

- 2026 2 16

RL (RLHF → DPO → GRPO → RLVR) **VLA/Agent**

1: FLAC - Maximum Entropy RL via Kinetic Energy Regularized Bridge Matching

: []

- : FLAC: Maximum Entropy RL via Kinetic Energy Regularized Bridge Matching
- **ArXiv ID:** 2602.12829
- : Xiao Ma, Fuchun Sun, Yu Luo, Yunfei Li, Lei Lv
- :
- : 2026-02-13

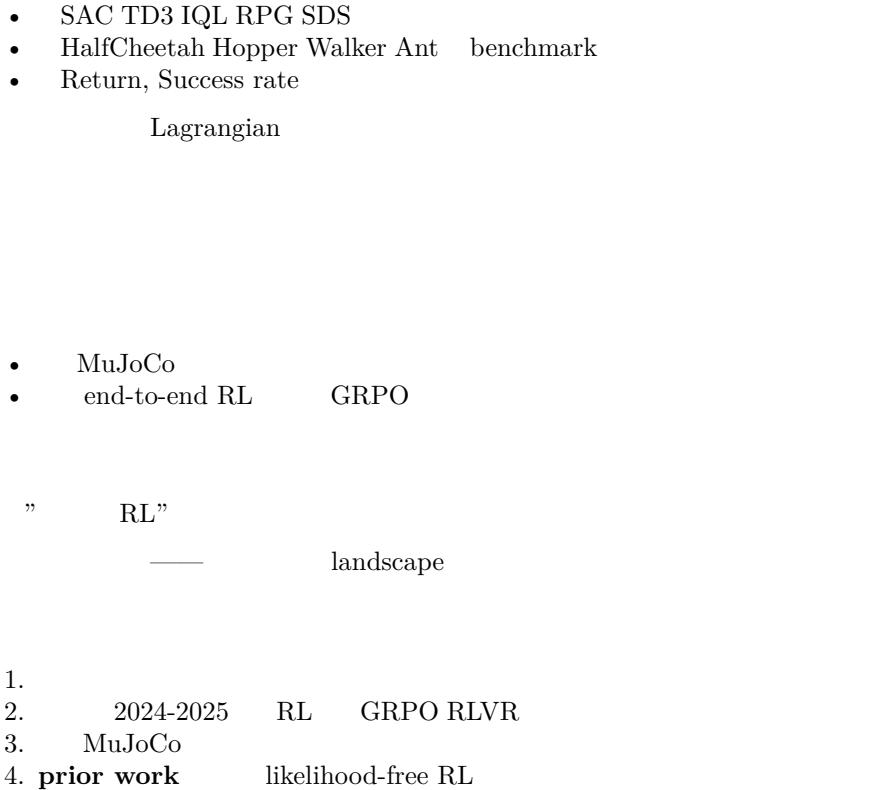
$$\frac{\pi_\theta(a|s)}{\log \pi_\theta(a|s)} / \text{log-density} \quad \text{RL}$$
$$\mathbf{E}_{\tau \sim \pi_\theta}[R(\tau)] + \alpha H(\pi_\theta) \quad H(\pi_\theta) = -\mathbf{E}_{a \sim \pi_\theta(\cdot|s)}[\log \pi_\theta(a|s)]$$

kinetic energy
Schrödinger Bridge (GSB)
" "

1. GSB
- 2.
3. Lagrangian
- 4.

Prior Work

- SAC Soft Actor-Critic
- Diffusion Policy
- Flow Matching RL



2: On Robustness and Chain-of-Thought Consistency of RL-Finetuned VLMs

: []

- : On Robustness and Chain-of-Thought Consistency of RL-Finetuned VLMs
- **ArXiv ID:** 2602.12506
- : Rosie Zhao
- : Apple
- : 2026-02-13

RL LLM VLM RL VLM grounding

caption CoT VLM
VLM

1. RL VLM
2. - (accuracy-faithfulness trade-off)
- 3.
4. +

Prior Work

- RLHF/VLM
- CoT
- RL VLM

- multimodal reasoning models
- caption CoT
- CoT

- RL - - - +

-
- CoT
-
- " "

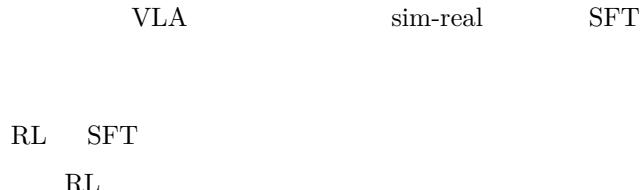
RL VLM
" " " " _____

1. trade-off
 2. ” ”
 3. RL VLM
-

3: RLinf-Co - Reinforcement Learning-Based Sim-Real Co-Training for VLA Models

: []

- : RLinf-Co: Reinforcement Learning-Based Sim-Real Co-Training for VLA Models
- **ArXiv ID:** 2602.12628
- : Liangzhi Shi
- :
- : 2026-02-13



1. RL-based Sim-Real Co-Training (RL-Co)
2. SFT → RL +
- 3.

- 4
- OpenVLA, .
-

- OpenVLA: +24%
- . : +20%
-

Prior Work

- SFT
-
- sim-real

VLA

” ” ——

- 1.
2. RL sim-real
3. 4 2
4. real-to-sim

4: GeoAgent - Learning to Geolocate Everywhere with Reinforced Geographic Characteristics

: []

- : GeoAgent: Learning to Geolocate Everywhere with Reinforced Geographic Characteristics
- **ArXiv ID:** 2602.12617
- : Modi Jin, Yiming Zhang, Boyuan Sun, Dingwen Zhang, MingMing Cheng, Qibin Hou
- : /NKU
- : 2026-02-13

RL AI CoT

1. **GeoSeek:** CoT
2. **geo-similarity reward:**
3. **consistency reward:**

- VLLM
-

- RL +
 - SFT
 - reward hacking
-

5: DICE - Diffusion Large Language Models Excel at Generating CUDA Kernels

: []

- : DICE: Diffusion Large Language Models Excel at Generating CUDA Kernels
- **ArXiv ID:** 2602.11715
- : Haolei Bai, Lingcheng Kong, Xueyi Chen, Jianmian Wang, Zhiqiang Tao, Huan Wang
- :
- : 2026-02-12

dLLM token AR LLM dLLM CUDA kernel

1. **CuKe:** CUDA kernel
2. **BiC-RL:** RL
 - CUDA kernel infilling
 - CUDA kernel
3. **DICE:** 1.7B, 4B, 8B

- KernelBench
- AR Diffusion LLM SOTA

- diffusion
 - CuKe
 - AR CodeGen
 - code generation
-

6: What does RL improve for Visual Reasoning? A Frankenstein-Style Analysis

: []

- : What does RL improve for Visual Reasoning? A Frankenstein-Style Analysis
- ArXiv ID: 2602.12395
- : Xirui Li
- : UMD Tianyi Lab
- : 2026-02-12

RL with verifiable rewards RL

RL RL
RL mid-to-late transformer

1. **Frankenstein-style**
 - via
 - via
 - via
2. RL **mid-to-late layers**
3. mid-to-late

- VLM
-

Prior Work

- RLHF VLM
- RL

RL VLM
 " "

- 1.
- 2.
- 3.

RL VLM

FLAC	[]	RL
RL VLM	[]	RL VLM
RLinf-Co	[]	
GeoAgent	[]	
DICE	[]	Diffusion +
RL	[]	RL

1. Apple VLM - RL+VLM 2. UMD RL - 3.
FLAC -