

Artificial Intelligence Newsletter

人工智能 前沿快报

2025 年第 09 期
总第 27 期



学会官方公众号

广东省图象图形学会

GDSIG

学术交流、技术咨询、科学普及

目录

学术动态

一. Mastering diverse control tasks through world models.....	1
二. YOLOv12: Attention-Centric Real-Time Object Detectors.....	3
三. Foundation model of neural activity predicts response to new stimulus types.....	5
四. LLMs are Bayesian, In Expectation, Not in Realization.....	7
五. Neural-Driven Image Editing.....	9
六. Disentangling the Factors of Convergence between Brains and Computer Vision Models.....	11
七. R-4B: Incentivizing General-Purpose Auto-Thinking Capability in MLLMs via Bi-Mode Annealing and Reinforce Learning.....	13
八. Implicit Reasoning in Large Language Models: A Comprehensive Survey.....	15
九. The Landscape of Agentic Reinforcement Learning for LLMs: A Survey.....	17
十. REFRAG: Rethinking RAG based Decoding.....	19
十一. Attention as an Adaptive Filter.....	21
十二. Why Language Models Hallucinate.....	23
十三. A Comprehensive Survey on Trustworthiness in Reasoning with Large Language Models.....	24
十四. Towards a Unified View of Large Language Model Post-Training.....	27
十五. An AI system to help scientists write expert-level empirical software.....	29
十六. K2-Think: A Parameter-Efficient Reasoning System.....	31
十七. A Survey of Reinforcement Learning for Large Reasoning Models.....	33
十八. Teaching LLMs to Plan: Logical Chain-of-Thought Instruction Tuning for Symbolic Planning.....	35
十九. Single-stream Policy Optimization.....	37
二十. AToken: A Unified Tokenizer for Vision.....	39
二十一. Discovery of Unstable Singularities.....	40
二十二. MANZANO: A Simple and Scalable Unified Multimodal Model with a Hybrid Vision Tokenizer.....	43
二十三. Open-World Drone Active Tracking with Goal-Centered Rewards.....	45

目录

技术动态

一. 英伟达开源通用深度研究大模型 Universal Deep Research	47
二. 腾讯开源 WTM2025 冠军大模型	48
三. 阿里推出最新旗舰模型 Qwen3-Max-Preview	49
四. Google 发布面向嵌入式应用的大模型 EmbeddingGemma	51
五. 类脑脉冲大模型“瞬悉 1.0”发布	52
六. 通义千问发布 Qwen3 ASR 模型：智能语音识别新标杆，支持歌声与定制化	53
七. 腾讯开源图像生成大模型混元 Image2.1	55
八. 月之暗面 Kimi-K2 大模型正式开源	56
九. 英伟达发布 Rubin-CPX，为大规模上下文推理量身打造的新型 GPU	57
十. 阿里巴巴发表 MOE 新模型 Qwen3-Next 系列	58
十一. Meta 发布 MobileLLM-R1 端侧模型	60
十二. World Labs 发布 3D 世界生成模型 Marble	61
十三. 通义千问 Qwen3-Omni 开源，引领原生全模态 AI 新范式	63
十四. 阿里通义大模型家族七连发，Qwen3-Max 领衔刷新 SOTA	64

资源分享

一. AI Agents for Beginners - A Course	66
---------------------------------------	----

INTRODUCTION

导读

在人工智能技术飞速发展的时代背景下，我们将不定期梳理人工智能领域的部分前沿学术与技术动态，并以简报的形式分享给读者，以帮助关注此领域的人士了解最新的学术前沿发展与产业技术动态，促进学术交流和技術繁荣。本期摘选了近期全球人工智能领域的 23 篇最新学术动态、14 篇技术动态、以及 1 个资源推荐给广大读者。

ACADEMIC TRENDS

学术动态

● 论文一

Mastering diverse control tasks through world models

日期

2025-04-02

Article

Mastering diverse control tasks through world models

<https://doi.org/10.1038/s41586-025-08744-2>

Danijar Hafner^{1,2}, Jurgis Pasukonis¹, Jimmy Ba² & Timothy Lillicrap¹

Received: 5 April 2024

Accepted: 5 February 2025

Published online: 2 April 2025

Open access

Developing a general algorithm that learns to solve tasks across a wide range of applications has been a fundamental challenge in artificial intelligence. Although current reinforcement-learning algorithms can be readily applied to tasks similar to what they have been developed for, configuring them for new application domains

内容摘要

本文提出了第三代 Dreamer 算法，一种通用的模型为中心（model-based）强化学习方法，旨在以固定超参数在大量不同领域任务中表现优于专门调优的算法。Dreamer 通过学习环境的潜在表示（world model）并在潜在空间中“想象”（即模拟未来情景）来规划行动。为提高跨任务的稳定性与泛化性，作者设计了归一化、平衡机制和变换技术等一系列鲁棒性策略，使模型能在视觉输入、连续控制、离散动作、稀疏奖励等多种环境下稳定训练。实验证明，Dreamer 在超过 150 个控制任务上，使用同一套超参数就能超越多个为特定域调优的最先进方法。最具标志性的是，Dreamer 是首次在完全从零开始、不依赖人类示范或课程策略的情况

下，在 Minecraft 环境中自主收集钻石的算法。该工作表明，用一个统一的模型即可处理极为多样的强化学习任务，降低超参数调试成本，扩大 RL 在现实任务中的可适用性。

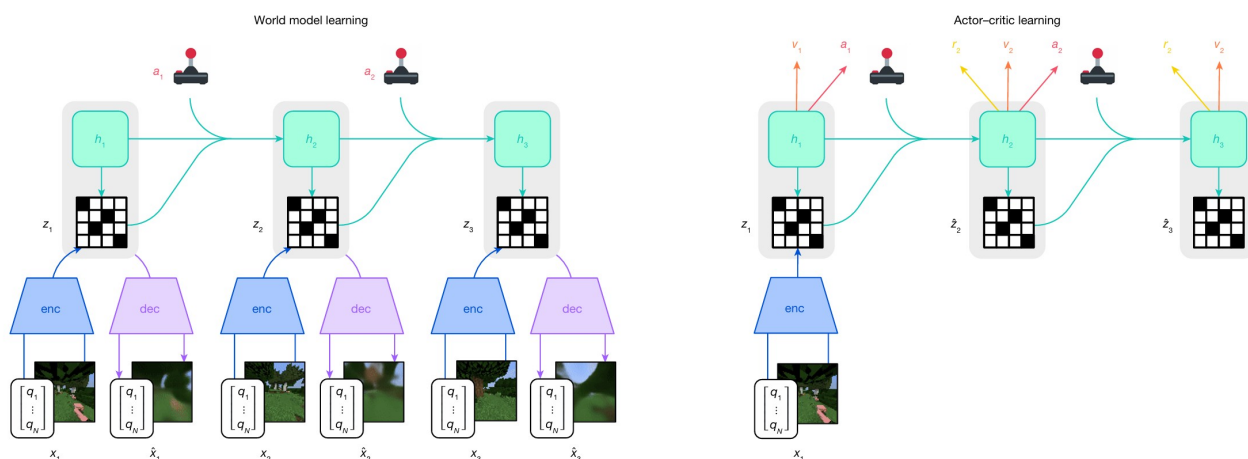


Fig. 1 | Training process of Dreamer. The world model encodes sensory inputs x_t using the encoder (enc) into discrete representations z_t that are predicted by a sequence model with recurrent state h_t given actions a_t . The inputs are

reconstructed as $\hat{x}_t$ using the decoder (dec) to shape the representations. The actor and critic predict actions a_t and values v_t and learn from trajectories of abstract representations $\hat{z}_t$ and rewards r_t predicted by the world model.

主要亮点

- **跨领域的一致性能：** Dreamer 在超过 150 个环境任务上，以固定的超参数配置，就能超越各个领域里为之专门调优的最先进算法，体现了其极高的通用性。
- **首次从零开始在 Minecraft 中完成挑战性任务：** 在 Minecraft 这样一个具有稀疏奖励、长远规划需求、开放世界探索性质的环境中，Dreamer 无需任何人类示范或课程设计，即能自主学会收集钻石，被认为是 AI 控制领域的重大里程碑。
- **鲁棒性机制与潜在空间推演策略：** 为使模型在各种环境条件下学习稳定，作者引入了多种鲁棒性技术（如归一化、损失项平衡、变换策略等），并让 actor/critic 思考未来情景（模拟潜在轨迹）以优化策略，从而在潜在表示空间中高效规划行为。

相关链接

Paper: <https://www.nature.com/articles/s41586-025-08744-2.pdf>

Github: <https://github.com/danijar/dreamerv3>

● 论文二

YOLOv12: Attention-Centric Real-Time Object Detectors

日期

2025-02-18

YOLOv12: Attention-Centric Real-Time Object Detectors

Yunjie Tian
University at Buffalo
yunjiet1@buffalo.edu

Qixiang Ye
University of Chinese Academy of Sciences
qxye@ucas.ac.cn

David Doermann
University at Buffalo
doermann@buffalo.edu

内容摘要

本文介绍了 YOLOv12，一个以注意力机制为核心的实时目标检测框架。该模型旨在解决传统上注意力模型速度不及 CNN-based 模型的问题，在保持高准确率的同时，实现了与现有 CNN-based 模型相当的推理速度。YOLOv12 引入了 Area Attention (A2) 模块和 Residual Efficient Layer Aggregation Networks (R-ELAN) 等创新，以优化计算效率和训练稳定性。在实验中，YOLOv12 在标准目标检测基准上表现出色，其性能超越了 YOLOv10/11-N 和 RT-DETR/RT-DETRv2 等流行模型。



Figure 2. **Comparison of the representative local attention mechanisms with our area attention.** Area Attention adopts the most straightforward equal partitioning way to divide the feature map into l areas vertically or horizontally. (default is 4). This avoids complex operations while ensuring a large receptive field, resulting in high efficiency.

主要亮点

- **注意力机制与速度的平衡**: YOLOv12 框架开创性地将注意力机制融入 YOLO 系列, 同时保持了与传统 CNN 模型相媲美的速度, 打破了 CNN 在此领域的长期主导地位。
- **创新的模块化设计**: 论文引入了 Area Attention (A2) 模块, 它能在降低计算复杂性的同时保持大感受野。此外, 还提出了 Residual Efficient Layer Aggregation Networks (R-ELAN), 通过残差连接和优化的特征聚合方法, 解决了大规模模型训练中的稳定性问题。
- **全面的卓越性能**: YOLOv12 在准确性和速度之间取得了显著的平衡。例如, YOLOv12-N 在 T4 GPU 上实现了 40.6% 的 mAP, 推理延迟仅为 1.64 毫秒, 在性能上超越了 YOLOv10-N 和 YOLOv11-N, 同时还支持目标检测、实例分割和姿态估计等多种计算机视觉任务。

相关链接

Paper: <https://arxiv.org/pdf/2502.12524>

Github: <https://github.com/sunsmarterjie/yolov12/tree/main>

● 论文三

Foundation model of neural activity predicts response to new stimulus types

日期

2025-04-09

Foundation model of neural activity predicts response to new stimulus types

<https://doi.org/10.1038/s41586-025-08829-y>

Received: 24 April 2023

Accepted: 21 February 2025

Published online: 9 April 2025

Open access

Eric Y. Wang^{1,2}, Paul G. Fahey^{1,2,3,4,5}, Zhuokun Ding^{1,2,3,4,5}, Stelios Papadopoulos^{1,2,3,4,5}, Kayla Ponder^{1,2}, Marissa A. Weis⁶, Andersen Chang^{1,2}, Taliah Muhammad^{1,2}, Saumil Patel^{1,2,3,4,5}, Zhiwei Ding^{1,2}, Dat Tran^{1,2}, Jiakun Fu^{1,2}, Casey M. Schneider-Mizell⁷, MICrONS Consortium*, R. Clay Reid⁷, Forrest Collman⁷, Nuno Maçarico da Costa⁷, Katrin Franke^{1,2,3,4,5}, Alexander S. Ecker^{6,8}, Jacob Reimer^{1,2}, Xaq Pitkow^{1,2,9}, Fabian H. Sinz^{1,2,6,10} & Andreas S. Tolias^{1,2,3,4,5,11}✉

内容摘要

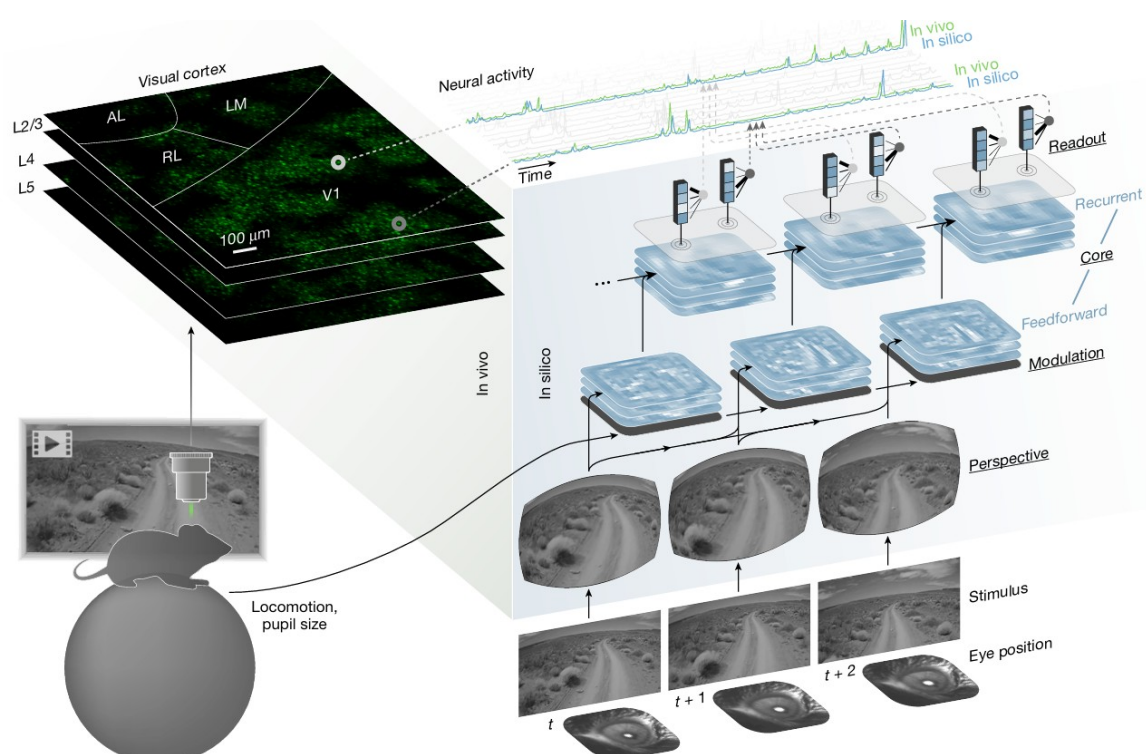


Fig. 1 | ANN model of the visual cortex. Top left, an in vivo recording session of excitatory neurons from several areas (V1, LM, RL and AL) and layers (layer 2/3 (L2/3), layer 4 (L4) and layer 5 (L5)) of the mouse visual cortex. Right, the architecture of the ANN model and the flow of information from inputs (visual stimulus, eye position, locomotion and pupil size) to outputs (neural activity). Underlined labels denote the four main modules of the ANN. For the modulation and core, the stacked planes represent feature maps. For the readout, the blue

boxes represent the output features of the core at the readout position of the neuron, and the fanning black lines represent readout feature weights. The top of the schematic displays the neural activity for a sampled set of neurons. In vivo and in silico responses are shown for two example neurons. Stimulus adapted from Sports-1M Dataset (Andrej Karpathy; <https://cs.stanford.edu/people/karpathy/deepvideo/>); copyright 2014, IEEE, reprinted with permission from IEEE Proceedings, IEEE (CC BY3.0).

本文提出了一种基于“基础模型”的小鼠视觉皮层神经活动模型，能够精准预测小鼠在面对不同刺激时的神经反应。该模型通过收集多个小鼠的神经活动数据，训练出了一个“基础核心”，使得模型能够迅速适应新小鼠的数据，仅需要较少的训练数据就能得到高精度预测。该模型不仅能够有效预测自然视频刺激下的神经活动，还能应用于全新的刺激领域，如噪声模式和静态图像。进一步，模型还成功预测了小鼠视觉皮层的解剖学特征，包括细胞类型和突触连接性，为理解神经活动与解剖结构之间的关系提供了新的视角。此研究为基础神经科学模型的建立奠定了基础，展示了数据驱动模型在系统神经科学中的巨大潜力。

主要亮点

- **跨领域的神经反应预测能力：**该模型不仅在自然视频刺激下表现出色，还能够成功预测其他类型刺激（如噪声模式和静态图像）下的神经活动，展示了强大的通用性。
- **基础模型核心的迁移学习优势：**通过训练“基础核心”并将其应用于新的小鼠，模型能在仅有少量训练数据的情况下实现高精度预测，相比于传统的个体化模型，显著提高了数据效率和泛化能力。
- **解剖学与功能性预测的结合：**该模型不仅能够预测神经活动，还能通过功能条形码预测解剖学特征，如细胞类型和突触连接性，为结构-功能关系的研究提供了新的工具和方法。

相关链接

Paper: <https://www.nature.com/articles/s41586-025-08829-y>

Github: <https://github.com/cajal/fnn>

● 论文四

LLMs are Bayesian, In Expectation, Not in Realization

日期

2025-07-15

LLMs are Bayesian, In Expectation, Not in Realization

Leon Chlon
Hassana Labs
leo@hassana.io

Sarah Rashidi
Hassana Labs
sarah@hassana.io

Zein Khamis
Hassana Labs
zein@hassana.io

MarcAntonio M. Awada
Harvard University
mawada@hbs.edu

内容摘要

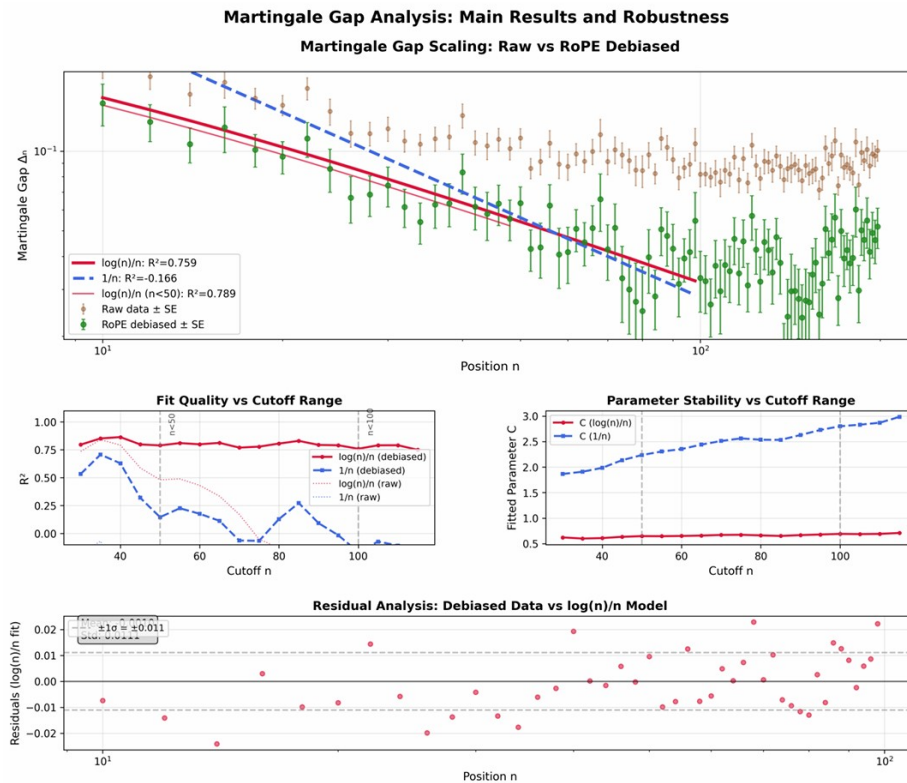


Figure 1: **Martingale violations follow theoretical predictions.** Raw and debiased gaps both exhibit $\Theta(\log n/n)$ scaling, with model comparison strongly favoring our theoretical prediction over classical $\Theta(1/n)$ scaling. Debiasing removes periodic artifacts while preserving fundamental scaling behavior.

该论文探讨了大型语言模型（LLMs）与贝叶斯推理之间的关系，并提出了一个重要的观点：LLMs 在“期望”（in Expectation）层面上表现出贝叶斯行为，但在“现实”（in Realization）层面上并非如此。研究发现，虽然 LLMs 能够执行类似贝叶斯定理所描述的任务（如更新信念），但它们所生成的概率分布并不完全符合贝叶斯理性的数学要求。这意味着 LLMs 在处理不确定性时，其内部机制并非严格遵循贝叶斯推理原则，这对于理解其能力和局限性具有重要意义。

主要亮点

- **澄清 LLMs 与贝叶斯推理的关系：**论文首次系统性地论证了 LLMs 的贝叶斯行为是存在于“期望”而非“现实”层面。这为理解 LLMs 如何处理不确定性和信息更新提供了精确的理论框架，避免了将 LLMs 简单等同于贝叶斯模型的误解。
- **构建贝叶斯基基准测试框架：**为了验证这一论点，研究者开发了一个新的贝叶斯推理基准测试框架，用于评估 LLMs 在复杂信念更新任务中的表现。该框架可以量化 LLMs 在多轮证据推理中的偏差，为未来的模型评估提供了新的工具。
- **揭示 LLMs 的内在局限性：**论文通过实验表明，尽管 LLMs 在表面上能够进行信念更新，但它们生成的概率分布往往存在“过度自信”或“非理性”的偏差。这一发现揭示了 LLMs 在严格逻辑推理和不确定性量化方面的内在局限性，对 LLM 在科学、金融等需要高度严谨性的领域中的应用具有重要的指导意义。

相关链接

Paper: <https://arxiv.org/abs/2507.11768>

Github: <https://github.com/leochlon/hallbayes>

● 论文五

Neural-Driven Image Editing

日期

2025-08-08

Neural-Driven Image Editing

Pengfei Zhou^{1*†}, Jie Xia^{2*†}, Xiaopeng Peng^{3†}, Wangbo Zhao¹, Zilong Ye², Zekai Li¹,
Suorong Yang^{1,4}, Jiadong Pan², Yuanxiang Chen⁵, Ziqiao Wang¹, Kai Wang¹, Qian Zheng²,
Xiaojun Chang^{5,6}, Gang Pan², Shurong Dong^{2†}, Kaipeng Zhang^{7,8†}, Yang You¹
¹NUS ²Zhejiang University ³RIT ⁴NJU ⁵USTC ⁶MBZUAI ⁷Shanghai AI Lab ⁸SII
<https://loongx1.github.io>

内容摘要

本文提出了 LoongX，一种基于神经驱动的图像编辑方法，旨在通过多模态神经生理信号实现免手动的图像编辑。LoongX 利用最新的扩散模型，训练了一个包含 23,928 对图像编辑对的数据集，每对数据都配有同步的脑电图（EEG）、功能性近红外光谱（fNIRS）、光电容积脉搏波描记图（PPG）和头部运动信号，以捕捉用户的意图。为了有效处理这些异质信号，LoongX 集成了两个关键模块：跨尺度状态空间（CS3）模块和动态门控融合（DGF）模块。这两个模块可以将不同模态的特征聚合到统一的潜在空间中，并通过微调扩散变换器（DiT）进行语义对齐。LoongX 通过这种创新的方法，不仅实现了图像编辑的自动化，还能够根据用户的神经信号和语音输入生成个性化的编辑效果。实验结果表明，LoongX 的图像编辑效果在性能上与基于文本的图像编辑方法相当，并且在结合神经信号和语音时表现出更优的性能，显示了其在个性化和实时性方面的优势。

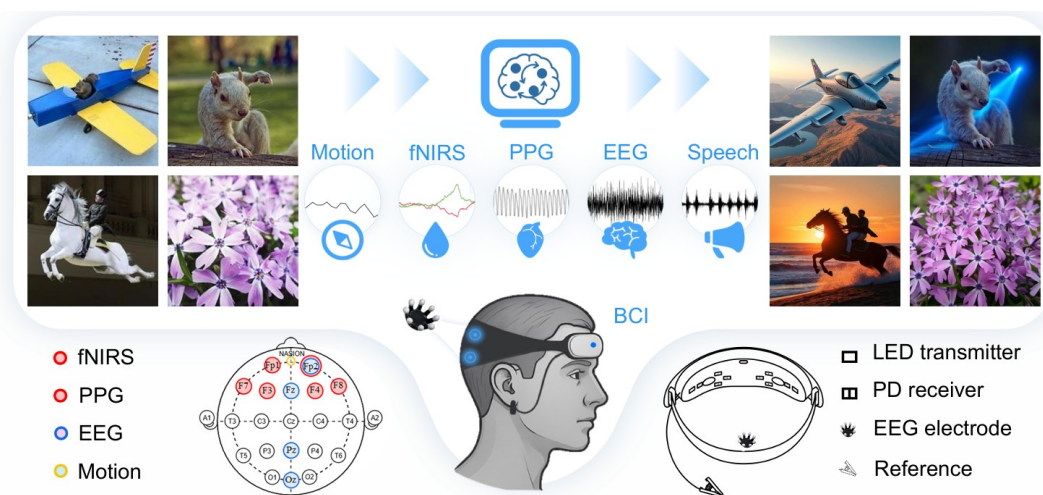


Figure 1: Illustration of LoongX for hands-free image editing via multimodal neural signals.

主要亮点

- **多模态信号融合**：LoongX 通过融合 EEG、fNIRS、PPG 和头部运动信号，实现了对用户意图的准确捕捉，为个性化图像编辑提供了新的方式。
- **跨尺度状态空间模块**：提出了跨尺度状态空间模块，有效地编码了来自不同模态的特征，并增强了图像编辑的准确性。
- **动态门控融合模块**：设计了动态门控融合模块，通过智能融合不同模态的特征，提升了图像编辑效果的实时性和个性化。

相关链接

Paper: <https://arxiv.org/abs/2507.05397>

Github: <https://loongx1.github.io/>

● 论文六

Disentangling the Factors of Convergence between Brains and Computer Vision Models

日期

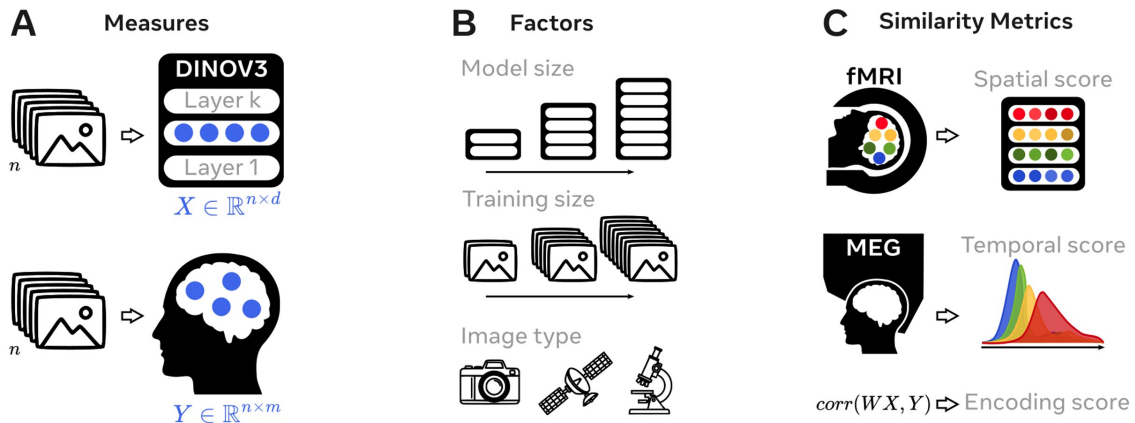
2025-08-25

Disentangling the Factors of Convergence between Brains and Computer Vision Models

Joséphine Raugel^{1,2}, Marc Szafraniec¹, Huy V. Vo¹, Camille Couprie¹, Patrick Labatut¹, Piotr Bojanowski¹, Valentin Wyart², Jean-Rémi King¹

¹Meta AI, ²Ecole Normale Supérieure - PSL Université

内容摘要



A. We compare the activation of DINOv3, a state-of-the-art self-supervised computer vision model trained on natural images, to the activations of the human brain in response to the same images. **B.** To understand the factors that make DINOv3 more-or-less similar to the brain, we train *from scratch* a variety of models on different image domains (pictures from human-centric cameras, satellite images or biological data), and with a varying amount of data. **C.** We compare each model to both functional Magnetic Resonance Imaging (fMRI, with high spatial resolution) and Magneto-Encephalography (MEG, with high temporal resolution) by computing the overall linear similarity of their representations (encoding score) and the similarity of their hierarchical organization (spatial and temporal scores).

本文深入探讨了人工智能模型在处理自然图像时，如何在表征层面与人类大脑产生相似性。研究者训练了一系列自监督视觉变换器（DINOv3），系统地调整模型架构、训练数据量和图像类型等因素，以分析这些因素如何独立或交互地影响模型的脑类比性。通过与人类大脑的功能性磁共振成像（fMRI）和脑磁图（MEG）数据进行比较，研究发现：模型规模、训练数据量和图像类型均独立且交互地影响模型的脑类比性；训练过程

中，模型首先与人类大脑的感觉皮层早期表征对齐，随后才对齐到大脑的晚期和前额叶表征；模型对晚期和前额叶表征的对齐与人类皮层的结构和功能特性相关，如皮层的发育扩展、厚度、髓鞘化程度和时间尺度等。这些发现揭示了人工神经网络在视觉表征上如何模仿人类大脑的过程，为理解人类大脑如何表征视觉世界提供了新的视角。

主要亮点

- **系统性实验设计：**通过系统地调整模型架构、训练数据量和图像类型等因素，深入分析了这些因素对模型脑类比性的影响。
- **多模态数据比较：**利用 fMRI 和 MEG 数据，与人工神经网络的表征进行比较，提供了高分辨率的空间和时间分析。
- **结构功能关联：**发现模型对晚期和前额叶表征的对齐与人类皮层的结构和功能特性相关，揭示了人工神经网络在视觉表征上如何模仿人类大脑的过程。

相关链接

Paper: <https://arxiv.org/abs/2507.05397>

● 论文七

R-4B: Incentivizing General-Purpose Auto-Thinking Capability in MLLMs via Bi-Mode Annealing and Reinforce Learning

日期

2025-08-28

-4B: INCENTIVIZING GENERAL-PURPOSE AUTO-THINKING CAPABILITY IN MLLMs VIA BI-MODE ANNEALING AND REINFORCE LEARNING

¹Tencent Hunyuan Team ²Institute of Automation, CAS

 <https://github.com/yannqi/R-4B>
 <https://huggingface.co/YannQi/R-4B>

内容摘要

本文提出了 R-4B 框架，旨在提高多模态大语言模型（MLLMs）的通用自动思维能力。当前的多模态大语言模型在处理复杂任务时，往往依赖于传统的推理路径，这些路径对于简单任务来说可能过于复杂，导致计算资源的浪费。为了解决这一问题，R-4B 引入了双模退火（Bi-Mode Annealing）和强化学习（Reinforcement Learning）机制，使模型能够根据任务的复杂度自适应地选择是否进行推理，进而优化计算效率。双模退火机制能够在推理过程中根据当前任务状态灵活调整推理策略，而强化学习则用来优化推理过程中的决策机制。作者通过大量实验验证了该方法在多任务推理中的表现，R-4B 在提高任务适应性和降低计算成本方面展现了显著的优势。实验结果表明，R-4B 能够有效地平衡任务复杂性与计算效率，在多个任务上获得了优异的性能表现，尤其在推理时间和资源消耗方面比传统方法更为高效。

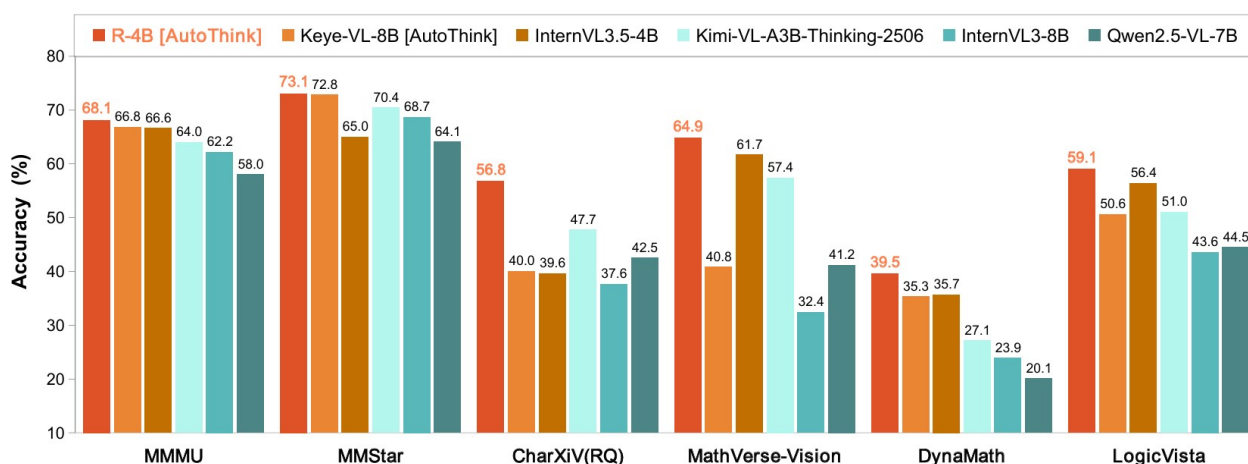


Figure 1: Comparison between R-4B-RL and frontier open-source MLLMs, including non-thinking MLLMs (e.g., InternVL3-8B, Qwen2.5-VL 7B), thinking MLLMs (e.g., Kimi-VL-A3B-Thinking-2506) and auto-thinking MLLMs (e.g., Keye-VL-8B), on different benchmarks.

主要亮点

- **双模退火机制**：通过双模退火机制，R-4B 能够动态调整推理策略，根据任务复杂度自适应选择推理路径，从而提高效率并降低计算资源消耗。
- **强化学习优化**：结合强化学习优化推理过程中的决策机制，确保模型能够在不同复杂度的任务中高效执行，并有效减少不必要的推理过程。
- **实验验证**：在多个标准任务和基准上进行验证，R-4B 在提高任务适应性和计算效率方面表现出色，尤其在推理时间和资源消耗方面比现有方法更具优势。

相关链接

Paper: <https://arxiv.org/abs/2508.21113>

GitHub: <https://github.com/yannqi/R-4B>

Model: <https://huggingface.co/YannQi/R-4B>

● 论文八

Implicit Reasoning in Large Language Models: A Comprehensive Survey

日期

2025-09-02

Implicit Reasoning in Large Language Models: A Comprehensive Survey

Jindong Li*	jli839@connect.hkust-gz.edu.cn
Hong Kong University of Science and Technology (Guangzhou)	
Yali Fu*	fuyi23@mails.jlu.edu.cn
Jilin University	
Li Fan	a213837054@gmail.com
Hong Kong University of Science and Technology (Guangzhou)	
Jiahong Liu	jiahong.liu21@gmail.com
The Chinese University of Hong Kong	
Yao Shu	yaoshu@hkust-gz.edu.cn
Hong Kong University of Science and Technology (Guangzhou)	
Chengwei Qin	chengweiqin@hkust-gz.edu.cn
Hong Kong University of Science and Technology (Guangzhou)	
Menglin Yang†	menglin.yang@outlook.com
Hong Kong University of Science and Technology (Guangzhou)	
Irwin King	king@cse.cuhk.edu.hk
The Chinese University of Hong Kong	
Rex Ying	rex.ying@yale.edu
Yale University	

内容摘要

本文全面回顾了大型语言模型（LLMs）中的隐式推理，介绍了三种主要的执行范式：潜在优化、信号引导控制和层间递归执行。作者通过结构性、行为性和表示性证据，支持 LLMs 在推理过程中的隐式推理能力，并对现有的推理模型进行了分类和评估。隐式推理指的是模型在进行推理时，能通过内部机制有效地处理和理解信息，而不需要外部明确的推理步骤或规则。该文详细分析了这些推理机制如何通过内部模型结构进行信息处理，并为 LLMs 提供了一种新的视角来理解其推理过程。文章还讨论了当前方法的优缺点，并提出了未来改进隐式推理模型的研究方向。

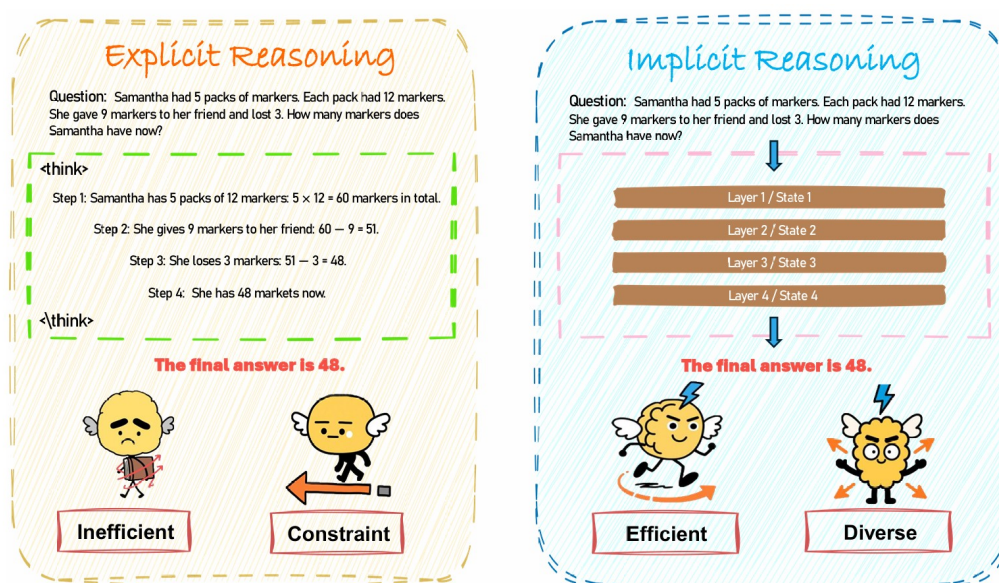


Figure 1: Comparison between explicit and implicit reasoning in LLMs. Explicit reasoning shows each step by producing natural language explanations, as illustrated on the left. The model describes the problem-solving process one step at a time. In contrast, implicit reasoning, shown on the right, handles the process internally across different layers or states without writing out any steps. Explicit reasoning is less efficient because generating text takes time and resources. *Implicit reasoning happens inside the model by hidden representations, supporting faster processing.* Also, explicit reasoning is limited by the structure of language, while *implicit reasoning allows many types of internal computation without needing to be described in words.*

主要亮点

- **三种推理范式：**介绍了潜在优化、信号引导控制和层间递归执行三种执行范式，并分析了它们在隐式推理中的作用。
- **隐式推理的支持证据：**通过结构性、行为性和表示性证据，支持 LLMs 在推理过程中存在隐式推理机制的观点。
- **未来研究方向：**提出了在隐式推理模型的改进方向，并为进一步提升 LLMs 推理能力提供了新的视角。

相关链接

Paper: <https://arxiv.org/abs/2509.02350>

GitHub: <https://github.com/digailab/awesome-llm-implicit-reasoning>

● 论文九

The Landscape of Agentic Reinforcement Learning for LLMs: A Survey

日期

2025-09-02



内容摘要

本文对智能体强化学习（Agentic RL）在大语言模型（LLMs）中的应用进行了全面调查。传统的 LLM 强化学习方法主要聚焦于单一的决策步骤，通常用于解决简单的任务。然而，随着任务复杂性的增加，智能体强化学习的优势愈加显现。智能体强化学习不仅关注单步决策，还强调在复杂、动态的环境中进行长期规划和决策，以实现长期目标。作者基于代理 RL 的核心能力，提出了一个两维度的分类框架：一方面，基于核心能力如规划、工具使用、记忆、推理、自我改进和感知；另一方面，基于应用领域，如对话系统、游戏、机器人控制等。通过这个框架，本文总结了智能体强化学习在多个实际应用中的潜力，并讨论了当前研究中的挑战和未解问题。通过对现有开源环境、基准和框架的汇总，本文为未来的智能体 RL 研究提供了系统的指导和参考。

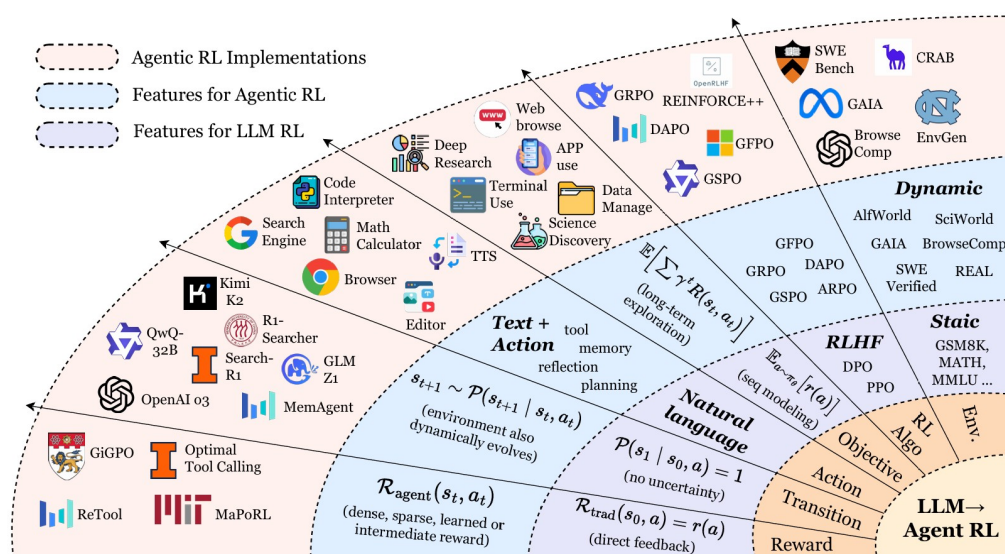


Figure 2: Paradigm shift from LLM-RL to agentic RL.

主要亮点

- **智能体强化学习的定义**：通过对智能体强化学习的系统阐述，明确了 LLMs 从被动的生成模型转变为主动的决策代理，为复杂任务提供了新的解决方案。
- **分类框架**：提出了基于核心能力和应用领域的两维度分类框架，为研究者提供了深入理解智能体 RL 的系统视角，涵盖了规划、记忆、推理等多个关键能力。
- **资源汇总**：汇总了当前在智能体强化学习领域中的开源环境、基准和框架，为学者们提供了丰富的研究资源，促进了该领域的进一步发展。

相关链接

Paper: <https://arxiv.org/abs/2509.02547>

GitHub: <https://github.com/xhyumiracle/Awesome-AgenticLLM-RL-Papers>

● 论文十

REFRAG: Rethinking RAG based Decoding

日期

2025-09-03

REFRAG: Rethinking RAG based Decoding

Xiaoqiang Lin^{1,2,*}, Aritra Ghosh¹, Bryan Kian Hsiang Low², Anshumali Shrivastava^{1,3}, Vijai Mohan¹

¹Meta Superintelligence Labs, ²National University of Singapore, ³Rice University

*Work done at Meta

内容摘要

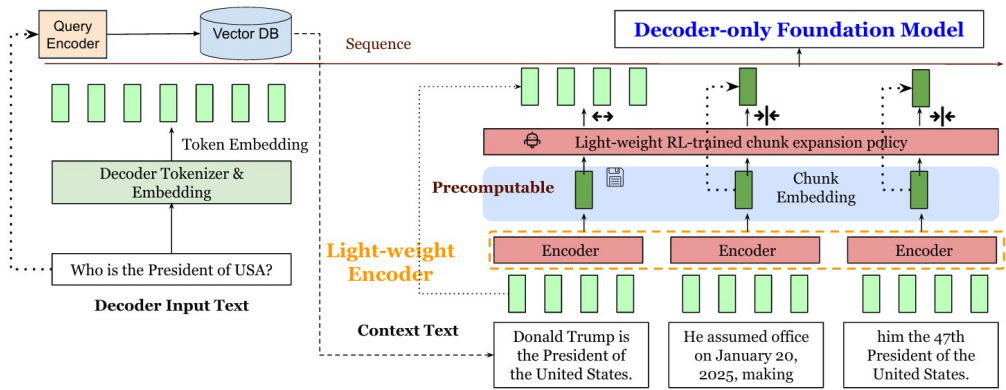


Figure 1 The main design of REFRAG. The input context is chunked and processed by the light-weight encoder to produce chunk embeddings, which are precomputable for efficient reuse. A light-weight RL policy decide few chunks to expand. These chunk embeddings along with the token embeddings of the question input are fed to the decoder.

本论文针对检索增强生成在解码阶段的效率瓶颈提出了 REFRAG (Representation For RAG) 框架。作者观察到，RAG 模型在生成输出时会对检索到的大量上下文进行重复计算，而实际上大部分这些计算是冗余的。为此，REFRAG 采用“压缩—感知—扩展” (compress → sense → expand) 三步机制：首先对检索上下文进行压缩，保留关键信息；然后在压缩表示上进行快速感知与筛选；最后在必要时扩展回原始上下文进行精细计算。这样可以显著减少解码时对上下文的计算负担，而无需改变基础的 LLM 架构或引入额外的解码器参数。实验证明，REFRAG 在常见的 RAG 任务上可以将 time-to-first-token 加速约 30.85 倍 (比之前

最优方法提升 3.75 倍)，同时在困惑度 (perplexity) 等性能指标上几乎无损失。作者还展示了其在内存使用和延迟方面的优越性，从而为高效 RAG 系统在实际部署中的性能优化提供了可行路径。

主要亮点

- **压缩—感知—扩展三步机制**：REFRAG 提出一个新的解码流程：先压缩检索上下文至紧凑表示，再在紧凑表示上快速筛选与感知，最后仅在必要时扩展到全上下文计算。这一设计有效去除冗余计算，同时保留关键语义信息。
- **无需修改模型结构**：REFRAG 不依赖于对基础 LLM 模型结构的改动，也不引入新的解码器参数，而是以一种框架式方式插入解码流程，具有良好的兼容性和易部署性。
- **显著提升解码效率且性能无损**：在各类 RAG 任务中，REFRAG 实现了 time- to- first- token 加速约 30.85×，比已有最优方法提升约 3.75 倍，同时在 perplexity 等指标上几乎无性能下降。

相关链接

Paper: <https://arxiv.org/abs/2509.01092>

GitHub: <https://github.com/vaisax/refrag>

● 论文十一

Attention as an Adaptive Filter

日期

2025-09-04

Attention as an Adaptive Filter

Peter Racioppo

Independent Researcher, formerly UCLA
pcracioppo@ucla.edu

September 5, 2025

内容摘要

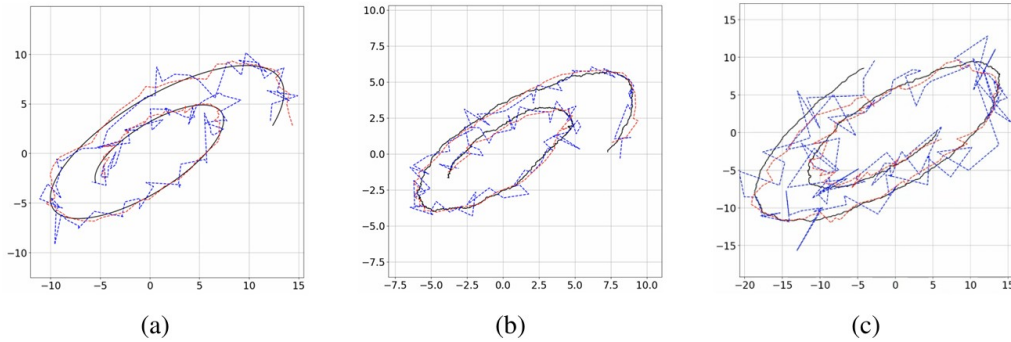


Figure 1: Filter performance on different 2D systems: ground-truth trajectory (black), measured (blue), and predicted (red). (a) system with only measurement noise ($\sigma^2 = 0.0$, $\eta^2 = 1.0$). (b) system with both process noise and measurement noise ($\sigma^2 = 0.3$, $\eta^2 = 0.5$). (c) higher noise ($\sigma^2 = 0.5$, $\eta^2 = 2.0$).

本文提出了一种新的注意力机制——自适应滤波注意力（Adaptive Filter Attention, AFA），旨在通过对输入序列建模不确定性，提升注意力机制的鲁棒性和准确性。与传统的基于固定权重的注意力机制不同，AFA将输入序列视为线性随机微分方程（SDE）的离散观测，从而引入可学习的动态模型来调整注意力权重。该方法通过对信息的不确定性进行建模，允许模型在推理过程中自然地传播配对不确定性，使得注意力权重能够自

适应地调整。这种新的注意力机制不仅提升了模型在处理复杂任务时的鲁棒性，还增强了对输入数据不确定性的适应能力。作者通过一系列实验验证了 AFA 在多个基准任务中的表现，结果显示，AFA 在提高模型的可靠性和泛化能力方面具有显著优势，尤其在处理噪声和不确定性较高的数据时，表现优于传统方法。

主要亮点

- **自适应滤波机制：**通过引入自适应滤波机制，AFA 能够自适应地调整注意力权重，从而提升模型在复杂任务中的鲁棒性。
- **不确定性建模：**AFA 通过建模输入序列中的不确定性，提高了模型对复杂任务和不确定数据的适应能力。
- **实验验证：**在多个标准任务中验证了 AFA 的有效性，结果显示其在处理噪声和不确定性较高的数据时，具有明显的优势。

相关链接

Paper: <https://arxiv.org/abs/2509.04154>

● 论文十二

Why Language Models Hallucinate

日期

2025-09-04

Why Language Models Hallucinate

Adam Tauman Kalai*
OpenAI

Ofir Nachum
OpenAI

Santosh S. Vempala†
Georgia Tech

Edwin Zhang
OpenAI

September 4, 2025

内容摘要

本文深入探讨了大型语言模型（LLMs）在生成答案时为何会出现“幻觉”现象，即在缺乏可靠证据或信息时产生虚假或不准确的回答。研究发现，幻觉的根源主要在于模型的训练机制和现有的评估标准。在传统的训练和评估过程中，模型通常会被奖励生成“有意义”的回答，即便这些回答在事实层面并不准确。这种机制使得模型即使对某些问题没有充分的信息，也会倾向于“猜测”并生成看似合理的内容。作者指出，模型在没有外部信息或不确定的情况下，依赖于上下文中的统计关系和模式进行推测，这种过度自信的行为导致了幻觉现象。为了应对这一挑战，作者提出了一种新的评估标准，建议减少对不确定回答的惩罚，并鼓励模型在不确定的情况下表达保留意见。这种方法有助于构建更具可靠性和可解释性的 AI 系统，减少误导性输出的发生。

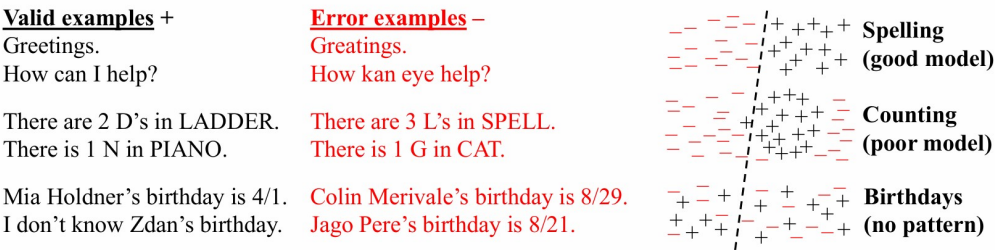


Figure 1: Is-It-Valid requires learning to identify valid generations using labeled $\pm$ examples (left). Classifiers (dashed lines) may be accurate on certain concepts like spelling (top) but errors often arise due to poor models (middle) or arbitrary facts when there is no pattern in the data (bottom).

主要亮点

- **幻觉的统计学根源：**通过对幻觉现象的分析，揭示了这种现象源于模型的过度自信和缺乏有效的不确定性表达。
- **评估标准的影响：**现有评估机制过于注重生成有意义的回答，忽视了输出的准确性和真实性，导致了幻觉现象的普遍存在。
- **提出的解决方案：**提议修改评估基准，减少对不确定性回答的惩罚，以促进更为可靠的模型生成，并避免误导性信息的传播。

相关链接

Paper:

<https://cdn.openai.com/pdf/d04913be-3f6f-4d2b-b283-ff432ef4aaa5/why-language-models-hallucinate.pdf>

● 论文十三

A Comprehensive Survey on Trustworthiness in Reasoning with Large Language Models

日期

2025-09-04

A Comprehensive Survey on Trustworthiness in Reasoning with Large Language Models

Yanbo Wang^{1,2}, Yongcan Yu^{1,2}, Jian Liang^{1,2,‡} and Ran He^{1,2}
¹School of Artificial Intelligence, University of Chinese Academy of Sciences
²NLPR & MAIS, Institute of Automation, Chinese Academy of Sciences

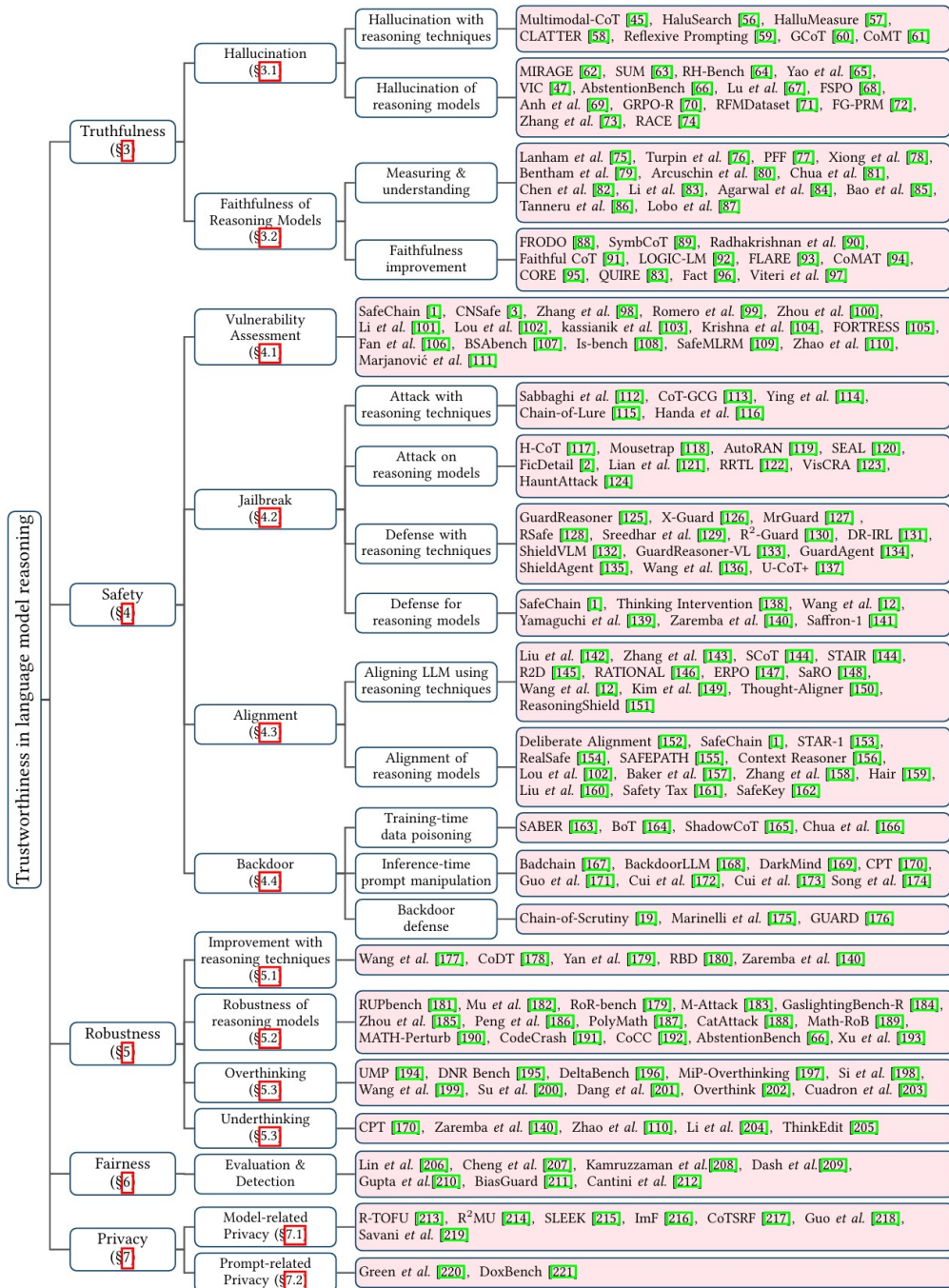


Figure 2: Taxonomy of trustworthiness in reasoning with large language models.

本文对大语言模型 (LLMs) 在推理过程中的可信度进行了全面的调研, 分析了影响推理可信度的五个核心维度: 真实性、安全性、鲁棒性、公平性和隐私性。随着 LLMs 在多个任务中的表现越来越好, 其在推理任务中的可靠性问题变得愈加重要。研究表明, 虽然推理技术提高了模型的性能, 但仍存在多方面的挑战, 特别是

在生成推理过程中，模型容易受到训练数据中的偏差和错误信息的影响，导致生成的推理结果可能存在不真实、缺乏安全性或鲁棒性差等问题。文章还讨论了如何解决这些问题，并提出了针对每一维度的未来研究方向。总的来说，本文为 AI 领域的学者提供了一个全面的框架，以指导未来在 LLMs 推理可信度方面的研究和实践。

主要亮点

- **推理可信度的五大维度：**本文深入探讨了推理过程中的真实性、安全性、鲁棒性、公平性和隐私性等核心问题，并提供了对这些问题的全面分析。
- **推理模型的挑战：**尽管推理技术提升了 LLMs 的性能，但仍面临诸如数据偏差和推理结果不准确的问题，影响了推理模型的可信度。
- **未来研究方向：**提出了针对推理可信度的未来研究方向，重点关注如何改善推理模型的安全性、鲁棒性及公平性，为 LLMs 的发展提供了重要的指导。

相关链接

Paper: <https://arxiv.org/abs/2509.03871>

GitHub: <https://github.com/ybwang119/Awesome-reasoning-safety>

● 论文十四

Towards a Unified View of Large Language Model Post-Training

日期

2025-09-04

Towards a Unified View of Large Language Model Post-Training

Xingtai Lv^{1*}, Yuxin Zuo^{1*}, Youbang Sun^{1†}, Hongyi Liu¹, Yuntian Wei¹,
Zhekai Chen¹, Lixuan He¹, Xuekai Zhu¹, Kaiyan Zhang¹, Bingning Wang³,
Ning Ding^{1,2†}, Bowen Zhou^{1,2†}

¹Tsinghua University, ²Shanghai AI Laboratory, ³WeChat AI

🔗 Code: [TsinghuaC3I/Unify-Post-Training](https://github.com/TsinghuaC3I/Unify-Post-Training)

✉ Mail: lvxt24@mails.tsinghua.edu.cn

内容摘要

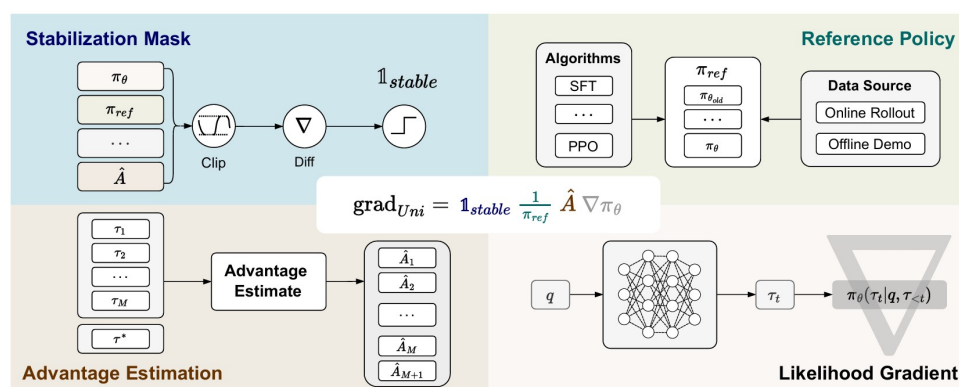


Figure 1: Illustration of the Unified Policy Gradient Estimator. The “ ∇ ” in the background of the Likelihood Gradient part refers to the calculation of the gradient with respect to the π_θ .

本文提出了一个统一的后训练框架，旨在整合监督微调（SFT）和强化学习（RL）这两种后训练方法。作者通过引入一个统一的策略梯度估计器，将这两种方法看作同一个优化过程的不同实例，从而使得模型可以在后训练阶段灵活地选择不同的训练信号，以提高推理和任务适应性。特别是，本文提出了一种名为“混合后训练”（Hybrid Post-Training, HPT）的算法，它可以在训练过程中根据任务的不同需求，动态地选择训练信号并进行有效的任务适应。通过多种实验验证，HPT 算法在数学推理任务和分布外测试中的表现均超过了现有

的强基线，证明了该方法的有效性。此外，作者还介绍了该算法的训练过程，并分析了不同训练策略对推理效率和准确性的影响。

主要亮点

- **统一框架**：提出了一个统一的后训练框架，通过一个策略梯度估计器将 SFT 和 RL 整合为一个优化过程，为模型训练提供了更灵活的选择。
- **混合后训练 (HPT) 算法**：介绍了 HPT 算法，它能够根据不同任务的需求，动态选择训练信号，从而提升推理任务的适应性。
- **实验验证**：在多个数学推理基准和分布外测试中，HPT 表现超越了现有的强基线模型，显示出该方法在提高模型性能和推理效率上的优势。

相关链接

Paper: <https://arxiv.org/abs/2509.04419>

GitHub: <https://github.com/TsinghuaC3I/Unify-Post-Training>

● 论文十五

An AI system to help scientists write expert-level empirical software

日期

2025-09-08

An AI system to help scientists write expert-level empirical software

Eser Aygün^{1,*}, Anastasiya Belyaeva^{2,*}, Gheorghe Comanici^{1,*}, Marc Coram^{2,*}, Hao Cui^{2,*}, Jake Garrison^{3,*}, Renee Johnston^{2,*}, Anton Kast^{2,*}, Cory Y. McLean^{2,*}, Peter Norgaard^{2,*}, Zahra Shamsi^{2,*}, David Smalling^{1,*}, James Thompson^{2,*}, Subhashini Venugopalan^{2,*}, Brian P. Williams^{2,*}, Chujun He^{2,4,**}, Sarah Martinson^{2,5,**}, Martyna Plomecka^{2,6,**}, Lai Wei², Yuchen Zhou², Qian-Ze Zhu^{2,5,**}, Matthew Abraham², Erica Brand², Anna Bulanova¹, Jeffrey A. Cardille^{2,7}, Chris Co², Scott Ellsworth², Grace Joseph², Malcolm Kane², Ryan Krueger^{2,5,**}, Johan Kertiwa², Dan Liebling², Jan-Matthis Lueckmann², Paul Raccuglia², Xuefei (Julie) Wang^{2,8,**}, Katherine Chou², James Manyika², Yossi Matias², John C. Platt², Lizzie Dorfman², Shibl Mourad^{1,‡} and Michael P. Brenner^{2,5,‡}

¹Google DeepMind, ²Google Research, ³Google Platforms and Devices, ⁴Massachusetts Institute of Technology, ⁵School of Engineering and Applied Sciences, Harvard University, ⁶Google Cloud, ⁷Faculty of Agricultural and Environmental Sciences, McGill University, ⁸California Institute of Technology

内容摘要

本文介绍了一个 AI 系统，旨在帮助科学家编写专家级的软件，以加速科学发现过程。该系统结合了大语言模型（LLM）和树搜索（Tree Search, TS）方法，系统地探索和整合复杂的研究思想，生成高质量的科学软件。在生物信息学领域，该系统发现了 40 种新的单细胞数据分析方法，在公共排行榜上超越了顶级人类开发的方法；在流行病学领域，生成了 14 个模型，超越了 CDC 的集成模型和所有其他单一模型。此外，该方法还在地理空间分析、斑马鱼神经活动预测、时间序列预测和数值积分求解等任务中表现出色。这些成果展示了 AI 系统在加速科学进展方面的潜力。

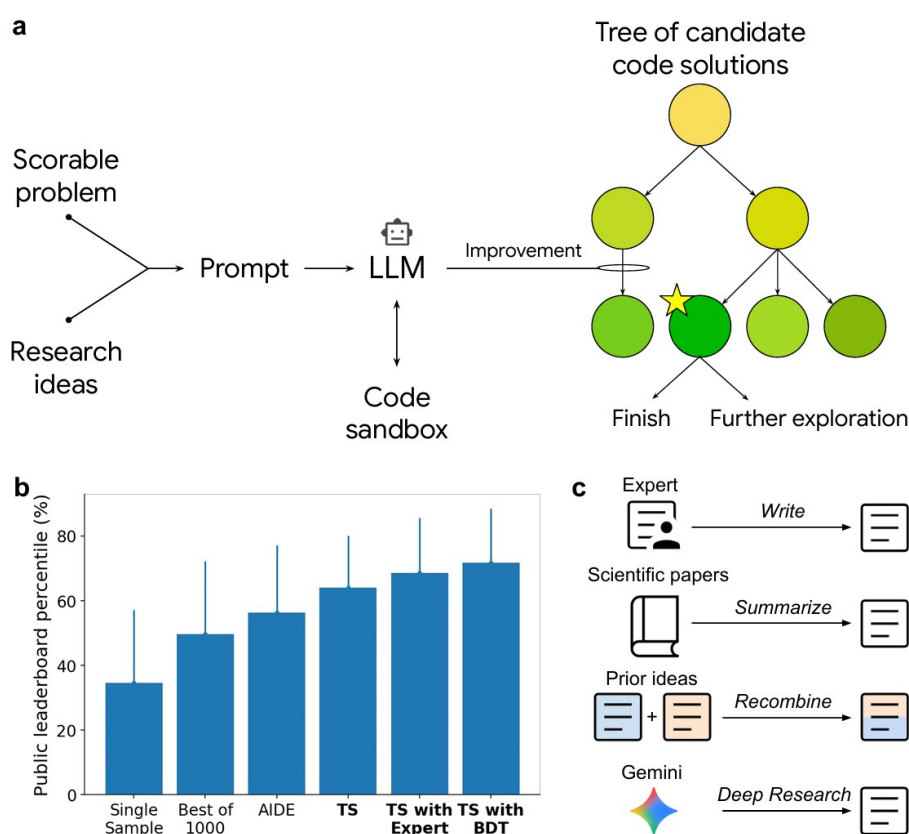


Figure 1 | Schematic and performance of our method. **a**, Schematic of our method algorithm. A scorable task, together with research ideas proposing methods to solve the task, are fed to an LLM, which produces code to evaluate the scorable task in a sandbox. This is then embedded within a tree search algorithm, whereby new nodes are chosen balancing exploitation and exploration, sampling from the LLM (Methods). **b**, Performance of code generation methods on Kaggle Playground benchmark. Results report the average public leaderboard percentile performance over 16 tasks. Methods based on our method are listed in bold. Error bars indicate standard deviation. BDT, boosted decision tree. **c**, Mechanisms used to create initial research ideas to solve scientific problems.

主要亮点

- **AI 辅助科学软件生成**: 结合 LLM 和 TS 方法，生成高质量的科学软件，帮助科学家加速研究进程。
- **跨领域应用**: 在生物信息学、流行病学、地理空间分析等多个领域取得了显著成果，展示了 AI 系统的广泛应用潜力。
- **性能超越人类专家**: 在多个任务中，AI 系统生成的软件超越了人类专家开发的方法，体现了 AI 在科学研究中的优势。

相关链接

Paper: <https://arxiv.org/abs/2509.06503>

● 论文十六

K2-Think: A Parameter-Efficient Reasoning System

日期

2025-09-09

K2-Think: A Parameter-Efficient Reasoning System

Zhoujun Cheng^{*}, Richard Fan^{*}, Shibo Hao^{*}, Taylor W. Killian^{*O}, Haonan Li^{*}, Suqi Sun^{*}
Hector Ren, Alexander Moreno, Daqian Zhang, Tianjun Zhong, Yuxin Xiong, Yuanzhe Hu, Yutao Xie
Xudong Han, Yuqi Wang, Varad Pimpalkhute, Yonghao Zhuang, Aaryamonvikram Singh, Xuezhi Liang
Anze Xie, Jianshu She, Desai Fan, Chengqian Gao, Liqun Ma, Mikhail Yurochkin, John Maggs
Xuezhe Ma, Guowei He, Zhiting Hu, Zhengzhong Liu^{*O}, Eric P. Xing^O

Institute of Foundation Models, Mohamed bin Zayed University of Artificial Intelligence

^{*}Core Contributors (listed alphabetically), ^OCorresponding Authors

内容摘要

本文提出了 K2-Think，一个参数高效的推理系统，旨在通过结合先进的后训练和测试时计算技术，实现与更大模型（如 GPT-OSS 120B 和 DeepSeek v3.1）相媲美的性能。该系统基于 Qwen2.5 基础模型，采用了六个关键技术支柱：长链推理监督微调、可验证奖励的强化学习（RLVR）、推理前的代理规划、测试时扩展、推测解码和优化硬件推理。K2-Think 在数学推理方面取得了开源模型的最新成绩，同时在代码和科学等其他领域也表现出色。实验结果表明，K2-Think 32B 模型通过集成的后训练方案，在推理速度和性能上均表现优异。

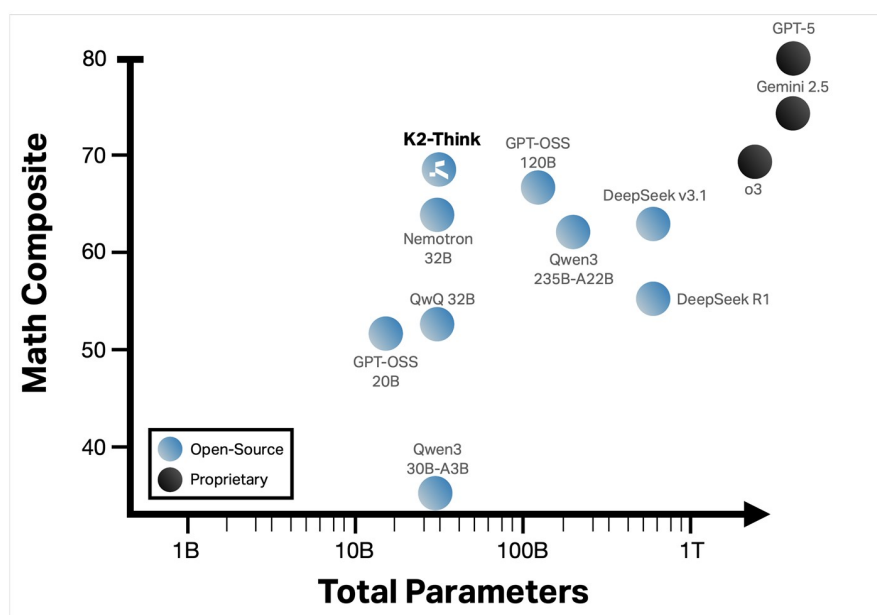


Figure 1: K2-THINK exhibits remarkable parameter efficiency, providing comparable or superior performance to frontier reasoning models in complex math domains with an order of magnitude smaller model. The composite score here is the micro-average for each model over four complex math benchmarks, weighted by the number of questions in each benchmark (AIME 2024, AIME 2025, HMMT 2025, and Omni-MATH-HARD; see Section 3 for the benchmark details). Note: parameter counts for proprietary models are speculative.

主要亮点

- **参数高效性:** K2-Think 32B 模型通过集成的后训练方案, 实现了与更大模型相媲美的性能, 展示了参数高效性的优势。
- **多技术集成:** 结合了长链推理监督微调、RLVR、推理前的代理规划等多种技术, 提升了推理系统的整体性能。
- **开源可用性:** K2-Think 系统是开源的, 提供了高效的推理速度, 促进了推理系统的普及和应用。

相关链接

Paper: <https://arxiv.org/abs/2509.07604>

GitHub: <https://github.com/MBZUAI-IFM/K2-Think-SFT>

Demo: <https://www.k2think.ai/>

Model: <https://huggingface.co/LLM360/K2-Think>

● 论文十七

A Survey of Reinforcement Learning for Large Reasoning Models

日期

2025-09-11

A Survey of Reinforcement Learning for Large Reasoning Models

Kaiyan Zhang^{1*†}, Yuxin Zuo^{1*†}, Bingxiang He^{1*}, Youbang Sun^{1*}, Runze Liu^{1*}, Che Jiang^{1*}, Yuchen Fan^{2,3*}, Kai Tian^{1*}, Guoli Jia^{1*}, Pengfei Li^{2,6*}, Yu Fu^{9*}, Xingtai Lv^{1*}, Yuchen Zhang^{2,4*}, Sihang Zeng^{7*}, Shang Qu^{1,2*}, Haozhan Li^{1*}, Shijie Wang^{2*}, Yuru Wang^{1*}, Xinwei Long¹, Fangfu Liu¹, Xiang Xu⁵, Jiaze Ma¹, Xuekai Zhu³, Ermo Hua^{1,2}, Yihao Liu^{1,2}, Zonglin Li², Huayu Chen¹, Xiaoye Qu², Yafu Li², Weize Chen¹, Zhenzhao Yuan¹, Junqi Gao⁶, Dong Li⁶, Zhiyuan Ma⁸, Ganqu Cui², Zhiyuan Liu¹, Biqing Qi^{2‡}, Ning Ding^{1,2‡}, Bowen Zhou^{1,2‡}

¹ Tsinghua University ² Shanghai AI Laboratory ³ Shanghai Jiao Tong University ⁴ Peking University
⁵ University of Science and Technology of China ⁶ Harbin Institute of Technology ⁷ University of Washington
⁸ Huazhong University of Science and Technology ⁹ University College London

[†] Project Lead. ^{*} Core Contributors. [‡] Corresponding Authors.

✉ zhang-ky22@mails.tsinghua.edu.cn 🔗 TsinghuaC3I/Awesome-RL-for-LRMs

内容摘要

本文系统综述了将强化学习（Reinforcement Learning, RL）用于提升大语言模型（LLMs）推理能力的研究方向，称为“RL for Large Reasoning Models (RL-for-LRMs)” 。作者指出，随着 LLMs 在复杂任务（如数学、编程、对话）中的广泛应用，如何通过 RL 机制提升其推理深度、自主性与交互智能，已成为前沿问题。文章从奖励设计、策略优化、采样机制三个基础组件切入，梳理了当前 RL 方法在推理任务中的关键挑战，包括稀疏奖励、探索-利用困境、奖励欺骗与样本效率等；同时总结了 RL-for-LRMs 所依赖的训练资源，包括静态语料（如 MATH、GSM8K）与动态交互环境（如 MiniWoB、WebArena），以及支撑这些方法的大规模系统与算力基础。在此基础上，作者回顾了 RL 在代码生成、检索增强生成、对话智能体、科学发现等具体场景的应用实践，并指出未来在构建更高层智能（如可验证奖励、协同学习、环境构造等）方面的核心研究方向。

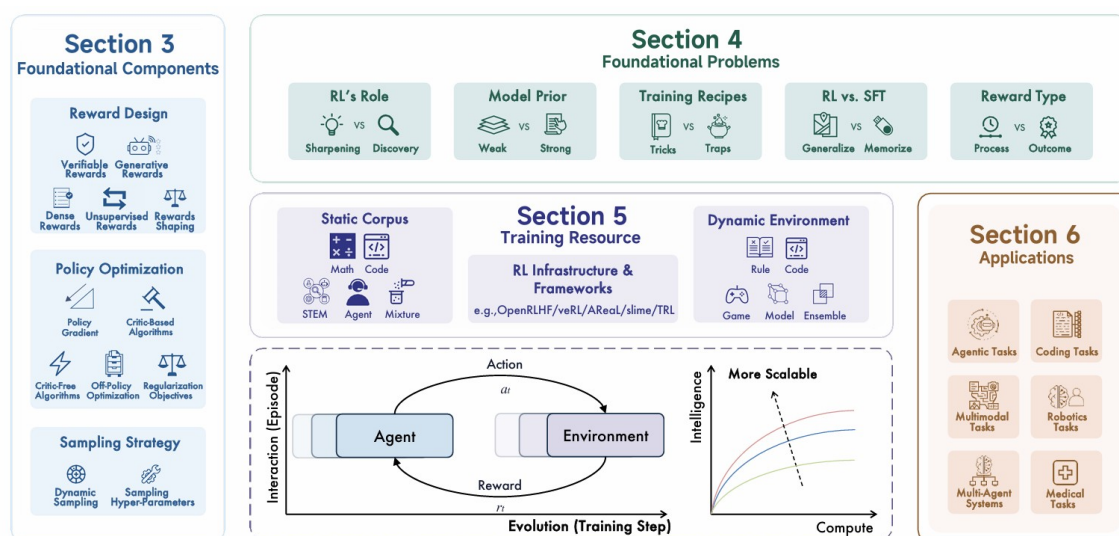


Figure 1 | Overview of the survey. We introduce the foundational components of RL for LRMs, along with open problems, training resources, and applications. Central to this survey is a focus on large-scale interactions between language agents and environments throughout long-term evolution.

主要亮点

- **全景式梳理 RL + 推理 模型架构**：这篇综述将 RL 驱动的推理模型方向（即将 LLM 变为 LRM）作为一个整体进行结构化拆分，覆盖从奖励、策略、采样到训练资源与应用的多个模块，为研究者提供了“从零到一”的系统视角。
- **深入揭示 RL 在推理任务中的核心难题**：文中针对稀疏奖励、探索效率、奖励误导、样本效率瓶颈等问题，在大模型语境下做了深入讨论，指出这些问题与传统 RL 场景中的区别与挑战，并通过案例分析体现其实际影响。
- **面向未来的路线图与研究方向**：除了总结现有成果外，作者重点提出了多条未来研究路径：如何设计可验证的奖励函数 (verifiable rewards)、如何提高探索效率、如何构建可扩展训练架构、如何让 RL-for-LRMs 向更高智能层次演进等，为后续研究指明方向。

相关链接

Paper: <https://arxiv.org/abs/2509.08827>
GitHub: <https://github.com/TsinghuaC3I/Awesome-RL-for-LRMs>

● 论文十八

Teaching LLMs to Plan: Logical Chain-of-Thought Instruction Tuning for Symbolic Planning

日期

2025-09-14

Teaching LLMs to Plan: Logical Chain-of-Thought
Instruction Tuning for Symbolic Planning

Pulkit Verma MIT CSAIL Cambridge, USA pulkitv@mit.edu	Ngoc La MIT CSAIL Cambridge, USA ntmla@mit.edu	Anthony Favier MIT CSAIL Cambridge, USA antfav24@mit.edu
Swaroop Mishra* Microsoft AI Mountain View, USA swaromishra@microsoft.com	Julie A. Shah MIT CSAIL Cambridge, USA julie_a_shah@csail.mit.edu	

内容摘要

本文介绍了一种名为 “PDDL-Instruct” 的新型指令微调框架，旨在通过逻辑思维链 (chain-of-thought) 推理来增强大语言模型 (LLMs) 的符号规划能力。该框架通过开发指令提示，引导模型进行精确的逻辑推理，以确定在给定状态下何时可以应用动作，从而使 LLMs 能够通过结构化反思来自动纠正其规划过程。该方法将规划验证分解为原子推理步骤，并将其融入指令微调中，使 LLMs 不仅能生成语法正确的计划，还能通过分步

推理来验证其逻辑有效性。实验结果表明,该方法在多个规划领域表现出色,在标准基准测试中实现了高达 94% 的规划准确率,相对于基线模型有 66% 的绝对提升。

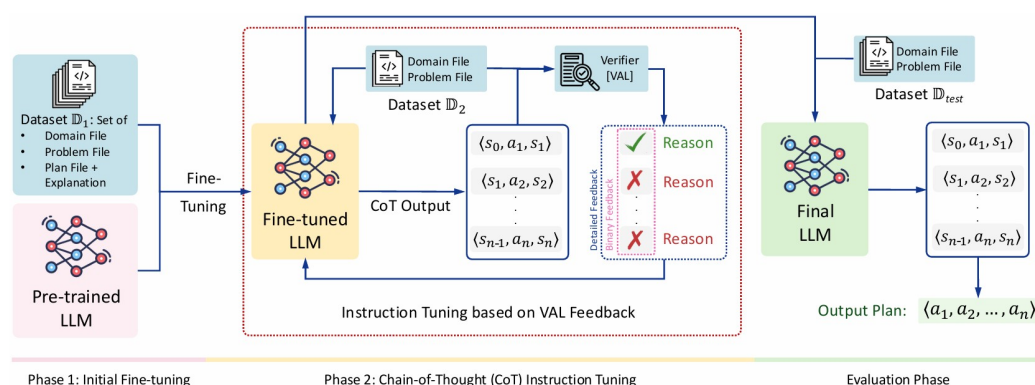


Figure 1: The PDDL-INSTRUCT approach consists of three phases: Two training phases (Initial and CoT Instruction Tuning) and evaluation phase. The main innovation lies in the second phase: CoT Instruction Tuning (highlighted by the red boundary). The initially tuned LLM is further trained using a structured instruction process that emphasizes complete logical reasoning chains.

主要亮点

- **新颖的指令微调框架, 增强 LLM 的符号规划能力:** 论文提出了 PDDL-Instruct 框架, 通过逻辑思维链推理来专门解决 LLMs 在需要形式化表示的结构化符号规划方面的局限性。
- **引入逻辑推理实现自我纠正:** 该框架通过指导模型进行精确的逻辑推理, 使其能够系统地检查先决条件、应用效果并验证不变量, 从而实现规划过程的自我纠正和逻辑有效性验证。
- **显著超越基线模型的卓越性能:** 通过指令微调, LLMs 在规划方面表现出显著提升, 在标准基准测试中的规划准确率高达 94%, 相较于基线模型实现了 66% 的绝对性能提升。

相关链接

Paper: <https://arxiv.org/abs/2509.13351>

GitHub: <https://github.com/ParallelLLC/symbolic-planning>

● 论文十九

Single-stream Policy Optimization

日期

2025-09-16

Single-stream Policy Optimization

Zhongwen Xu^{1*} and Zihan Ding^{1*}

¹Tencent, ^{*}Equal contribution.

内容摘要

本文对当前用于大型语言模型（LLMs）的策略梯度优化方法进行了反思，指出主流的基于“组”（group-based）方法（例如 GRPO）虽然能在一定程度减小方差，却存在固有缺陷：当一个 prompt 的所有采样结果落在同一类别（即“退化组”）时，会导致学习信号被完全抹除；而且在分布式训练设置中，这些方法还会因为各子任务的响应速度不一致而出现同步瓶颈。为克服这些问题，作者提出了 Single-stream Policy Optimization（SPO），它放弃了对 prompt 分组的思路，而是为每个采样独立维护一个 KL 自适应价值估计器（value tracker）作为基线，并在整个批次范围内统一归一化优势值。该设计避免了退化组的信号丢失、简化了训练流程，并且具备更好的可扩展性。进一步地，SPO 自然支持一种优先采样（prioritized sampling）机制，使得训练过程能够聚焦于更具学习价值的 prompt。作者在 Qwen3- 8B 模型上于多个困难数学基准任务（如 BRUMO25、AIME25、HMMT25 等）进行实验证明，SPO 相比 GRPO 在收敛稳定性和准确率上都有显著提升。消融研究也表明，SPO 的优势确实源自于其基线估计与归一化设计。总体而言，SPO 提出了一条更简洁、稳健且高效的策略梯度路径，有望成为 LLM 推理强化学习中的新范式。

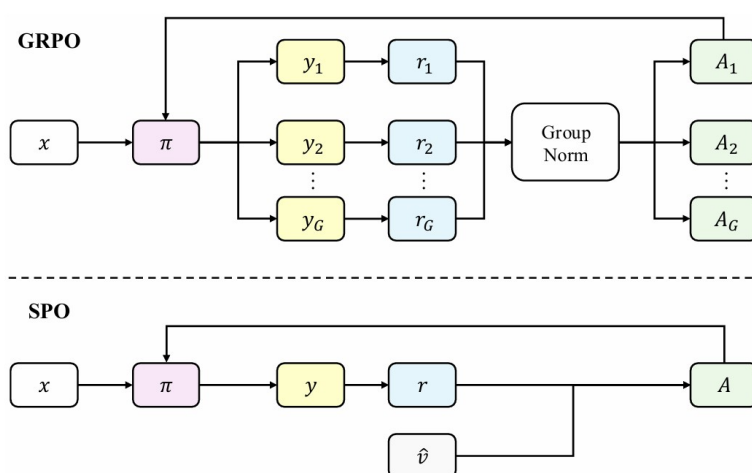


Figure 1: Illustrations of GRPO and the proposed SPO.

主要亮点

- **摆脱“组”机制带来的局限**：SPO 彻底放弃了按 prompt 分组生成多个采样然后构造即时 baseline 的做法，从架构层面消除“退化组”（所有样本同类导致优势归零）的风险，并避免了因为各子采样速度不同造成的训练同步瓶颈。
- **KL 自适应价值估计 + 批量优势归一化**：SPO 为每个 prompt 维护一个持久的 KL 自适应价值 tracker 作为 baseline，并在整个训练批次层面统一归一化优势值，这样每条样本都有稳定低方差的学习信号，从而提升训练效率和稳定性。
- **优先采样与更好扩展性**：由于 SPO 是无组（group-free）的，它在长 horizon 推理、工具调用或异质响应时间的场景下更具扩展性。此外，价值 tracker 自然支持优先采样（prioritized prompt sampling），使训练更集中于高信息量的 prompt，从而加速学习。

相关链接

Paper: <https://arxiv.org/abs/2509.13232>

● 论文二十

AToken: A Unified Tokenizer for Vision

日期

2025-09-17

ATOKEN: A UNIFIED TOKENIZER FOR VISION

Jiasen Lu* Liangchen Song* Mingze Xu Byeongjoo Ahn

Yanjun Wang Chen Chen Afshin Dehghan Yinfei Yang

Apple

内容摘要

本文提出了 AToken，这是首个能够对图像、视频和三维资产进行高保真重建和语义理解的统一视觉分词器。与现仅有针对单一模态或任务的分词器不同，AToken 利用纯 Transformer 架构将不同模态的视觉输入编码到一个共享的四维潜在空间中，从而在单个框架内统一了重建和理解任务。AToken 在下游应用中表现出色，支持多种视觉生成和理解任务，如文生视频和图生三维合成，并在所有基准测试中取得了有竞争力的性能。

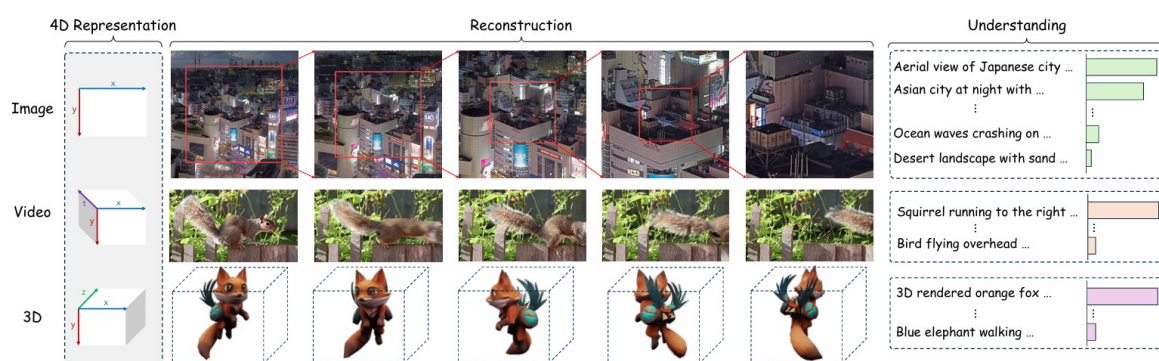


Figure 1: **Illustration of our method on different visual modalities.** Given images, videos, and 3D assets, ATOKEN leverages a shared 4D latent space (left) to produce high-fidelity reconstructions (middle: zoomed regions with red boxes for images, temporal frames for videos, multiple viewpoints for 3D) while preserving strong semantic understanding (right: showing text-aligned representations for zero-shot text retrieval).

主要亮点

- **首个统一的视觉分词器**: AToken 首次在单个框架内, 同时实现了跨图像、视频和三维资产等多种视觉模态的高保真重建和语义理解。
- **创新的四维 Transformer 架构**: 该模型引入了纯 Transformer 架构, 并采用四维旋转位置嵌入, 使其能够灵活处理任意分辨率和时长的视觉输入。
- **任务与模态的统一**: AToken 不仅在重建任务中表现出色, 还使能了多种下游应用, 例如文生视频、图生三维合成以及多模态大型语言模型 (LLMs), 展现了其卓越的通用性和性能。

相关链接

Paper: <https://arxiv.org/abs/2509.14476>

GitHub: <https://github.com/sanowl/ATOKEN-A-UNIFIED-TOKENIZER-FOR-VISION>

● 论文二十一

Discovery of Unstable Singularities

日期

2025-09-17

Discovery of Unstable Singularities

Yongji Wang^{1, 2}, Mehdi Bennani³, James Martens³, Sébastien Racanière³, Sam Blackwell³, Alex Matthews³, Stanislav Nikolov³, Gonzalo Cao-Labora^{1, 4}, Daniel S. Park⁵, Martin Arjovsky^{3, †}, Daniel Worrall^{3, †}, Chongli Qin^{3, †}, Ferran Alet^{3, †}, Borislav Kozlovskii^{3, †}, Nenad Tomašev^{3, †}, Alex Davies³, Pushmeet Kohli³, Tristan Buckmaster^{1, *}, Bogdan Georgiev^{3, *}, Javier Gómez-Serrano^{6, *}, Ray Jiang^{3, *} and Ching-Yao Lai^{2, *}

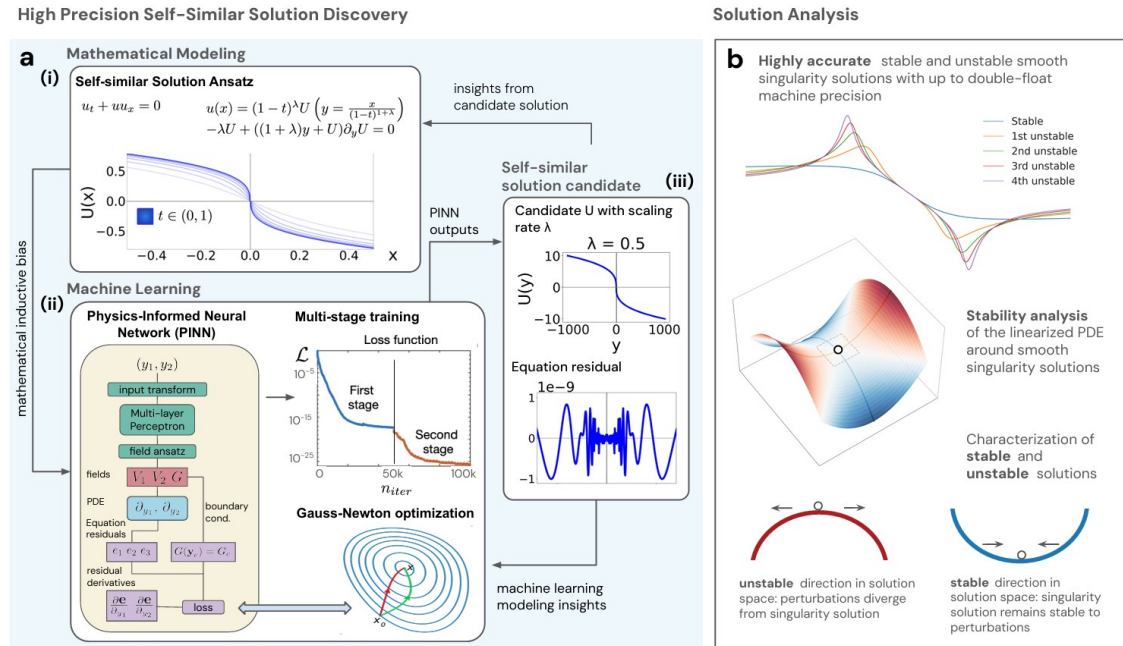


Figure 1. Research flowchart. Our research methodology consists of two main stages. **a) Solution discovery.** We start with a candidate solution that allows us to search for self-similar spatial profiles of blow-up solutions characterized by a self-similar scaling rate λ , illustrated in (i) with Burgers' Equation. We then refine our machine learning pipeline (ii) and solution accuracy using an iterative approach. Empirical results from candidate solutions (iii) and their accuracy guide the mathematical modeling and neural network architecture. Mathematical modeling in turn guides the inductive biases built into the network architecture, such as input coordinate transforms and the shaping of the output fields. We use a Physics-informed Neural Network (PINN) with a Gauss-Newton optimiser and a multi-stage refinement training scheme to generate a highly accurate candidate solution while finding the right scaling rate λ . **b) Solution analysis.** We analyze the stability of a high-precision candidate solution by linearising the PDE around it. We discover unstable modes, along which any slight perturbation causes the system to deviate from the blow-up trajectory. Thus, we characterize the degree of stability and confirm the discovery of highly accurate stable and unstable singularities.

本文首次系统地发现了流体动力学中不稳定奇点的存在，挑战了传统对奇点稳定性的理解。在流体动力学中，奇点是指解在有限时间内从光滑初始条件发展出无限梯度的现象。传统上，数值方法主要识别稳定奇点，即使初始状态稍有扰动，奇点也能稳定形成。然而，对于无边界的欧拉方程和 Navier-Stokes 方程等关键开放问题，假设存在不稳定奇点被认为至关重要。本研究通过结合精心设计的机器学习架构和高精度的 Gauss-Newton 优化器，首次发现了多种新的不稳定自相似解。这些解仅在初始条件精确到无限位数时才能形成，且对微小扰动极为敏感，立即偏离其爆炸轨迹。特别地，研究揭示了不可压缩多孔介质方程和具有边界的三维欧拉方程的新不稳定解，并提出了一个简单的经验性渐近公式，将爆炸速率与不稳定性阶数联系起来。对于某些解（如 Córdoba-Córdoba-Fontelos (CCF) 稳定解和第 1 不稳定解），研究达到了接近双精度浮点数

的机器精度，精度仅受限于 GPU 硬件的舍入误差。这一精度水平满足了通过计算辅助证明进行严格数学验证的严格要求。

主要亮点

- **首次发现不稳定奇点：**系统地发现了流体动力学中不稳定奇点的存在，挑战了传统对奇点稳定性的理解。
- **高精度数值方法：**结合机器学习架构和高精度优化器，达到了接近双精度浮点数的机器精度，为严格的数学验证提供了可能。
- **新解的物理意义：**揭示了不可压缩多孔介质方程和三维欧拉方程中的新不稳定解，为理解流体动力学中的奇点形成提供了新的视角。

相关链接

Paper: <https://arxiv.org/pdf/2509.14185>

● 论文二十二

MANZANO: A Simple and Scalable Unified Multimodal Model with a Hybrid Vision Tokenizer

日期

2025-09-19

MANZANO: A Simple and Scalable Unified Multimodal Model with a Hybrid Vision Tokenizer

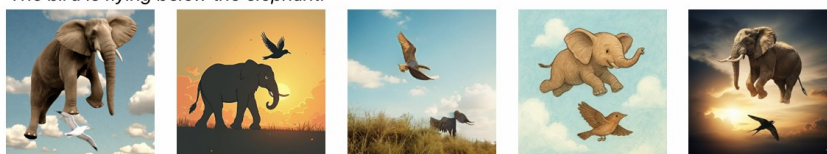
Yanghao Li^{o†}, Rui Qian^o, Bowen Pan^o, Haotian Zhang^o, Haoshuo Huang^o, Bowen Zhang^{o†}, Jialing Tong^{*}, Haoxuan You^{*}, Xianzhi Du^{*}, Zhe Gan^{*}, Hyunjik Kim^{*}, Chao Jia^{*}, Zhenbang Wang^{*}, Yinfei Yang^{*}, Mingfei Gao, Zi-Yi Dou, Wenze Hu, Chang Gao, Dongxu Li, Philipp Dufter, Zirui Wang, Guoli Yin, Zhengdong Zhang, Chen Chen, Yang Zhao, Ruoming Pang[‡], Zhifeng Chen

Apple

^oFirst authors ^{*}Core authors [†]Project lead [‡]Work done at Apple

内容摘要

The bird is flying below the elephant.



A photo of a corgi with a sign that says 'I am not a real corgi'.



A hand-drawn blueprint for a time machine, with the caption 'Time Traveling Device'.



Manzano

Bagel

Janus Pro

GPT-4o

Nano Banana

Figure 1 Qualitative text-to-image generation on challenging prompts. Manzano handles counterintuitive, physics-defying prompts (e.g., 'The bird is flying below the elephant') comparably to GPT-4o [35] and Nano Banana [17].

Manzano 是一个简单且可扩展的统一多模态大型语言模型 (LLM) 框架, 旨在通过混合视觉分词器 (hybrid vision tokenizer) 和精心设计的训练配方, 解决现有开源模型在视觉理解和生成能力之间存在的性能权衡问题。该模型采用单个共享的视觉编码器, 连接到两个轻量级适配器, 分别用于生成用于理解的连续嵌入和用于生成

的分立标记。这种架构与统一的训练方法相结合，实现了两种能力的联合可扩展学习，并能最小化任务冲突。实验结果表明，Manzano 在统一模型中达到了最先进的水平，并且与专业模型相比也具有竞争力，尤其是在处理富文本评估任务时表现突出。

主要亮点

- **统一的简单可扩展多模态框架：**Manzano 框架通过统一的设计，有效解决了现有模型在视觉理解和生成任务中性能相互制约的问题，实现了简单而可扩展的联合学习。
- **创新的混合视觉分词器：**该模型的核心创新在于其混合视觉分词器，它通过单个共享编码器为理解和生成任务分别生成连续和分立的标记，从而在统一的语义空间内高效地处理不同任务。
- **卓越的性能与可扩展性：**Manzano 不仅在统一多模态模型中取得了最先进的成果，并且能够与专业模型相媲美。研究表明，该模型设计能将任务冲突降至最低，且能从模型规模的扩展中持续获得性能提升，验证了其设计的优越性。

相关链接

Paper: <https://arxiv.org/abs/2509.16197>

GitHub: https://github.com/AkihikoWatanabe/paper_notes/issues/2925

● 论文二十三

Open-World Drone Active Tracking with Goal-Centered Rewards

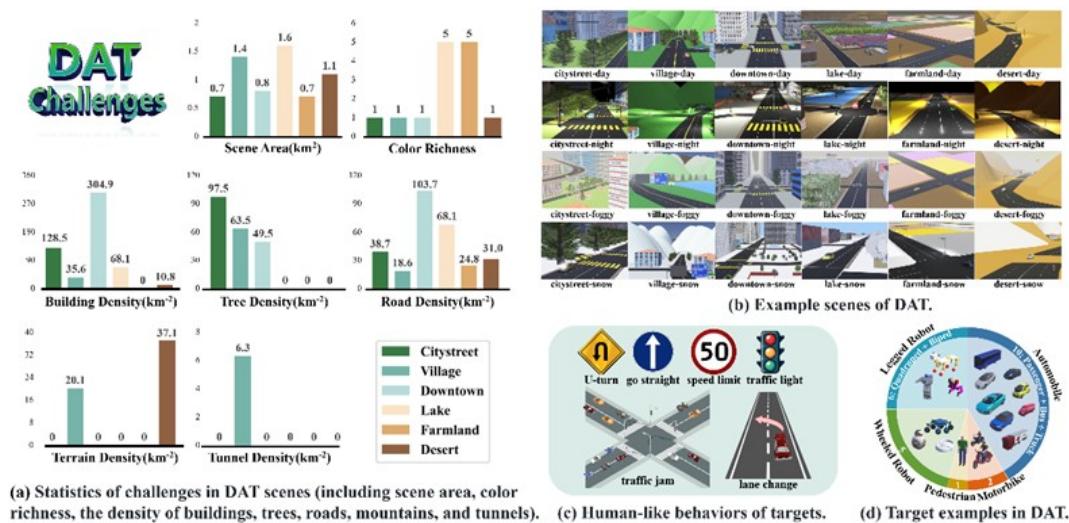
日期

2025-09-19

Open-World Drone Active Tracking with Goal-Centered Rewards

Haowei Sun^{1,2*} Jinwu Hu^{1,3*} Zhirui Zhang^{1,2} Haoyuan Tian^{1,2} Xinze Xie^{1,2}
Yufeng Wang^{1,4} Xiaohua Xie⁵ Yun Lin⁶ Zhuliang Yu^{1,2 [8]} Mingkui Tan^{1 [8]}

内容摘要



华南理工大学，超级机器人研究院（黄埔），琶洲实验室与鹏城实验室联合发布了统一的无人机视觉主动跟踪基准 DAT。该基准涵盖 6 种场景与 4 种天气条件，共构建了 24 个大范围、高复杂度的仿真场景，包含多种跟踪目标和高精度的动力学模拟，可充分验证无人机主动跟踪智能体在跨场景和跨域条件下的迁移能力。针对当前主动跟踪领域存在的视角受限问题，该工作提出了以目标为中心的视觉主动跟踪算法 GC-VAT。该方法

设计了以目标为中心的奖励函数，能够在视角变化时提供准确的反馈；同时，采用基于课程学习的训练策略，逐步提升智能体在复杂场景中的跟踪性能。在仿真器和真实图像数据集上的实验结果表明，该方法在复杂场景无人机视觉主动跟踪任务上显著优于现有最优算法。

主要亮点

- **提出了统一的无人机主动跟踪基准 DAT：**能够充分验证智能体在跨场景和跨域条件下的迁移能力。
- **提出了以目标为中心的主动跟踪算法 GC-VAT：**利用以目标为中心的奖励函数和基于课程学习的训练策略，显著提升复杂场景下无人机的跟踪性能。
- **全面实验说明基准和方法的有效性：**在仿真器和真实图像上的实验充分验证了 DAT 基准和 GC-VAT 方法的有效性。

相关链接

Paper: <https://arxiv.org/pdf/2412.00744>

GitHub: https://github.com/SHWplus/DAT_Benchmark

TECHNICAL UPDATES

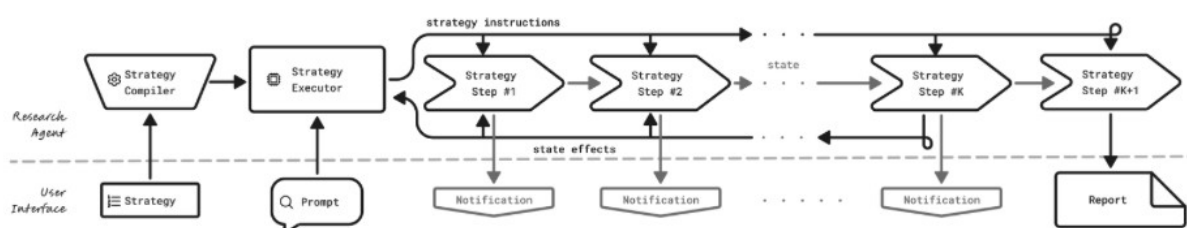
技术动态

● 技术动态一

英伟达开源通用深度研究大模型 Universal Deep Research

日期

2025-08-29



An illustration of the UDR setup. The deep research agent takes both the research strategy and research prompt as inputs, compiles the strategy from natural language into an actionable research orchestration code snippet, and then carries out the strategy, delivering the final report to the user.

内容摘要

英伟达研究院正式推出通用 Dixit-Stiglitz 表征 (Universal Dixit-Stiglitz Representation, UDR)，一种为强化学习 (RL) 设计的全新世界模型。该模型基于诺贝尔经济学奖得主 Avinash Dixit 和 Joseph Stiglitz 的理论，通过将环境状态分解为一系列可区分的实体及其属性，从而学习环境的动态变化。UDR 在多种模拟环境中展现了卓越的性能和样本效率，仅用十分之一的训练数据便能在 Crafter 基准测试中超越现有最强模型，并在 Atari 100k 基准测试中取得了与领域内领先方法相当的成绩。UDR 的核心技术亮点在于其创新的实体化表征学习方法和高效的架构设计。模型通过一个基于 Transformer 的实体提议模块来识别场景中的关键实体，

并利用其动态预测器来模拟这些实体间的相互作用和未来状态。这种基于实体的方法不仅提升了模型的泛化能力和计算效率，还增强了模型的可解释性，使其能够更好地理解和预测复杂的环境动态。英伟达已将 UDR 的官方 PyTorch 实现代码在 GitHub 上开源，并发布了相关的技术论文。此举旨在推动强化学习领域的研究，为开发更高效、更通用的智能体提供强大的新工具，并促进其在机器人、自动驾驶和游戏等领域的应用发展。

相关链接

技术博客: <https://research.nvidia.com/labs/lpr/udr/>

GitHub: <https://github.com/NVlabs/UniversalDeepResearch>

● 技术动态二

腾讯开源 WTM2025 冠军大模型

日期

2025-09-01



内容摘要

Hunyuan-MT 系列模型正式开源,包含翻译模型 Hunyuan-MT-7B 和集成模型 Hunyuan-MT-Chimera, 支持 33 种语言的双向互译, 其中涵盖 5 种中国少数民族语言。Hunyuan-MT-7B 在同规模模型中实现行业

领先表现，而 Hunyuan-MT-Chimera 是业界首个开源的翻译集成模型，显著提升了翻译结果的整体质量。

在 WMT25 国际翻译评测中，Hunyuan-MT 在 31 个参评语种方向中斩获 30 个第一，展现出卓越的跨语言能力。该系列模型提出了完整的训练体系，包括预训练、持续预训练、监督微调、翻译强化学习以及集成强化学习，在同规模模型中取得了最先进水平。模型权重已全面开源，可广泛支持学术研究与产业应用，为多语言交流与全球化场景提供高质量支撑。

相关链接

Hugging Face: <https://huggingface.co/tencent/Hunyuan-MT-7B>

GitHub: <https://github.com/Tencent-Hunyuan/Hunyuan-MT/>

在线体验: <https://hunyuan.tencent.com/modelSquare/home/list>

● 技术动态三

阿里推出最新旗舰模型 Qwen3-Max-Preview

日期

2025-09-01

内容摘要

阿里巴巴通义千问团队正式发布最新旗舰大语言模型 Qwen3-Max-Preview。这是 Qwen3 系列中规模最大的模型，参数量超过 1 万亿，支持 100 多种语言，具备更强的推理、指令跟随和多语言理解能力。Qwen3-Max-Preview 在数学、编程和科学推理等任务中表现突出，能更准确地执行复杂指令，显著减少幻觉，生成更高质量的内容。其上下文窗口扩展至 256,000 tokens，可处理长篇文档和复杂文本，适用于问答、写

作、编程辅助、跨语言交流等多场景应用。用户可通过 Qwen Chat 网页免费体验，或在阿里云百炼平台通过 API 接入，实现灵活的企业级部署。



相关链接

在线链接: <https://chat.qwen.ai/>

● 技术动态四

Google 发布面向嵌入式应用的大模型 EmbeddingGemma

日期

2025-09-04



内容摘要

谷歌 EmbeddingGemma 系列模型正式发布，包含 2B 和 7B 两种规模。该系列模型是谷歌最新推出的文本嵌入模型，基于 Gemma 模型架构，在性能和效率上均有显著提升。EmbeddingGemma 在 MTEB 基准测试中取得了领先的成绩，同时保持了较小的模型体积和较低的计算成本，使其能够在 CPU 上高效运行。其核心技术亮点包括采用 Matryoshka 表示法，可在不牺牲准确性的前提下缩短嵌入维度，以及支持在不同设备上进行微调和部署。该系列模型已在 Kaggle 和 Colab 上提供，并通过 Vertex AI 和 Hugging Face 向开发者开放，旨在推动文本嵌入技术在学术研究和产业应用中的发展。

相关链接

博客: <https://developers.googleblog.com/en/introducing-embeddinggemma/>

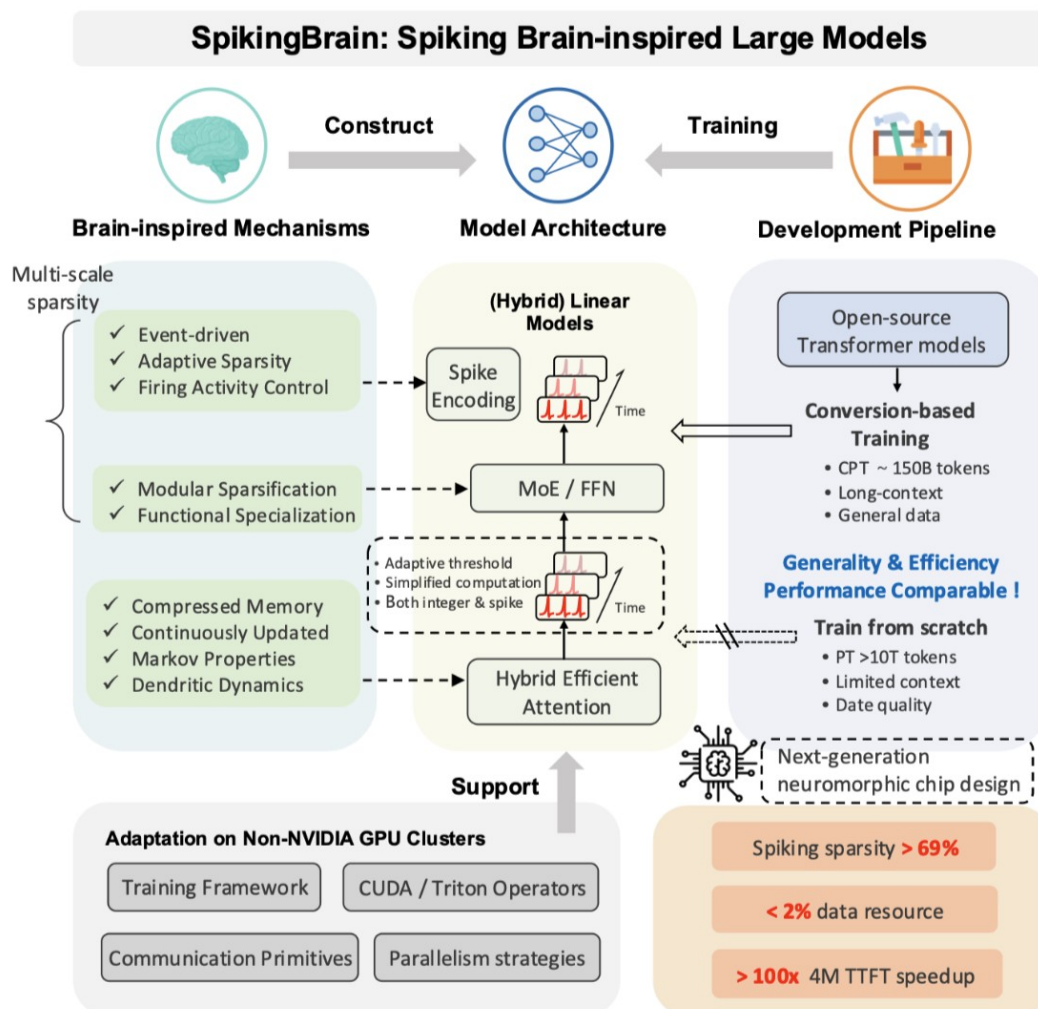
Hugging Face: <https://huggingface.co/google/embeddinggemma-300m>

技术动态五

类脑脉冲大模型 “瞬息 1.0” 发布

日期

2025-09-05



内容摘要

清华大学脑启发计算实验室（BICLab）正式发布 SpikingBrain-7B 模型。该模型受人脑机制启发，将高效混合注意力机制、专家混合（MoE）模块和脉冲编码融入其架构，旨在突破传统大语言模型在处理长序列信息时面临的效率瓶颈。SpikingBrain-7B 及其混合专家版本 SpikingBrain-76B 展现了在长序列训练和推理上

的显著效率提升，并成功在非英伟达（NVIDIA）计算平台（天数智芯 MetaX GPU）上实现了大规模模型的稳定训练与部署。该模型的核心技术亮点在于其创新的脉冲神经网络设计和高效的训练策略。通过独特的基于转换的训练流程，SpikingBrain 能够利用开源模型（如 Qwen2.5-7B）作为基础，仅使用约 1500 亿个词元进行持续预训练，便可达到与主流开源模型相当的性能。其脉冲机制实现了高达 69.15% 的稀疏性，大幅降低了功耗。在处理 400 万词元的超长序列时，“首字生成时间”相较于传统模型加速超过 100 倍。SpikingBrain-7B 的完整实现、模型权重及量化版本均已在 GitHub 上开源，并提供了包括 HuggingFace 版本和 vLLM 推理版本在内的灵活部署选项，为学术研究和产业应用提供了有力支持。

相关链接

技术报告：https://github.com/BICLab/SpikingBrain-7B/blob/main/SpikingBrain_Report_Chi.pdf

GitHub：<https://github.com/BICLab/SpikingBrain-7B>

试用链接：<https://controller-fold-injuries-thick.trycloudflare.com>

● 技术动态六

通义千问发布 Qwen3 ASR 模型：智能语音识别新标杆，支持歌声与定制化

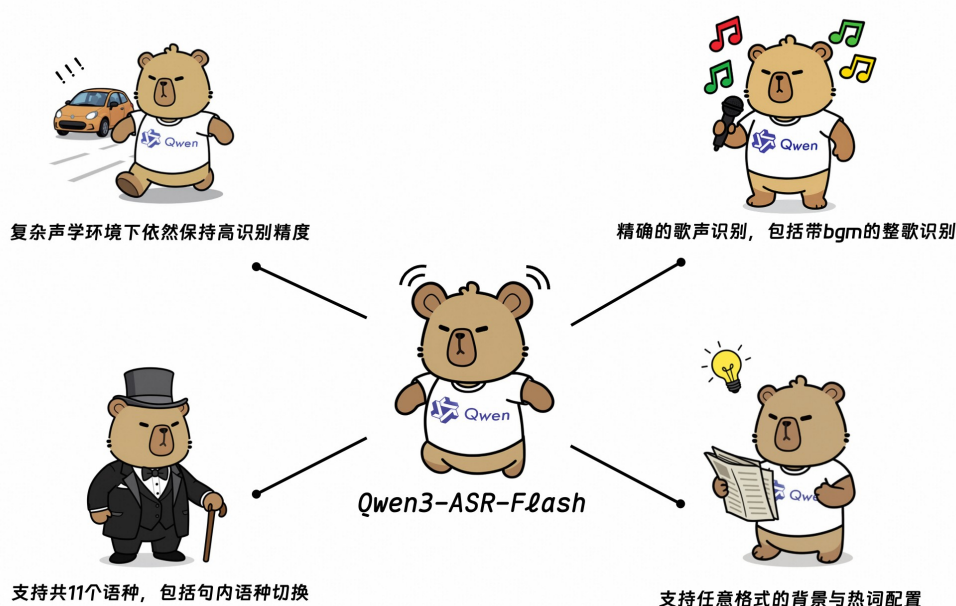
日期

2025-09-08

内容摘要

阿里巴巴通义千问团队正式发布全新的 Qwen3-ASR-Flash 语音识别服务。该模型基于 Qwen3 大语言模型，并结合海量多模态数据及千万小时级别的 ASR 数据进行构建，实现了高精度、高鲁棒性的语音识别性能。Qwen3-ASR-Flash 在多个中英文及多语言基准测试中表现出色，达到了业界领先水平。该模型的核心技

术亮点在于其强大的定制化识别能力和歌声识别功能。用户可以提供任意格式的文本（如词汇表、段落、文档）作为背景上下文，模型能智能利用这些信息，精准识别并匹配命名实体等关键术语，输出定制化的转写结果。此外，模型在歌声识别方面取得了突破，无论是清唱还是带背景音乐的歌曲，实测错误率均低于 8%。Qwen3-ASR-Flash 单模型支持包括中、英、法、德、日、韩在内的 11 种语言及其多种方言和口音，并能精确进行语种识别和非人声过滤，在复杂声学环境下依然保持高准确率。



相关链接

博客：

<https://qwen.ai/blog?id=41e4c0f6175f9b004a03a07e42343eaaf48329e7&from=research.latest-advancements-list>

试用链接：<https://huggingface.co/spaces/Qwen/Qwen3-ASR-Demo>

● 技术动态七

腾讯开源图像生成大模型混元 Image2.1

日期

2025-09-09



内容摘要

腾讯正式发布并开源混元图像大模型 2.1 版本（Hunyuan-Image 2.1）。该模型在中文原生文生图领域实现了业界领先水平，特别是在中国元素、古诗词意境、人像及动漫风格的图像生成方面表现突出。

Hunyuan-Image 2.1 凭借其对中文语义的深度理解和高质量的图文对齐，能够生成更符合中文语境和审美需求的图像作品，在多个权威文生图评测基准中取得了优异成绩。该模型的核心技术亮点在于其采用了与 Sora 和 Stable Diffusion 3 相似的 Diffusion Transformer (DiT) 架构，并结合了腾讯自研的“T5+CLIP”双文本编码器，显著提升了对中文长文本和复杂语义的理解能力。此外，模型通过优化训练策略，实现了高效的训练和推理，并引入了美学评分系统和 LoRA 微调支持，让用户能够更灵活地生成和定制高质量图像。

Hunyuan-Image 2.1 的模型权重已在 GitHub 上开源，并提供了便捷的在线体验，旨在为开发者和创作者提供强大的文生图工具，推动相关领域的研究与应用创新。

相关链接

GitHub: <https://github.com/Tencent-Hunyuan/HunyuanImage-2.1>

Hugging Face: <https://huggingface.co/tencent/HunyuanImage-2.1>

● 技术动态八

月之暗面 Kimi-K2 大模型正式开源

日期

2025-09-05



内容摘要

月之暗面 (Moonshot AI) 正式开源其最新的 Kimi K2 大模型。该模型是一个拥有 1 万亿总参数和 320 亿激活参数的混合专家 (MoE) 模型，不仅在前沿知识、数学和编码等基准测试中达到了业界顶尖水平，更是针对“智能体” (Agentic) 任务进行了深度优化，使其不仅能“回答”，更能通过自主调用工具来“行动”。在多个权威评测中，Kimi K2 的性能已追平甚至超越了部分领先的开源及闭源模型。Kimi K2 的核心技术亮点在于其创新的训练方法和智能体能力构建。首先，它采用自研的 MuonClip 优化器，通过 qk-clip 技术有效解决了大规模训练中的不稳定性难题，实现了在 15.5T tokens 数据上的稳定预训练。其次，模型通过大规模智能体数据合成和通用强化学习 (General RL) 系统，学习了复杂的工具使用能力，能够自主完成从数据分析、代码开发到行程规划等一系列复杂任务，展现了卓越的规划与执行能力。此次开源包括 Kimi-K2-Base 基础模型和 Kimi-K2-Instruct 指令微调模型，支持 256K 超长上下文。开发者可通过 Kimi 官网、API 平台或使用 vLLM 等框架进行本地部署，推动开放式智能体技术的发展和应用。

相关链接

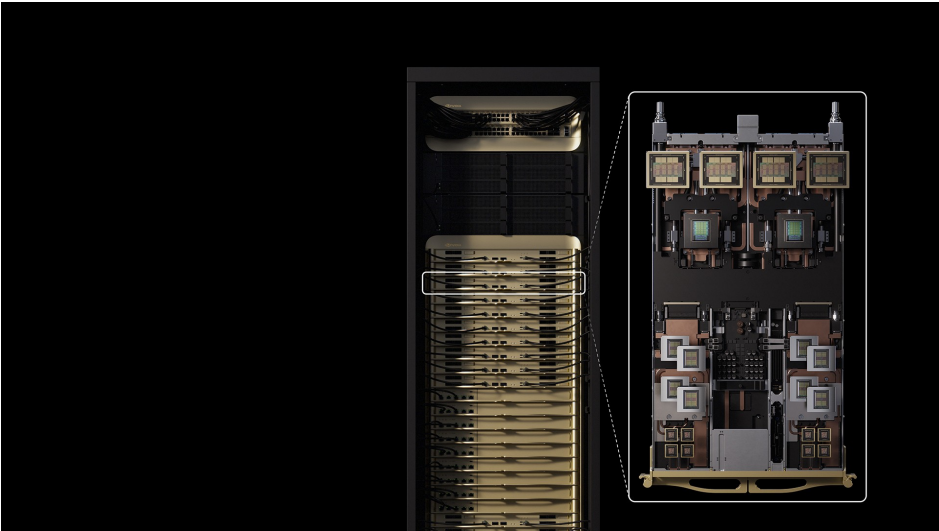
技术博客: <https://moonshotai.github.io/Kimi-K2/>
技术报告: <https://arxiv.org/abs/2507.20534>
Github: <https://github.com/moonshotai/kimi-K2>

● 技术动态九

英伟达发布 Rubin-CPX，为大规模上下文推理量身打造的新型 GPU

日期

2025-09-09



内容摘要

英伟达 (NVIDIA) 正式发布了 Rubin-CPX，一款专为处理大规模上下文推理而设计的全新 GPU。该处理器旨在高效处理百万级 token 的软件编码和生成式视频等复杂 AI 任务，为长上下文处理提供了前所未有的性能与效率。Rubin-CPX 将与 NVIDIA Vera CPU 和 Rubin GPU 协同工作于全新的 NVIDIA Vera Rubin

NVL144 CPX 平台中，该平台在一个机架内集成了高达 8 exaflops 的 AI 算力、100TB 的高速内存和 1.7 PB/s 的内存带宽，性能相较于 NVIDIA GB300 NVL72 系统提升了 7.5 倍。Rubin-CPX 的核心技术亮点在于其专为大规模上下文 AI 设计的 CUDA GPU 架构，能够让模型一次性处理和理解数百万 token 的知识。该 GPU 采用 NVFP4 精度可提供高达 30 petaflops 的算力，并配备 128GB 的 GDDR7 内存，其注意力机制处理速度比 GB300 NVL72 系统快 3 倍，大幅提升了处理长序列的能力。此外，Rubin-CPX 还集成了视频编解码器，为视频搜索和高质量生成式视频等长格式应用提供了单芯片解决方案。英伟达的这一发布标志着 AI 计算领域的又一次飞跃，通过解耦推理（disaggregated inference）的方式，由 Rubin-CPX 专注于处理计算密集的上下文阶段，从而显著提升效率和投资回报率。该产品预计将于 2026 年底上市，并已获得 Cursor、Runway 和 Magic 等 AI 创新公司的关注，有望加速其应用的开发进程。

相关链接

博客：

<https://nvidianews.nvidia.com/news/nvidia-unveils-rubin-cpx-a-new-class-of-gpu-designed-for-massive-context-inference>

● 技术动态十

阿里巴巴发表 MOE 新模型 Qwen3-Next 系列

日期

2025-09-11

内容摘要

通义千问团队正式发布并开源 Qwen3-Next 系列模型，该系列模型旨在解决未来大模型在长上下文扩展和总参数扩展两大趋势下的效率瓶颈。基于全新的 Qwen3-Next 架构，此次发布的 Qwen3-Next-80B-A3B

模型总参数量达 800 亿,但推理时仅激活 30 亿参数。该模型以不到 Qwen3-32B 模型十分之一的训练成本,实现了相近甚至更优的性能,并在 32k 以上长上下文的推理吞吐量上提升了十倍以上,达到了极致的训练与推理性价比。Qwen3-Next 的核心技术亮点在于其创新的模型结构:首先,它采用了 Gated DeltaNet 与标准注意力相结合的混合注意力机制,在保证长序列建模效率的同时,维持了强大的上下文学习能力;其次,通过将专家数量扩展至 512 个,实现了高达 3.7% 激活率的极致稀疏 MoE 架构,大幅提升了资源利用率;此外,模型还引入了多项训练稳定性优化(如 Zero-Centered RMSNorm)和多令牌预测机制(Multi-Token Prediction),在保证训练稳定性的同时,显著提高了推理效率。此次同步开源了

Qwen3-Next-80B-A3B-Instruct 和 Qwen3-Next-80B-A3B-Thinking 两个对齐版本。其中,Instruct 模型在 256K 超长上下文任务中表现出显著优势,性能媲美旗舰模型 Qwen3-235B;Thinking 模型则在复杂推理任务上表现卓越,在多个基准测试中超越了同量级的开源模型及部分闭源模型。目前,Qwen3-Next 系列模型已在 Hugging Face 和 ModelScope 开源,并可在阿里云百炼及 NVIDIA API Catalog 上体验,为开发者提供了全流程的支持。



相关链接

Hugging Face: <https://huggingface.co/collections/Qwen/qwen3-next-68c25fd6838e585db8eeea9d>

博客: <https://modelscope.cn/collections/Qwen3-Next-c314f23bd0264a>

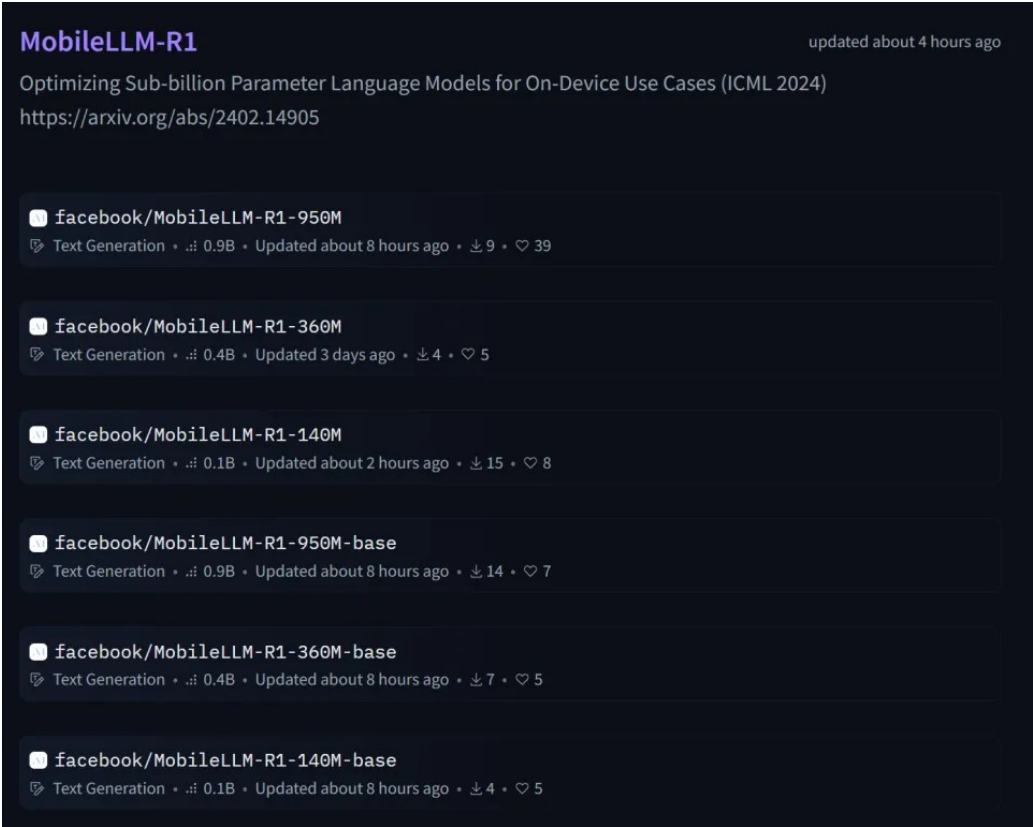
Kaggle: <https://kaggle.com/models/qwen-lm/qwen3-next-80b>

● 技术动态十一

Meta 发布 MobileLLM-R1 端侧模型

日期

2025-09-13



内容摘要

Meta 公司近日正式开源其 MobileLLM-R1 系列模型，这是一组专为在手机等移动设备上高效运行而设计的轻量化语言模型。该系列包含 1.4 亿、3.6 亿和 9.5 亿三种参数规模的版本，旨在以更低的计算资源消耗，为端侧 AI 应用提供强大的推理能力。尽管模型参数量均小于 10 亿，但在多个基准测试中，其性能表现优于部分更大规模的模型，展现了卓越的效率与潜力。MobileLLM-R1 的核心技术亮点在于其“精简”与“专攻”的设计理念。该模型并非通用的聊天机器人，而是通过监督式微调（SFT）技术，专注于解决数学、编程和科学问题等特定的推理任务。模型采用了精心设计的 Transformer 架构，在保持模型表达能力的同时，显

著降低了计算和内存需求，使其能够原生在移动设备上部署运行。此外，尽管训练数据量远小于一些主流模型，MobileLLM-R1 依然通过高效的训练策略和高质量的数据集，在 MATH、GSM8K 等多项评测中取得了领先成绩。Meta 已通过 Hugging Face 平台，以 Apache 2.0 授权协议的形式开放了 MobileLLM-R1 系列模型，并支持 vLLM 等主流推理引擎，方便开发者快速部署和集成。此次开源将进一步推动 AI 模型在移动设备和边缘计算场景的应用，为开发者构建创新的端侧智能应用提供了更多可能性。

相关链接

Hugging Face: <https://huggingface.co/collections/facebook/mobilellm-r1-68c4597b104fac45f28f448e>

试用链接: <https://huggingface.co/spaces/akhaliq/MobileLLM-R1-950M>

● 技术动态十二

World Labs 发布 3D 世界生成模型 Marble

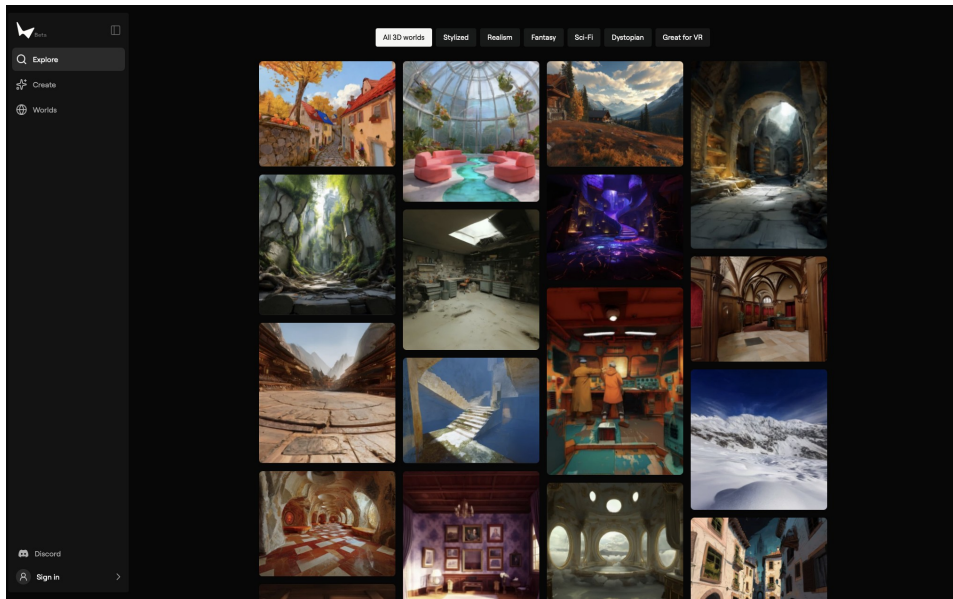
日期

2025-09-18

内容摘要

由斯坦福大学 AI 领域先驱李飞飞联合创立的空间智能公司 World Labs，近日发布了其最新的 3D 世界生成式模型的有限访问测试版 Marble。该模型旨在突破当前人工智能主要基于文本和二维图像的局限，向能够感知、生成并与三维物理世界交互的“大型世界模型”（Large World Models）出关键一步。用户只需提供一张图片或一段文字，Marble 就能生成一个可供用户在其中自由导航和探索的、持久且一致的 3D 世界，且没有时间限制。Marble 的核心技术亮点在于其卓越的生成能力和对空间智能的深度探索。相比之前的版本，新模型生成的 3D 世界规模更大、风格更多样，并且拥有更清晰的几何结构。它能够从一张图片“想象”出画

面之外的场景，构建出比输入信息更丰富的完整 3D 环境。用户可以将生成的 3D 世界导出为高斯泼溅（Gaussian splats）格式，以便在游戏引擎、VR/AR 应用等下游项目中使用。为此，World Labs 还开源了其渲染库 Spark，以方便开发者将模型无缝集成到各类 3D 体验中。目前，用户可以通过 marble.worldlabs.ai 申请体验 Marble 的测试版。World Labs 的这一进展为电影制作、游戏开发、虚拟现实等领域提供了强大的创作工具，有望彻底改变创作者构建和交互数字内容的方式。



相关链接

博客地址: <https://www.worldlabs.ai/blog/bigger-better-worlds>

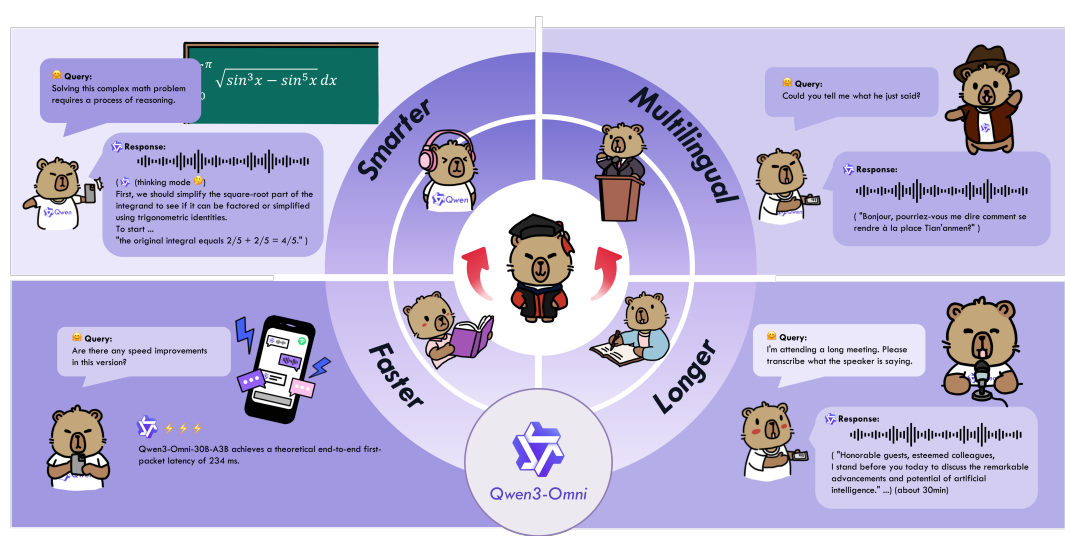
试用地址: <https://marble.worldlabs.ai/>

● 技术动态十三

通义千问 Qwen3-Omni 开源，引领原生全模态 AI 新范式

日期

2025-09-22



内容摘要

阿里巴巴通义千问团队正式开源其新一代原生全模态大模型 Qwen3-Omni。该模型能够无缝处理和理解文本、图像、音频和视频等多种形式的任意交错输入，在 36 项公开多模态评测基准中，有 32 项取得了开源模型中的最佳性能，其中 22 项达到了世界顶尖水平（SOTA）。Qwen3-Omni 不仅能进行复杂的图文音视混合内容的问答和推理，还支持实时的多语言语音对话，展现了强大的多任务处理和交互能力。Qwen3-Omni 的核心技术亮点在于其原生的全模态架构，它将不同模态的数据统一编码到一个共享的表示空间，实现了深度跨模态融合与理解，而非简单的模态拼接。该模型支持高达 100 万 token 的超长上下文处理，并能同时理解多达 40 张图片或 40 段音频。此外，模型在语音处理上表现突出，其语音识别（ASR）性能超越了 Whisper-v3 和 SeamlessM4T-v2 等领先模型，并支持多种语言的实时语音合成（TTS），为构建下一代多模态 AI 助理和应用提供了强大的基础。目前，Qwen3-Omni 的模型权重、代码以及在线试玩 Demo 均已在 GitHub 和 Hugging Face 上发布，并采用通义千问开源协议，为学术研究和商业应用提供了全方位的支持，旨在推动全

模态 AI 技术的创新与普及。

相关链接

技术报告: <https://arxiv.org/abs/2509.17765>

Github: <https://github.com/QwenLM/Qwen3-Omni>

Hugging Face: <https://huggingface.co/spaces/Qwen/Qwen3-Omni-Demo>

魔塔社区: <https://modelscope.cn/studios/Qwen/Qwen3-Omni-Demo>

● 技术动态十四

阿里通义大模型家族七连发, Qwen3-Max 领衔刷新 SOTA

日期

2025-09-24



内容摘要

阿里巴巴正式发布了通义大模型家族的七款新成员，覆盖语言、视觉、语音、多模态、代码等多个方向，旨在构建“AI 时代的 Android”。本次发布由总参数超过万亿的旗舰模型 Qwen3-Max 领衔，该模型在代码、推理及智能体能力上均达到全球顶尖水平，在多个基准测试中表现优异。此外，备受瞩目的下一代架构模型 Qwen3-Next 也同步亮相，以仅 3B 的激活参数实现了媲美 235B 密集模型的性能，训练成本降低 90%，长文本推理效率提升十倍以上。此次发布的模型阵容还包括：Qwen3-VL：视觉语言模型，从“识图”进化到“推理与执行”，能够操作电脑手机界面，甚至将草图直接生成代码。Qwen3-Omni：全模态模型，能同时处理文本、图像、音视频，并在 36 项公开评测中取得 32 项开源第一。Qwen3-Coder：编程模型全面升级，提升了代码生成速度与安全性。Wan2.5-preview：视觉生成模型，首次实现了音画同步的电影级视频生成，时长提升至 10 秒，并支持 1080P 高清画质。通义百聆 Fun：全新的语音大模型家族，提供上百种音色，覆盖客服、直播、有声书等多种场景。阿里巴巴集团 CEO 吴泳铭表示，通义千问将坚持开源开放，目标成为“AI 时代的 Android”，并计划在未来三年投入超 3800 亿元用于云和 AI 基础设施建设。目前，通义大模型全球下载量已超 6 亿次，衍生模型超 17 万个，是中国应用最广泛、最受全球开发者欢迎的开源模型。

相关链接

博客地址：<https://qwen.ai/research>

RESOURCE SHARING

资源分享

● 资源分享一

AI Agents for Beginners - A Course

类型：课程资源

内容提要

微软 AI 智能体入门课程，涵盖构建 AI 代理的基础知识，包含多个主题的课程，支持多种语言。用户可通过克隆或分叉仓库获取课程材料，运行、测试和修改代码，还能在 GitHub Codespaces 中操作课程代码。目前，该项目的 start 数超过 40k。

资源链接

<https://github.com/microsoft/ai-agents-for-beginners>

人工智能前沿快报

Artificial Intelligence
Newslettler



2025 年第 09 期

总第 27 期

广东省图象图形学会

主办

广东省图象图形学会

本期编委

金连文	华南理工大学
谭明奎	华南理工大学
钟亦武	香港中文大学
林上港	华南农业大学
李斌	深圳大学
谢晓华	中山大学
王泽宇	香港科技大学 (广州)
李冠彬	中山大学
熊龙飞	金山办公
段纪伟	金山办公
黄家荣	金山办公
李元满	深圳大学
严沛	深圳大学
周越	深圳大学
余阳鑫	深圳大学
詹天帅	深圳大学
许毅平	华南农业大学
陈嘉杜	华南农业大学

人工智能 前沿快报

2025 年第 09 期

总第 27 期

— 2025. 9.30

声明：

- (1) 本快报内容提要、文章亮点等部分内容由 AI 生成，不一定准确及全面，文章完整内容及思想应以原文为准。
- (2) 本文大部分插图来源于相关论文或文章，版权归原作者或原出版社所有。
- (3) 本快报摘录文章观点不代表编委会及主办方立场。



学会官方公众号

广东省图象图形学会

GDSIG

学术交流、技术咨询、科学普及