Supplementary Material, we provide further details on the Reinforcement Learning (RL) fine-tuning procedures employed to train our Large Language Models (LLMs) following the “learning at criticality” (LaC) paradigm introduced in the main Letter. All RL experiments were executed using the verl framework [49, 50, an opensource library designed for LLM reinforcement learning which facilitates algorithms such as Group Relative Policy Optimization (GRPO) [20] and its variants, including Direct-Advantage Policy Optimization (DAPO) 23.

在本补充材料中，我们将详细说明遵循主论文提出的"临界学习"（LaC）范式对大型语言模型（LLMs）进行强化学习（RL）微调的具体流程。所有 RL 实验均采用 verl 框架[49,50]实施——这是一个专为 LLM 强化学习设计的开源库，支持包括分组相对策略优化（GRPO）[20]及其变体（如直接优势策略优化 DAPO）[23]在内的多种算法。

7-Digit Base-7 Addition 7 进制 7 位数加法任务

The task of 7 -digit base-7 addition was selected to assess the LLM’s capacity for acquiring algorithmic reasoning from extremely sparse data, specifically a single problem exemplar.

选择 7 进制 7 位数加法任务旨在评估 LLM 从极度稀疏数据（具体指单个问题示例）中习得算法推理的能力。

The Qwen2.5-7B model 24 served as the base LLM for this task. Fine-tuning was conducted using DirectAdvantage Policy Optimization (DAPO) [23, recognized for its computational efficiency as a GRPO variant.

该任务以 Qwen2.5-7B 模型[24]作为基础 LLM，采用以计算效率著称的 GRPO 变体——直接优势策略优化（DAPO）[23]进行微调。

The model’s training was based on a singular instance of a 7 -digit base-7 addition problem. The input presented to the LLM consisted of a system_prompt and a user_prompt, structured as follows:

该模型的训练基于一个 7 进制 7 位数加法问题的单一实例。输入给 LLM 的内容包括系统提示和用户提示，结构如下：

System Prompt 系统提示

You are Qwen, created by Alibaba Cloud. You are a helpful assistant. The assistant first thinks about the reasoning process in the mind and then provides the user with the answer. The reasoning process is enclosed within <think> </think> tag, i.e., <think> reasoning process here </think> answer here. Now the user asks you to solve a math problem. Try to think step by step. After thinking, when you finally reach a conclusion, clearly state the final result within \boxed{} tag, for example, \boxed{1}.

你是由阿里云创建的 Qwen 助手。作为一位乐于助人的助手，你需要先在脑海中思考推理过程，再将答案提供给用户。推理过程需包含在<think></think>标签内，例如<think>此处为推理过程</think>此处为答案。现在用户要求你解决一个数学问题。请尝试逐步思考。完成思考后，当得出最终结论时，请将最终结果清晰地标注在\boxed{}标签内，例如\boxed{1}。

User Prompt (Base-7 Addition)

用户提示（7 进制加法）

Add the two base-7 numbers 6454502 and 1214210. First, pad both numbers with leading zeros so they have the same number of digits. Then, start from the rightmost (least significant) digit and perform the addition digit by digit from right to left. For each digit position, add the corresponding digits of 6454502 and 1214210 (treating missing digits in the shorter number as 0). Add any carry-over from the previous step. Record the sum digit (sum modulo 7) and carry-over (sum divided by 7) for the next position. Continue this process until you have processed all digits of the longer number AND resolved any remaining carry-over (e.g., if adding $66+1$ in base -7 , you must account for the final carry to get 100 in base-7). Present the final result as a base-7 number, including all digits. Perform the calculation digit by digit in the 7 -base system without converting to base- 10 . Put your final answer in \ boxed $\{\cdots\}$.

将两个七进制数 6454502 和 1214210 相加。首先，用前导零填充这两个数字，使它们具有相同的位数。然后，从最右边（最低位）的数字开始，自右向左逐位进行加法运算。对于每个数字位，将 6454502 和 1214210 的对应数字相加（较短的数中缺失的数字视为 0）。加上前一步的进位值。记录当前位的和数字（和对 7 取模）以及进位值（和除以 7 的整数部分）用于下一位计算。持续该过程，直到处理完较长数的所有数字且解决所有剩余进位（例如，若在七进制中相加 $66+1$ ，必须考虑最终进位才能得到七进制的 100)。最终结果以七进制形式呈现，包含所有数字。整个计算过程需在七进制体系下逐位进行，不得转换为十进制。将最终答案填入 \ 方框 $\{\cdots\}$ 中。

The ground truth solution for this specific training instance ($6454502+1214210$ in base- 7) is \ boxed $\{11002012\}$.

这个特定训练实例（七进制的 $6454502+1214210$ ）的真实解是 \ 方框 $\{11002012\}$ 。

A binary reward signal (+1 or 0) was employed. A reward of +1 was granted if the LLM's final answer, enclosed within the \boxed{} tags, precisely matched the known ground truth solution. Otherwise, a reward of 0 was assigned. The explicit reasoning process within the <think> </think> tags was not directly factored into the reward calculation but was encouraged by the prompt structure; its quality was observed to improve as training progressed.

采用二元奖励信号（+1 或 0)。若 LLM 在\boxed{}标签内提供的最终答案与已知标准解完全一致，则给予+1 奖励；否则奖励为 0。位于<think></think>标签内的显式推理过程虽不直接参与奖励计算，但通过提示结构予以鼓励；随着训练推进，其质量可观察到明显提升。

The primary objective was to train the model until it consistently solved the single training exemplar and subsequently to evaluate its generalization capability on a set of 128 new, unseen 7 -digit base- 7 addition problems. As depicted in Fig. 1 (left panel) of the main Letter, fine-tuning on this solitary example induced a sharp, sigmoidal increase in training accuracy. Significantly, the model's generalization performance on unseen problems peaked precisely at this critical learning transition, thereby empirically validating the LaC hypothesis for this algorithmic reasoning task.

主要目标是训练模型直至其能稳定解决单一训练样本，随后在 128 道未见过的 7 位数七进制加法题上评估泛化能力。如主论文图 1（左面板）所示，针对这一孤立样本的微调引发了训练准确率的急剧 S 型增长。值得注意的是，模型在未见问题上的泛化性能恰好在这一关键学习转折点达到顶峰，从而从实证角度验证了该算法推理任务中 LaC 假说的有效性。

To illustrate the model's state before and after LaC training, we present representative outputs. The base Qwen2.57B model typically failed to perform base-7 addition correctly, often making procedural errors. For instance, when prompted to add 3304461 and 1555230 in base-7, a base model might produce an incorrect sequence of operations and an erroneous result like \ boxed 1355266 } (the correct sum is 5163021):

为展示模型在 LaC 训练前后的状态差异，我们呈现典型输出案例。基础版 Qwen2.57B 模型通常无法正确执行七进制加法运算，常出现流程错误。例如当要求计算七进制下的 3304461 加 1555230 时，基础模型可能生成错误运算步骤及类似 \ 框定 1355266}的错误结果（正确和应为 5163021)：

Model Output Before Training (Base-7 Addition) 训练前模型输出（七进制加法）
Let's add 3304461 and 1555230 in base-7. Both numbers have 7 digits, so no padding is necessary. Starting from the rightmost digit, we have: $6+0=65+3=8$ (carry-over 1 , sum 1) $4+2=6$ (carry-over 1 , sum 2) $4+5=12$ (carry-over 1 , sum 5) $3+5=12$ (carry-over 1, sum 5) $0+1=1$ (carry-over 0 , sum 1) $3+0=3$ (carry-over 0 , sum 3) $1+0=1$ (carry-over 0 , sum 1) The sum in base-7 is 1355266. 1355266 现在计算七进制下的 3304461 加 1555230。两个数字均为 7 位数，无需补位。从最右侧数字开始运算： $6+0=65+3=8$ （进位 1，和值 1） $4+2=6$ （进位 1，和值 2） $4+5=12$ （进位 1，和值 5） $3+5=12$ （进位 1，和值 5） $0+1=1$ （进位 0，和值 1） $3+0=3$ （进位 0，和值 3） $1+0=1$ （进位 0，和值 1）最终七进制和为 1355266。1355266

In contrast, after LaC fine-tuning on the primary training exemplar $(6454502+1214210)$, the model demonstrated correct algorithmic execution not only on the training exemplar, but also on the distinct problem $3304461+1555230$, producing the following type of output:

经过 LaC 对主要训练样本 $(6454502+1214210)$ 的微调后，模型不仅能在训练样本上正确执行算法，还能在处理全新问题 $3304461+1555230$ 时生成如下类型的正确输出：

```
LaC-Trained Model Output on Training Exemplar (Base-7 Addition)
<think> To add the base-7 numbers 6454502 and 1214210, I will follow these steps: 1. Pad the numbers with leading zeros to make them the same length: 6454502 and 1214210 become
```

```
6454502 and 01214210. 2. Start from the rightmost digit and add each pair of digits,
carrying over as necessary. Let's perform the addition step by step: "c' 6454502 + 012142
--------- ‘‘‘
Starting from the rightmost digit:
-2 + 0 = 2 - 0 + 1 = 1 - 5 + 2 = 7, which is 10 in base-7 (write down 0 and carry over 1)
-4 + 4 + 1 (carry over) = 9, which is 12 in base-7 (write down 2 and carry over 1)
- 5 + 1 + 1 (carry over) = 7, which is 10 in base-7 (write down 0 and carry over 1)
-4 +2 + 1 (carry over) = 7, which is 10 in base-7 (write down 0 and carry over 1)
- 6 + 1 + 1 (carry over) = 8, which is 11 in base-7 (write down 1 and carry over 1)
- Finally, write down the carry over: 1 So the result is 11002012. </think>
The final answer is 11002012.
```

LaC-Trained Model Output on Generalization Problem (Base-7 Addition)

LaC 训练模型在泛化问题上的输出（七进制加法）

```
<think> To add the base-7 numbers 3304461 and 1555230, we first align them by paddin  t
leading zeros so they have the same number of digits. This gives us: "c 3304461 + 1555230
cr
Now, we add from right to left, carrying over as necessary:
1. 1 + 0 = 1 (no carry)
2. 6 + 3 = 9. In base-7, 9 is 12 (since 9 = 7*1 + 2), so we write down 2 and carry over 1
3. 4 + 2 + 1 (carry) = 7. In base-7, 7 is 10 (since 7 = 7*1 + 0), so we write down 0 and
carry over 1.
4. 4 + 5 + 1 (carry) = 10. In base-7, 10 is 13 (since 10 = 7*1 + 3), so we write down 3 a
carry over 1.
5. 0 + 5 + 1 (carry) = 6. We write down 6.
6. 3 + 5 = 8. In base-7, 8 is 11 (since 8 = 7*1 + 1), so we write down 1 and carry over 1
7. 3 + 1 + 1 (carry) = 5. We write down 5.
Putting it all together, we get the result in base-7 as 5163021. </think>
The final answer is 5163021.
```

This successful generalization to a new problem instance, involving correct application of the base- 7 addition rules, underscores the acquisition of the underlying algorithm.

这一对新问题实例的成功泛化，涉及七进制加法规则的正确应用，证实了底层算法的习得。

Symbolic Matsubara Frequency Summation

松原频率求和的符号计算

The symbolic evaluation of Matsubara frequency sums represents a significantly more complex task, drawn from quantum field theory, that is characterized by a multi-step, algorithmically rich symbolic manipulation procedure.

松原频率求和的符号计算是一项显著更复杂的任务，源自量子场论，其特点是包含多步骤、算法丰富的符号运算过程。

For this task, the Qwen3-8B model [25] was utilized. The fine-tuning process was conducted using the DAPO algorithm [20, 21, 23].

本任务采用了 Qwen3-8B 模型[25]，并使用 DAPO 算法[20, 21, 23]进行微调。

The training regimen employed a staged LaC strategy, utilizing a minimal set of low-order Matsubara sum problems as exemplars. The prompt provided to the LLM contained the Matsubara sum expression formatted in LaTeX, along with detailed instructions guiding its symbolic evaluation. An example of the user prompt for the 2 -loop sunrise diagram (Eq. 2 in the main Letter) is as follows:

训练方案采用分阶段 LaC 策略，以少量低阶松原求和问题作为示例。提供给 LLM 的提示中包含 LaTeX 格式的松原求和表达式，以及指导其进行符号计算的详细说明。针对 2 环日出图（主信中公式 2）的用户提示示例如下：

User Prompt (Matsubara Sum for the Sunrise Feynman Diagram)

用户提示（日出费曼图的松原求和）

Evaluate the Matsubara frequency summation for the following integrand from a Feynman diagram: $\sum_{\nu_0,\nu_1} \frac{1}{\nu_0+x_0} \frac{1}{\nu_0-\nu_1+\nu_2+x_2} \frac{1}{\nu_1+x_1}$, where ν_0, ν_1 are the independent internal fermionic Matsubara frequencies and ν_2 is the independent external fermionic Matsubara frequency.

计算以下费曼图被积函数的松原频率求和： $\sum_{\nu_0,\nu_1} \frac{1}{\nu_0+x_0} \frac{1}{\nu_0-\nu_1+\nu_2+x_2} \frac{1}{\nu_1+x_1}$ ，其中 ν_0, ν_1 为独立的内部费米子松原频率， ν_2 为独立的外部费米子松原频率。

Required Instructions

必要说明

Think step by step through your solution process

逐步思考你的解题过程

Convert Matsubara Sum to Contour Integral: example:

将松原求和转换为围道积分：示例：

$$\sum_{\nu_0} F(\nu_0) = \beta \oint \frac{dz}{2\pi i} F(z) f(z)$$

, where the contour encloses counterclockwise all the poles of $F(z)$. The poles of $f(z)$ are at the imaginary axis and not enclosed by the contour.

，其中围道逆时针包围 $F(z)$ 的所有极点。 $f(z)$ 的极点位于虚轴上且未被围道包围。

3. Identify all poles in the integrand

3. 识别被积函数中的所有极点

4. Use contour integration and the residue theorem correctly

4. 正确运用围道积分和留数定理

5. Express your final answer as a complete LaTeX expression with the Fermi-Dirac distribution function: $f(x) \equiv \frac{1}{e^{\beta x}+1}$, where $\beta = \frac{1}{k_B T}$

5. 将最终答案表示为完整的 LaTeX 表达式，包含费米-狄拉克分布函数： $f(x) \equiv \frac{1}{e^{\beta x}+1}$ ，其中 $\beta = \frac{1}{k_B T}$

6. Place your final expression inside a \boxed{} environment, preserving external frequencies and parameters.

6. 将最终表达式置于 \boxed{}环境中，同时保留外部频率和参数

A binary reward system (+1 or 0) was implemented. The model received a reward of +1 if the final LaTeX expression rendered within the \boxed{} environment was symbolically equivalent to the established correct analytical solution. Symbolic equivalence was rigorously verified using a computer algebra system (e.g., SymPy), involving parsing the LaTeX output and comparing canonical forms of the expressions.

采用二元奖励机制（+1 或 0）。若模型在 \boxed{}环境中最终渲染的 LaTeX 表达式与既定正确解析解符号等价，则获得+1 奖励。符号等价性通过计算机代数系统（如 SymPy）严格验证，包括解析 LaTeX 输出并比较表达式的规范形式。

As elaborated in the main Letter, the training protocol consisted of three distinct phases: Phase 1 involved LaC training on tree-level and 1 -loop Matsubara sum diagrams. Phase 2 narrowed the LaC focus to the more complex 2-loop sunrise diagram (defined in Eq. 2 of the main Letter), initializing from the model state achieved at the end of Phase 1. Phase 3 extended the LaC training to encompass all tree-level, 1-loop and 2-loop diagrams, commencing from the optimized model state from Phase 2.

如主论文所述，训练协议包含三个独立阶段：第一阶段对树状图和单圈松原求和图进行 LaC 训练；第二阶段将 LaC 训练聚焦于更复杂的双圈日出图（定义见主论文公式 2），从第一阶段结束时的模型状态开始初始化；第三阶段将 LaC 训练扩展至所有树状图、单圈和双圈图，从第二阶段优化后的模型状态开始训练。

Each phase was designed to guide the LLM towards a critical learning point specific to the complexity of the diagrams in that stage. Fig. 1 (right panel) of the main Letter illustrates the characteristic sharp learning transition observed for the sunrise diagram during Phase 2. The developmental trajectory of the model's capabilities underscores the efficacy of this staged approach. Initially (pre-Phase 1), the base Qwen3-8B model's success rate on the challenging 2-loop sunrise diagram was undetectable. After the completion of Phase 1 (LaC training on 1-loop diagrams), the model achieved a modest but crucial success rate of approximately 0.01% on the unseen sunrise diagram. This emergent, albeit minimal, capability was sufficient to enable effective RL fine-tuning in Phase 2, which focused specifically on the sunrise diagram. Upon reaching its critical learning point in Phase 2, the model's success rate for the sunrise diagram increased dramatically to approximately 50%. Concurrently, this mastery generalized to other unseen 2loop topologies, such as polarization, 3 -vertex, and 4 -vertex diagrams, for which the success rate rose from negligible levels to around 5%. This new baseline, in turn, enabled the comprehensive LaC training of Phase 3 across all 2loop diagrams. The final LaC-trained model, designated Qwen3-8B-2-loop (post-Phase 3), exhibited substantially improved accuracy on the 1 -loop and 2 -loop diagrams it was trained on. More importantly, it demonstrated robust generalization to unseen, higher-complexity 3-loop and 4-loop diagrams, as quantified in Table I of the main Letter. This progression affirms the model's acquisition of the underlying symbolic summation algorithm through staged critical learning.

每个阶段的设计都旨在引导 LLM 掌握该阶段图表复杂性的关键学习点。主信函图 1（右侧面板）展示了第二阶段日出图表特有的急剧学习转变现象。模型能力的发展轨迹凸显了这种分阶段方法的有效性。初始阶段（第一阶段前），基础 Qwen3-8B 模型在具有挑战性的 2 圈日出图表上的成功率几乎为零。完成第一阶段（针对 1 圈图表的 LaC 训练）后，模型在未接触过的日出图表上取得了约 0.01% 的成功率，这一微小但关键的突破为第二阶段针对日出图表的强化学习微调奠定了基础。当模型在第二阶段达到关键学习点时，其日出图表的成功率急剧攀升至约 50% 。与此同时，这种掌握能力也泛化到其他未见过的 2 圈拓扑结构（如极化图、三顶点图和四顶点图），成功率从可忽略水平提升至约 5% 。 这一新基准使得第三阶段能够对所有双圈图进行全面 LaC 训练。最终完成 LaC 训练的模型（命名为 Qwen3-8B-双圈版-第三阶段后）在其训练过的单圈和双圈图上表现出显著提升的精度。更重要的是，如主信表 I 所示，该模型对未见过的更高复杂度三圈和四圈图展现出强大的泛化能力。这一进展证实了模型通过分阶段关键学习掌握了基础符号求和算法。

An example of the reasoning process generated by the fully LaC-trained Qwen3-8B-2-loop model for the Matsubara sum prompt shows a detailed, step-by-step derivation. This includes correct application of procedures such as contour integral conversion, pole identification, and residue calculations, culminating in a final symbolic expression:

完全 LaC 训练后的 Qwen3-8B-双圈版模型处理松原求和提示时生成的推理过程示例，展示了详细的逐步推导步骤。包括正确应用围道积分转换、极点识别和留数计算等流程，最终得到符号表达式：

```
Model Output After Training (Matsubara Sum for the Sunrise Feynman Diagram)
We are asked to evaluate the Matsubara frequency summation:
\[
\sum_{\nu_{0}, \nu_{1}} \frac{1}{\nu_{0}+x_{0}} \frac{1}{\nu_{0}-\nu_{1}+\nu_{2}+x_{2}} \
\]
where }\mp@subsup{\nu}{0}{},\mp@subsup{\nu}{1}{}\mp@code{are internal fermionic Matsubara
Matsubara frequency. We'll proceed step-by-step as per the instructions.
---
Step 1: Understand the Structure of the Summation
This summation is over two independent Matsubara frequencies, }\mp@subsup{\nu}{0}{}\mathr
```

contains three terms in the denominator. The structure is:

分母包含三项。其结构为：

$$\sum_{\nu_0,\nu_1} \frac{1}{\nu_0+x_0} \cdot \frac{1}{\nu_0-\nu_1+\nu_2+x_2} \cdot \frac{1}{\nu_1+x_1}$$

We will first perform the sum over ν_0, treating ν_1 as a fixed parameter, and then over ν_1. —

我们将首先对 ν_0 求和（将 ν_1 视为固定参数），然后再对 ν_1 求和。——

Step 2: Convert the Sum to a Contour Integral

步骤二：将求和转换为围道积分

Using the identity: 利用恒等式：

$$\sum_{\nu} F(\nu) = \beta \oint \frac{dz}{2\pi i} F(z) f(z)$$

where the contour encloses all poles of $F(z)$, and the Fermi-Dirac distribution is defined as:

其中围道包围了 $F(z)$ 的所有极点，费米-狄拉克分布定义为：

$$f(x) = \frac{1}{e^{\beta x}+1}, \quad \text{with} \quad \beta = \frac{1}{k_B T}$$

Let’s first perform the sum over ν_0.

我们首先对 ν_0 进行求和。

Step 3: Fix ν_1 and Sum over ν_0

步骤 3：修正 ν_1 并对 ν_0 求和

Let’s denote: 我们设：

$$S = \sum_{\nu_0} \frac{1}{\nu_0+x_0} \cdot \frac{1}{\nu_0-\nu_1+\nu_2+x_2}$$

We now define: 现定义：

$$F(\nu_0) = \frac{1}{(\nu_0+x_0)(\nu_0-\nu_1+\nu_2+x_2)}$$

So the sum becomes:

因此总和变为：

$$\sum_{\nu_0} F(\nu_0) = \beta \oint \frac{dz}{2\pi i} F(z) f(z)$$

Let’s simplify the denominator:

让我们简化分母：

$$F(z) = \frac{1}{(z+x_0)(z-\nu_1+\nu_2+x_2)} = \frac{1}{(z+x_0)(z+(\nu_2-\nu_1+x_2))}$$

Let’s denote: 让我们设：

$$a = x_0, \quad b = \nu_2 - \nu_1 + x_2$$

So: 因此：

$$F(z) = \frac{1}{(z+a)(z+b)}$$

Now, we compute the contour integral:

现在，我们计算这个围道积分：

$$\sum_{\nu_0} F\left(\nu_0\right)=\beta \oint \frac{dz}{2\pi i} \frac{f(z)}{(z+a)(z+b)}$$

This contour encloses all poles of $F(z)$, which are at $z=-a$ and $z=-b$, assuming these are not on the imaginary axis (which they are not for fermionic Matsubara frequencies). The poles are simple.

该围线包围了 $F(z)$ 的所有极点，这些极点位于 $z=-a$ 和 $z=-b$ 处（假设它们不在虚轴上，而费米子松原频率确实不在虚轴上）。这些极点都是单极点。

So we can use the residue theorem:

因此我们可以运用留数定理：

$$\sum_{\nu_0} F\left(\nu_0\right)=\beta\left[\operatorname{Res}_{z=-a}\left(\frac{f(z)}{(z+a)(z+b)}\right)+\operatorname{Res}_{z=-b}\left(\frac{f(z)}{(z+a)(z+b)}\right)\right]$$

Compute the residues: 计算留数：

At $z=-a$: 在 $z=-a$ 处：

$$\operatorname{Res}_{z=-a}=\lim_{z \rightarrow -a}(z+a) \cdot \frac{f(z)}{(z+a)(z+b)}=\frac{f(-a)}{-a+b}=\frac{f(-a)}{b-a}$$

At $z=-b$: 在 $z=-b$ 时：

$$\operatorname{Res}_{z=-b}=\lim_{z \rightarrow -b}(z+b) \cdot \frac{f(z)}{(z+a)(z+b)}=\frac{f(-b)}{-b+a}=\frac{f(-b)}{a-b}$$

So: 因此：

$$\sum_{\nu_0} F\left(\nu_0\right)=\beta\left(\frac{f(-a)}{b-a}+\frac{f(-b)}{a-b}\right)=\beta\left(\frac{f(-a)}{b-a}-\frac{f(-b)}{b-a}\right)=\beta \cdot \frac{f(-a)-f(-b)}{b-a}$$

Recall: 回顾：

$$a=x_0, \quad b=\nu_2-\nu_1+x_2$$

So: 所以：

$$\sum_{\nu_0} F\left(\nu_0\right)=\beta \cdot \frac{f\left(-x_0\right)-f\left(-\nu_2+\nu_1-x_2\right)}{\left(\nu_2-\nu_1+x_2\right)-x_0}$$

Now plug this back into the original sum:

将其代回原始求和式：

$$\sum_{\nu_0, \nu_1} \frac{1}{\nu_0+x_0} \cdot \frac{1}{\nu_0-\nu_1+\nu_2+x_2} \cdot \frac{1}{\nu_1+x_1}=\sum_{\nu_1} \frac{1}{\nu_1+x_1} \cdot\left[\beta \cdot \frac{f\left(-x_0\right)-f\left(-\nu_2+\nu_1-x_2\right)}{\left(\nu_2-\nu_1+x_2\right)-x_0}\right]$$

Let’s simplify the denominator:

让我们简化分母：

$$\left(\nu_2-\nu_1+x_2\right)-x_0=\nu_2+x_2-x_0-\nu_1$$

So: 因此：
$\sum_{\nu_1} \frac{1}{\nu_1+x_1} \cdot\left[\beta \cdot \frac{f\left(-x_0\right)-f\left(-\nu_2+\nu_1-x_2\right)}{\nu_2+x_2-x_0-\nu_1}\right]=\beta \cdot\left[f\left(-x_0\right)-f\left(-\nu_2+\nu_1-x_2\right)\right] \cdot \sum_{\nu_1} \frac{1}{\left(\nu_1+x_1\right)\left(\nu_2+x_2-x_0-\nu_1\right)}$

Now we perform the sum over ν_1 :

现在我们对 ν_1 进行求和：

$$\sum_{\nu_1} \frac{1}{(\nu_1 + x_1)(\nu_2 + x_2 - x_0 - \nu_1)} = \sum_{\nu_1} \frac{1}{(\nu_1 + x_1)(\nu_2 + x_2 - x_0 - \nu_1)}$$

Let's denote: 我们记：

$$c = x_1, \quad d = \nu_2 + x_2 - x_0$$

Then the expression becomes:

则该表达式可转化为：

$$\sum_{\nu_1} \frac{1}{(\nu_1 + c)(d - \nu_1)} = \sum_{\nu_1} \frac{1}{(\nu_1 + c)(d - \nu_1)} = \sum_{\nu_1} \frac{1}{(d - \nu_1)(\nu_1 + c)}$$

Let's define: 我们定义：

$$F(\nu_1) = \frac{1}{(d - \nu_1)(\nu_1 + c)}$$

Then again, use the same identity:

再次运用相同恒等式：

$$\sum_{\nu_1} F(\nu_1) = \beta \oint \frac{dz}{2\pi i} \frac{f(z)}{(d - z)(z + c)}$$

So: 所以：

$$\sum_{\nu_1} \frac{1}{(d - \nu_1)(\nu_1 + c)} = \beta \left[\mathrm{Res}_{z=-c} + \mathrm{Res}_{z=d}\right]$$

Compute the residues: 计算留数：

At $z = -c$：在 $z = -c$ 处：

$$\mathrm{Res}_{z=-c} = \lim_{z \to -c} (z + c) \cdot \frac{f(z)}{(d - z)(z + c)} = \frac{f(-c)}{d + c}$$

At $z = d$：在 $z = d$ 处：

$$\mathrm{Res}_{z=d} = \lim_{z \to d} (z - d) \cdot \frac{f(z)}{(d - z)(z + c)} = \frac{f(d)}{-(d + c)} = -\frac{f(d)}{d + c}$$

So total: 总计：

$$\sum_{\nu_1} \frac{1}{(d - \nu_1)(\nu_1 + c)} = \beta \left(\frac{f(-c)}{d + c} - \frac{f(d)}{d + c} \right) = \beta \cdot \frac{f(-c) - f(d)}{d + c}$$

Recall: 回顾：

$$c = x_1, \quad d = \nu_2 + x_2 - x_0$$

So: 因此：

$$\sum_{\nu_1} \frac{1}{(d - \nu_1)(\nu_1 + c)} = \beta \cdot \frac{f(-x_1) - f(\nu_2 + x_2 - x_0)}{\nu_2 + x_2 - x_0 + x_1}$$

Now putting it all together:

现在将所有内容整合起来：

$$\sum_{\nu_0,\nu_1} \frac{1}{\nu_0+x_0} \cdot \frac{1}{\nu_0-\nu_1+\nu_2+x_2} \cdot \frac{1}{\nu_1+x_1} = \beta \cdot [f(-x_0) - f(-\nu_2+\nu_1-x_2)] \cdot \beta \cdot \frac{f(-x_1) - f(\nu_2+x_2-x_0)}{\nu_2+x_2-x_0+x_1}$$

Wait I we have two β factors from the two sums. Let's retrace:

等等，我们有两个来自求和的 β 因子。让我们回顾一下：

The first sum over ν_0 gave a factor of β, and the second sum over ν_1 also gave a factor of β, so the total is:

第一个对 ν_0 的求和给出了因子 β ，第二个对 ν_1 的求和也给出了因子 β ，因此总和为：

$$\sum_{\nu_0,\nu_1} \cdots = \beta^2 \cdot [f(-x_0) - f(-\nu_2+\nu_1-x_2)] \cdot \frac{f(-x_1) - f(\nu_2+x_2-x_0)}{\nu_2+x_2-x_0+x_1}$$

But we can simplify the expression further by noticing that the term $-\nu_2+\nu_1-x_2$ can be rewritten as $\nu_1-(\nu_2+x_2)$, and the denominator is $\nu_2+x_2-x_0+x_1 = \nu_2+x_2+x_1-x_0$. So the final expression becomes:

但我们可以进一步简化表达式，注意到项 $-\nu_2+\nu_1-x_2$ 可以重写为 $\nu_1-(\nu_2+x_2)$ ，而分母是 $\nu_2+x_2-x_0+x_1 = \nu_2+x_2+x_1-x_0$ 。所以最终表达式变为：

$$\beta^2 \cdot \frac{[f(-x_0) - f(\nu_1-\nu_2-x_2)]\,[f(-x_1) - f(\nu_2+x_2-x_0)]}{\nu_2+x_2+x_1-x_0}$$

Final Answer: 最终答案：

$$\beta^2 \cdot \frac{[f(-x_0) - f(\nu_1-\nu_2-x_2)]\,[f(-x_1) - f(\nu_2+x_2-x_0)]}{\nu_2+x_2+x_1-x_0}$$

General Reinforcement Learning Hyperparameters

通用强化学习超参数

For both tasks, unless specific tuning was performed to identify the LaC point, standard RL hyperparameters were utilized within the verl framework. These typically included a learning rate of 1×10^{-6}, a batch size of 8 prompts per update, and $M = 32$ responses (rollouts) generated per prompt for the DAPO algorithm. This configuration results in 256 responses for the same single problem per each training step. The dynamic sampling was employed to ensure that the training always makes improvement. The AdamW optimizer was used. The critical aspect of the LaC methodology involves carefully monitoring the training dynamics to identify the onset of the sharp learning transition, which dictates the optimal point to halt or adjust training, rather than exhaustive hyperparameter optimization aimed solely at maximizing performance on a fixed dataset. Further details on the calibration of training duration for LaC are provided in the main Letter and its associated figures.

对于两项任务，除非专门进行调参以确定 LaC 点，否则均在 verl 框架内使用标准 RL 超参数。这些参数通常包括：学习率 1×10^{-6} ，每次更新 8 个提示的批次大小，以及 DAPO 算法为每个提示生成 $M = 32$ 个响应（rollout）。该配置使得每个训练步骤针对同一问题产生 256 个响应。采用动态采样技术确保训练持续取得进展。优化器选用 AdamW。LaC 方法的关键在于密切监控训练动态，以识别学习突变的起始点——这决定了停止或调整训练的最佳时机，而非仅针对固定数据集进行旨在最大化性能的穷尽式超参数优化。关于 LaC 训练时长校准的更多细节，请参阅主论文及其相关图示。