Novelty Assessment Report

Paper: AdaReasoner: Dynamic Tool Orchestration for Iterative Visual Reasoning
PDF URL: https://openreview.net/pdf?id=nUGPEmQ2ut
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-01

Abstract

While augmenting Multimodal Large Language Models (MLLMs) with tools is a promising direction, current approaches face critical limitations. They often rely on single, atomic tools, failing to address the challenges of multi-turn planning, and they do not equip models with the ability to select effective tool combinations for complex tasks. To overcome these limitations, we introduce AdaReasoner, a framework that teaches models to perform dynamic tool orchestration for iterative visual reasoning. Our paradigm is designed to support a broad spectrum of tools, including computationally intensive, expert-model-based services. It features a comprehensive design that includes a new data curation methodology and a tailored Tool GRPO algorithm to optimize multi-turn tool-calling trajectories, which yields state-of-the-art models that achieve substantial gains over their baselines (+38.7\% average on 7B) and reach near-perfect accuracy on complex benchmarks like Visual Spatial Planning (97.6\%). This performance surpasses leading proprietary systems such as GPT-5 and Claude Sonnet 4, demonstrating that our approach can effectively overcome scale-based limitations by augmenting smaller models with powerful tool-use capabilities. Critically, we find that AdaReasoner develops emergent, self-adaptive behaviors: it learns to autonomously adopt beneficial tools, discard irrelevant ones, and modulate its usage frequency. This ability to curate its own optimal problem-solving strategies represents a significant step toward building more robust, scalable, and reliable reasoning agents.

Core Task Landscape

This paper addresses: **Dynamic Tool Orchestration for Iterative Visual Reasoning**

A total of **50 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Reinforcement Learning-Based Tool Selection and Orchestration**
- **Multi-Agent Collaboration and Orchestration**
- **Supervised and Hybrid Tool Integration**
- **Adaptive Visual Attention and Perception**
- **Domain-Specific Tool-Augmented Reasoning**
- **Workflow Automation and Interface Interaction**
- **Hierarchical and Self-Organizing Agent Architectures**
- **Supporting Infrastructure and Educational Tools**
- **Specialized Reasoning and Optimization Tasks**

Complete Taxonomy Tree

- Dynamic Tool Orchestration for Iterative Visual Reasoning Survey Taxonomy
- Reinforcement Learning-Based Tool Selection and Orchestration
  - End-to-End RL for Visual Tool Use ★ (5 papers)
  - [0] AdaReasoner: Dynamic Tool Orchestration for Iterative Visual Reasoning (Anon et al., 2026) View paper
  - [1] Openthinkimg: Learning to think with images via visual tool reinforcement learning (Su, 2025) View paper
  - [2] Chain-of-Focus: Adaptive Visual Search and Zooming for Multimodal Reasoning via RL (Zhang Xintong, 2025) View paper
  - [8] Visualtoolagent (vista): A reinforcement learning framework for visual tool selection (Huang Zeyi, 2025) View paper
  - [17] VTool-R1: VLMs Learn to Think with Images via Reinforcement Learning on Multimodal Tool Use (Wu Mingyuan, 2025) View paper
  - Self-Evolving and Self-Rewarding Agents (3 papers)
  - [22] Agent0-VL: Exploring Self-Evolving Agent for Tool-Integrated Vision-Language Reasoning (Jiaqi Liu, 2025) View paper
  - [25] AutoTool: Dynamic Tool Selection and Integration for Agentic Reasoning (Jiaru Zou, 2025) View paper
  - [30] Clova: A closed-loop visual assistant with tool usage and update (Zhi Gao, 2024) View paper
  - Adaptive Tool Selection via RL in Specialized Domains (2 papers)
  - [4] Ego-R1: Chain-of-Tool-Thought for Ultra-Long Egocentric Video Reasoning (Tian Shulin, 2025) View paper
  - [49] Tool-Augmented Policy Optimization: Synergizing Reasoning and Adaptive Tool Use with Reinforcement Learning (Wu, 2025) View paper
- Multi-Agent Collaboration and Orchestration
  - Hierarchical Multi-Agent Architectures (3 papers)
  - [5] PixelCraft: A Multi-Agent System for High-Fidelity Visual Reasoning on Structured Images (Zhang Shuoshuo, 2025) View paper
  - [14] EvoAgentX: An Automated Framework for Evolving Agentic Workflows (Yingxu Wang, 2025) View paper
  - [50] ZJUT-MM@ MUCG Challenge: Agent Network for Multimodal Video Understanding (Meiyi Lu, 2025) View paper
  - Collaborative Reasoning with Reflection and Memory (3 papers)
  - [3] Mmctagent: Multi-modal critical thinking agent framework for complex visual reasoning (Kumar, 2024) View paper

    - [20] EndoAgent: A Memory-Guided Reflective Agent for Intelligent Endoscopic Vision-to-Decision Reasoning (Tang Yi, 2025) View paper
    - [47] CXRAgent: Director-Orchestrated Multi-Stage Reasoning for Chest X-Ray Interpretation (Yang Yan, 2025) View paper
    - Multi-Agent Systems for Embodied and Spatial Reasoning (2 papers)
    - [7] CLiViS: Unleashing Cognitive Map through Linguistic-Visual Synergy for Embodied Visual Reasoning (Li Kailing, 2025) View paper
    - [28] Visual agentic ai for spatial reasoning with a dynamic api (Damiano Marsili, 2025) View paper
- Supervised and Hybrid Tool Integration
    - Code-Based Tool Orchestration (3 papers)
    - [11] Simple o3: Towards Interleaved Vision-Language Reasoning (Wang Ye, 2025) View paper
    - [12] Pyvision: Agentic vision with dynamic tooling (Zhao Shitian, 2025) View paper
    - [13] CodeDance: A Dynamic Tool-integrated MLLM for Executable Visual Reasoning (Qi Song, 2025) View paper
    - Prompting-Based and Test-Time Tool Use (3 papers)
    - [19] Recursive Visual Programming (Jiaxin Ge, 2023) View paper
    - [27] DoraemonGPT: Toward Understanding Dynamic Scenes with Large Language Models (Yang, 2024) View paper
    - [29] ToolScope: An Agentic Framework for Vision-Guided and Long-Horizon Tool Use (Deng Mengjie, 2025) View paper
    - Supervised Fine-Tuning for Tool-Augmented Reasoning (4 papers)
    - [37] Orion: A Unified Visual Agent for Multimodal Perception, Advanced Visual Reasoning and Execution (N Dinesh Reddy, 2025) View paper
    - [39] Incentivizing Tool-augmented Thinking with Images for Medical Image Analysis (Yankai Jiang, 2025) View paper
    - [41] ChartAgent: A Chart Understanding Framework with Tool Integrated Reasoning (Boran Wang, 2025) View paper
    - [44] ChartAgent: A Multimodal Agent for Visually Grounded Reasoning in Complex Chart Question Answering (Kaur Rachneet, 2025) View paper
- Adaptive Visual Attention and Perception
    - Iterative Focus Refinement for Visual Grounding (2 papers)
    - [9] GUI-Spotlight: Adaptive Iterative Focus Refinement for Enhanced GUI Visual Grounding (Lei Bin, 2025) View paper
    - Adaptive Sampling and Active Vision (3 papers)
    - [6] Emulating human-like adaptive vision for efficient and flexible machine visual perception (Yulin Wang, 2025) View paper
    - [35] Deep Active Visual Attention for Real-Time Robot Motion Generation: Emergence of Tool-Body Assimilation and Adaptive Tool-Use (Hiruma, 2022) View paper
    - [43] OmniAgent: Audio-Guided Active Perception Agent for Omnimodal Audio-Video Understanding (Keda Tao, 2025) View paper
- Domain-Specific Tool-Augmented Reasoning
    - Medical and Healthcare Applications (2 papers)
    - [16] MedOrch: Medical Diagnosis with Tool-Augmented Reasoning Agents for Flexible Extensibility (Li Ang, 2025) View paper
    - [23] Aura: A multi-modal medical agent for understanding, reasoning and annotation (Nima Fathi, 2025) View paper
    - Agricultural and Environmental Monitoring (1 papers)
    - [32] Dynamic Orchestration of Multi-Agent System for Real-World Multi-Image Agricultural VQA (Ke Yan, 2025) View paper
    - Autonomous Driving and Robotics (2 papers)
    - [42] Orchestrate, Generate, Reflect: A VLM-Based Multi-Agent Collaboration Framework for Automated Driving Policy Learning (Peng Zengqi, 2025) View paper
    - [48] Agentic AI Home Energy Management System: A Large Language Model Framework for Residential Load Scheduling (Zwickl-Bernhard, 2025) View paper
- Workflow Automation and Interface Interaction
    - Web and Browser Automation (1 papers)
    - [21] Recon-Act: A Self-Evolving Multi-Agent Browser-Use System via Web Reconnaissance, Tool Generation, and Task Execution (He, 2025) View paper
    - GUI Interaction and Scene Synthesis (2 papers)
    - [10] Sceneweaver: All-in-one 3d scene synthesis with an extensible and self-reflective agent (Yang, 2025) View paper
    - [26] Blenderalchemy: Editing 3d graphics with vision-language models (Ian Huang, 2024) View paper
- Hierarchical and Self-Organizing Agent Architectures (1 papers)
    - [18] Hiva: Self-organized hierarchical variable agent via goal-driven semantic-topological evolution (Tang Jinzhou, 2025) View paper
- Supporting Infrastructure and Educational Tools (6 papers)
    - [24] IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems (Gazzola, 2019) View paper
    - [31] Exploring students' computational practice, design and performance of problem-solving through a visual programming environment (Poâ□□Yao Chao, 2016) View paper
    - [33] Integration and Orchestration of Analysis Tools (Robert Heinrich, 2021) View paper
    - [34] Adaptive Medical Visualization in AR: From Volumetric Clarity to Context-Aware Interaction (Zou, 2025) View paper
    - [45] Educational Robotics and Agile Learning: A Review (Antonios Konstantaras, 2025) View paper
    - [46] Rethinking UX Education (Ryan G Wilson, 2025) View paper
- Specialized Reasoning and Optimization Tasks (3 papers)
    - [15] MuaLLM: A Multimodal Large Language Model Agent for Circuit Design Assistance with Hybrid Contextual Retrieval-Augmented Generation (Pravallika Abbineni, 2025) View paper
    - [36] Agentic AI with Orchestrator-Agent Trust: A Modular Visual Classification Framework with Trust-Aware Orchestration and RAG-Based Reasoning (Sapkota, 2025) View paper
    - [38] Fusing lane-level flow inference and multi-step RL for adaptive traffic signal coordination (Haoran Cheng, 2025) View paper

Narrative

Core task: dynamic tool orchestration for iterative visual reasoning. The field addresses how agents can adaptively select and coordinate external tools—ranging from vision modules and code interpreters to domain-specific APIs—to solve complex visual tasks that require multiple reasoning steps. The taxonomy reveals several major branches: Reinforcement Learning-Based Tool Selection and Orchestration explores end-to-end RL methods that learn which tools to invoke and when; Multi-Agent Collaboration and Orchestration examines systems where multiple specialized agents coordinate their capabilities; Supervised and Hybrid Tool Integration focuses on training regimes that combine demonstration data with learned policies; Adaptive Visual Attention and Perception investigates how agents dynamically adjust their perceptual focus; Domain-Specific Tool-Augmented Reasoning targets applications in medicine, agriculture, and other specialized fields; Workflow Automation and Interface Interaction deals with GUI agents and process automation; Hierarchical and

Self-Organizing Agent Architectures studies modular designs that decompose tasks; Supporting Infrastructure and Educational Tools provides benchmarks and teaching frameworks; and Specialized Reasoning and Optimization Tasks covers niche problem settings. Representative works such as MMCTAgent[3] and VisualToolAgent[8] illustrate how tool libraries can be integrated into reasoning pipelines, while approaches like Ego-R1[4] and PixelCraft[5] demonstrate diverse strategies for managing iterative perception and action.

A particularly active line of work centers on end-to-end RL for visual tool use, where agents learn orchestration policies directly from task rewards rather than relying solely on supervised demonstrations. AdaReasoner[0] exemplifies this direction by training an RL-based controller that iteratively selects tools to refine visual understanding, closely aligning with OpenThinking[1] and Chain-of-Focus[2], which similarly emphasize learned decision-making over fixed pipelines. In contrast, VTool-R1[17] and VisualToolAgent[8] blend RL with more structured reasoning traces, highlighting a trade-off between flexibility and interpretability. AdaReasoner[0] distinguishes itself by focusing on adaptive iteration—dynamically deciding when to invoke perception modules versus reasoning steps—whereas Chain-of-Focus[2] prioritizes attention mechanisms and OpenThinking[1] explores transparent reasoning chains. These differences reflect broader questions in the field: how much structure should be imposed on tool selection, whether to optimize end-to-end or modularize components, and how to balance sample efficiency with generalization across diverse visual tasks.

Related Works in Same Category

The following **4 sibling papers** share the same taxonomy leaf node with the original paper:

1. Openthinkimg: Learning to think with images via visual tool reinforcement learning

Authors: Su, Zhaochen, Li, Linjie, Zhao-yu Su, et al. (26 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

While humans can flexibly leverage interactive visual cognition for complex problem-solving, enabling Large Vision-Language Models (LVLMs) to learn similarly adaptive behaviors with visual tools remains challenging. A significant hurdle is the current lack of standardized infrastructure, which hinders integrating diverse tools, generating rich interaction data, and training robust agents effectively. To address these gaps, we introduce OpenThinkIMG, the first open-source, comprehensive end-to-en...

Relationship Analysis

Both papers belong to the End-to-End RL for Visual Tool Use category, training vision-language models with reinforcement learning to orchestrate external tools for visual reasoning tasks. They overlap in using RL (GRPO-based methods) to optimize multi-turn tool invocation policies and both demonstrate significant performance gains over supervised baselines on visual reasoning benchmarks. The key difference is that AdaReasoner focuses on dynamic tool orchestration with adaptive tool selection behaviors across diverse tasks (VSP, Jigsaw, GUIQA) using a custom Tool GRPO algorithm, while OpenThinkIMG introduces V-ToolRL as a comprehensive open-source framework with standardized tool interfaces, primarily validated on chart reasoning tasks.

---

2. Chain-of-Focus: Adaptive Visual Search and Zooming for Multimodal Reasoning via RL

Authors: Zhang Xintong, Gao Zhi, Xintong Zhang, Zhang Bo-fei, Zhi Gao, et al. (24 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Vision language models (VLMs) have achieved impressive performance across a variety of computer vision tasks. However, the multimodal reasoning capability has not been fully explored in existing models. In this paper, we propose a Chain-of-Focus (CoF) method that allows VLMs to perform adaptive focusing and zooming in on key image regions based on obtained visual cues and the given questions, achieving efficient multimodal reasoning. To enable this CoF capability, we present a two-stage training...

Relationship Analysis

Both papers belong to the End-to-End RL for Visual Tool Use category, training vision-language models with reinforcement learning to orchestrate tools for iterative visual reasoning. They overlap in using two-stage pipelines (SFT followed by RL) to train models on adaptive tool selection for complex visual tasks, with both employing outcome-based rewards to refine reasoning strategies. The key difference is that AdaReasoner focuses on multi-turn tool orchestration across diverse tool types (perception, manipulation, calculation) with explicit reflection and backtracking mechanisms, while Chain-of-Focus specifically targets adaptive visual search and zooming operations for efficient region-based reasoning across different image resolutions.

---

3. Visualtoolagent (vista): A reinforcement learning framework for visual tool selection

Authors: Huang Zeyi, Ji, Yuyang, Zeyi Huang, Yuyang Ji, et al. (15 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

We introduce VisTA, a new reinforcement learning framework that empowers visual agents to dynamically explore, select, and combine tools from a diverse library based on empirical performance. Existing methods for tool-augmented reasoning either rely on training-free prompting or large-scale fine-tuning; both lack active tool exploration and typically assume limited tool diversity, and fine-tuning methods additionally demand extensive human supervision. In contrast, VisTA leverages end-to-end rei...

Relationship Analysis

Both papers belong to the End-to-End RL for Visual Tool Use category, employing GRPO to train vision-language models for dynamic tool orchestration across multi-turn reasoning tasks. They overlap in using reinforcement learning to optimize tool selection strategies without extensive human supervision, and both demonstrate adaptive tool-use behaviors on visual reasoning benchmarks. However, AdaReasoner focuses on multi-turn trajectory optimization with a comprehensive data curation pipeline including reflection and backtracking mechanisms, while VisTA emphasizes autonomous tool exploration and selection from diverse libraries using task outcomes as feedback signals, with particular attention to out-of-distribution generalization.

---

4. VTool-R1: VLMs Learn to Think with Images via Reinforcement Learning on Multimodal Tool Use

Authors: Wu Mingyuan, Yang JingCheng, Mingyuan Wu, Jiang, Jize, et al. (23 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Reinforcement Learning Finetuning (RFT) has significantly advanced the reasoning capabilities of large language models (LLMs) by enabling long chains of thought, self-correction, and effective tool use. While recent works attempt to extend RFT to vision-language models (VLMs), these efforts largely produce text-only reasoning conditioned on static image inputs, falling short of true multimodal reasoning in the response. In contrast, test-time methods like Visual Sketchpad incorporate visual step...

Relationship Analysis

Both papers belong to the End-to-End RL for Visual Tool Use category, training vision-language models with reinforcement learning to orchestrate tools for visual reasoning tasks. They overlap in using RL (GRPO-based methods) to optimize tool selection and multi-turn reasoning trajectories for complex visual question answering. However, AdaReasoner focuses on dynamic tool orchestration with diverse tool types (perception, manipulation, calculation) and adaptive tool selection behaviors, while VTool-R1 specifically emphasizes generating multimodal chains of thought by interleaving text and intermediate visual reasoning steps using Python-based visual editing tools for structured visual QA over charts and tables.

Contributions Analysis

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: AdaReasoner framework for dynamic tool orchestration

Description: The authors propose a comprehensive framework that enables multimodal large language models to dynamically select and combine tools for complex visual reasoning tasks. The framework includes a data curation methodology for multi-turn tool planning and a tailored Tool GRPO algorithm to optimize multi-turn tool-calling trajectories.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Towards robust multi-modal reasoning via model selection

URL: View paper

Brief Assessment

Robust Multi-modal Reasoning[55] focuses on model selection among predefined task-specific models for multi-step reasoning robustness, not on teaching models to dynamically orchestrate and combine tools through reinforcement learning as in the original paper.

---

2. Deep research agents: A systematic examination and roadmap

URL: View paper

Brief Assessment

Deep Research Agents[52] focuses on autonomous research agents for multi-turn informational tasks with browser-based exploration and report generation, not on multimodal visual reasoning with dynamic tool selection for vision-language models.

---

3. Beyond seeing: Evaluating multimodal llms on tool-enabled image perception, transformation, and reasoning

URL: View paper

Brief Assessment

Tool-enabled Perception[54] focuses on evaluating MLLMs' ability to use general-purpose tools for image manipulation and reasoning across diverse domains, rather than proposing a training framework for dynamic tool orchestration with multi-turn planning and reinforcement learning optimization.

---

4. SILMM: Self-Improving Large Multimodal Models for Compositional Text-to-Image Generation

URL: View paper

Brief Assessment

SILMM[56] focuses on self-improvement for text-to-image generation through direct preference optimization, not on dynamic tool orchestration for visual reasoning tasks. The technical approaches and problem domains are fundamentally different.

---

5. PixelCraft: A Multi-Agent System for High-Fidelity Visual Reasoning on Structured Images

URL: View paper

Brief Assessment

PixelCraft[5] focuses on multi-agent systems with high-fidelity image processing for structured images (charts, geometry), while the original paper addresses general visual reasoning with dynamic tool selection across diverse tasks. The architectural approaches differ fundamentally: PixelCraft uses a multi-agent architecture (dispatcher, planner, reasoner, critics) with image memory for adaptive reasoning, whereas the original emphasizes a single-model framework with GRPO-based optimization for multi-turn tool trajectories.

---

6. Chain-of-Focus: Adaptive Visual Search and Zooming for Multimodal Reasoning via RL

URL: View paper

Brief Assessment

Chain-of-Focus[2] focuses on adaptive visual search and zooming within images for multimodal reasoning, not on dynamic tool orchestration across multiple external tools. The candidate's approach involves focusing on image regions rather than selecting and combining diverse external tools for complex reasoning tasks.

---

7. Openthinkimg: Learning to think with images via visual tool reinforcement learning

URL: View paper

Brief Assessment

OpenThinking[1] focuses on chart reasoning tasks with a different RL approach (V-ToolRL), while the original paper addresses multi-turn tool planning across diverse visual reasoning tasks (VSP, Jigsaw, GUI-QA) with Tool GRPO and adaptive tool selection behaviors.

---

8. Benchmarking Multimodal Retrieval Augmented Generation with Dynamic VQA Dataset and Self-adaptive Planning Agent

URL: View paper

Brief Assessment

Dynamic VQA[53] focuses on adaptive retrieval planning for knowledge-seeking VQA tasks, not on general visual reasoning with tool orchestration. The candidate addresses retrieval strategies for answering questions, while the original paper develops a framework for multi-turn tool planning across diverse visual reasoning tasks.

---

9. Mmctagent: Multi-modal critical thinking agent framework for complex visual reasoning

URL: View paper

Brief Assessment

MMCTAgent[3] focuses on critical thinking and self-reflection mechanisms for visual reasoning, not on dynamic tool orchestration or multi-turn tool planning trajectories as in the original paper.

---

10. ReAgent-V: A Reward-Driven Multi-Agent Framework for Video Understanding

URL: View paper

Brief Assessment

ReAgent-V[51] focuses on video understanding with frame selection and reward-driven refinement for video-specific tasks, while the original paper addresses general multimodal visual reasoning with dynamic tool orchestration across diverse image-based tasks like spatial planning and jigsaw puzzles.

Contribution 2: Data curation methodology for multi-turn tool planning

Description: The authors introduce a three-stage data curation process that generates high-quality, human-like reasoning trajectories. This methodology deliberately incorporates reflection and backtracking scenarios, as well as explicit tool failure cases, to teach models robust problem-solving strategies beyond simply following optimal paths.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. GenoTEX: An LLM Agent Benchmark for Automated Gene Expression Data Analysis

URL: View paper

Brief Assessment

GenoTEX[62] focuses on gene expression data analysis with programming workflows, not multi-turn tool planning with reflection and backtracking for visual reasoning tasks.

2. Collecting metrics for continuous platform monitoring

URL: View paper

Brief Assessment

Platform Monitoring Metrics[57] focuses on collecting numeric measurements for platform stability monitoring and alerting systems, not on generating reasoning trajectories with reflection and backtracking for tool-augmented language models.

3. Systematic review of metadata-driven data orchestration in modern analytics engineering

URL: View paper

Brief Assessment

Metadata-driven Orchestration[63] focuses on metadata management for data pipeline orchestration in analytics engineering, not on data curation for multi-turn tool planning with reflection and backtracking in AI agent systems.

4. A survey of reasoning and agentic systems in time series with large language models

URL: View paper

Brief Assessment

Time Series Reasoning[60] is a survey paper focused on time series analysis with LLMs, not on data curation methodologies for multi-turn tool planning in visual reasoning tasks. The domains and objectives are fundamentally different.

5. Domain-Oriented Time Series Inference Agents for Reasoning and Automated Analysis

URL: View paper

Brief Assessment

Time Series Inference[64] focuses on time series analysis workflows with statistical and domain-specific operators, not multi-turn visual reasoning tool planning with reflection and backtracking scenarios as in the original paper.

6. Towards Standardization of GenAI-Driven Agentic Architectures for Radio Access Networks

URL: View paper

Brief Assessment

GenAI Radio Networks[58] focuses on standardizing agentic architectures for radio access networks in telecommunications, not on data curation methodologies for training multimodal models with tool-use capabilities. The domains and technical objectives are fundamentally different.

7. Reflection-Driven Control for Trustworthy Code Agents

URL: View paper

Brief Assessment

Reflection-Driven Control[66] focuses on self-supervision and self-correction in code agents, not on multi-turn tool planning with visual reasoning tasks. The candidate's approach is domain-specific to code generation rather than the original paper's visual reasoning framework with explicit reflection/backtracking scenarios and tool failure cases.

8. Domaino1s: Guiding llm reasoning for explainable answers in high-stakes domains

URL: View paper

Brief Assessment

DomainO1s[59] focuses on formal reasoning in high-stakes domains with self-correction mechanisms, not on multi-turn tool planning with reflection and backtracking for visual reasoning tasks.

9. Flexible and Reproducible RF Calibration using Google Cloud Workflows

URL: View paper

Brief Assessment

RF Calibration Workflows[65] focuses on automating radio frequency measurement device calibration using workflow orchestration systems. It does not address data curation for multi-turn tool planning with reflection and backtracking in multimodal reasoning contexts.

10. Generator-assistant stepwise rollback framework for large language model agent

URL: View paper

Brief Assessment

Generator-assistant Rollback[61] focuses on a rollback framework for error correction during execution, not on data curation methodologies for generating training trajectories with reflection and backtracking scenarios. The candidate's limited context does not address data generation processes.

Contribution 3: Tool GRPO algorithm for multi-turn tool interaction

Description: The authors develop an adaptive reinforcement learning paradigm that extends the GRPO framework to handle multi-turn tool-calling scenarios. This includes multi-turn reward accumulation and an adaptive reward mechanism with asymmetric incentive structure to guide models in learning when and how to use tools effectively.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Demystifying reinforcement learning in agentic reasoning

URL: View paper

Brief Assessment

Agentic Reasoning RL[68] focuses on general agentic reasoning with tool use across mathematical and coding tasks, while the original paper specifically develops an adaptive GRPO framework for multi-turn visual tool-calling trajectories with asymmetric reward structures for multimodal reasoning tasks.

---

2. ReTool: Reinforcement Learning for Strategic Tool Use in LLMs

URL: View paper

Brief Assessment

ReTool[67] focuses on code interpreter integration for mathematical reasoning, while the original paper addresses visual tool orchestration with perception/manipulation tools. The technical domains and tool types are fundamentally different.

---

3. Agentic reinforced policy optimization

URL: View paper

Brief Assessment

Agentic Policy Optimization[72] focuses on entropy-based adaptive rollout mechanisms for balancing reasoning and tool interactions, rather than the multi-turn reward accumulation and asymmetric incentive structure described in the original paper's Tool GRPO.

---

4. Steptool: A step-grained reinforcement learning framework for tool learning in llms

URL: View paper

Brief Assessment

StepTool[73] focuses on step-grained rewards for individual tool invocations in sequential tasks, not on multi-turn tool-calling trajectories with adaptive rewards for tool orchestration as in the original paper.

---

5. Reinforcing multi-turn reasoning in llm agents via turn-level reward design

URL: View paper

Prior Art Analysis

Multi-turn Reasoning[69] demonstrates prior work on extending GRPO to multi-turn scenarios with turn-level reward design. The candidate paper explicitly presents 'multi-turn grpo (mt-grpo)' that extends GRPO for multi-turn agentic tasks with turn-level advantages and intermediate rewards. Both papers address the same core challenge: adapting GRPO from trajectory-level to turn-level credit assignment in multi-turn tool-calling scenarios. The candidate's formulation of turn-level advantages (amt-gpro_i,1 = ai_i + αao_i, amt-gpro_i,2 = ao_i) and its focus on 'fine-grained credit assignment' directly parallels the original paper's 'adaptive reward mechanism with asymmetric incentive structure' for multi-turn tool trajectories.

Evidence

Evidence 1 - **Rationale**: Both papers describe extending GRPO for multi-turn scenarios with reward accumulation mechanisms. The candidate's work on 'multi-turn variants' with 'intermediate rewards' directly addresses the same problem space as the original's 'multi-turn reward accumulation'. - **Original**: To train our model for complex multi-turn tool-planning scenarios, we extend the grpo framework to effectively handle multi-turn tool-calling reasoning trajectories. concretely, we use multi-turn reward accumulation and adaptive tool reward to ensure the efficacy of the rl procedure. - **Candidate**: we investigate turn-level reward design for both multi-turn rl algorithms and agent applications... To train multi-turn llm agents effectively under our mdp formulation, we propose to extend grpo and ppo to their multi-turn variants by incorporating both outcome and intermediate rewards, enabling fi...

---

6. Tool-Augmented Policy Optimization: Synergizing Reasoning and Adaptive Tool Use with Reinforcement Learning

URL: View paper

Brief Assessment

Tool-Augmented Policy[49] focuses on integrating tool-calling with reasoning using a modified DAPO framework for search APIs and Python interpreters, rather than the multi-turn visual tool orchestration and adaptive reward mechanisms described in the original paper's Tool GRPO.

---

7. Fathom-deepresearch: Unlocking long horizon information retrieval and synthesis for slms

URL: View paper

Brief Assessment

Fathom-DeepResearch[71] focuses on web search and information retrieval tasks using RAPO (an extension of GRPO), not on visual reasoning or multi-turn tool orchestration for multimodal tasks like the original paper's Tool GRPO for visual tools.

---

8. ToRL: Scaling Tool-Integrated RL

URL: View paper

Brief Assessment

ToRL[75] focuses on tool-integrated reinforcement learning for mathematical reasoning with code execution, while the original paper develops a multi-turn tool GRPO framework specifically for visual reasoning tasks with diverse perception and manipulation tools.

---

9. Reinforcement learning foundations for deep research systems: A survey

URL: View paper

Brief Assessment

RL Deep Research[70] is a survey paper that systematizes RL methods for agentic research systems broadly. It does not present a specific GRPO extension for multi-turn tool-calling with adaptive rewards as described in the original paper's contribution.

---

10. Reinforcing multi-turn reasoning in llm agents via turn-level credit assignment

URL: View paper

Prior Art Analysis

Turn-level Credit[74] demonstrates that prior work exists on extending GRPO for multi-turn tool-calling scenarios with adaptive reward mechanisms. Both papers develop GRPO-based frameworks for multi-turn tool interaction with reward accumulation across multiple turns. The candidate paper explicitly addresses multi-turn tool-calling with turn-level rewards and adaptive advantage estimation, which directly overlaps with the original paper's claimed novelty of extending GRPO to handle multi-turn tool-calling scenarios with multi-turn reward accumulation and adaptive reward mechanisms.

Evidence

Evidence 1 - **Rationale**: Both papers implement adaptive reward mechanisms that adjust based on tool usage and answer correctness, demonstrating prior work on asymmetric incentive structures for tool learning. - **Original**: adaptive reward for encouraging tool use. to guide the model to use tools as a reliable aid when uncertain, we introduce an adaptive reward mechanism with an asymmetric incentive structure, where the reward calculation is contingent on the final answer's correctness. correct trajectories automaticall... - **Candidate**: turn-level verifiable rewards:these depend solely on the first turn performed by the llm agent. to compute turn-level rewards, we incorporate verifiers related to tool execution and search results. these verifiers ensure that the search engine is correctly invoked and that the ground-truth answer ap...

Evidence 2 - **Rationale**: Both papers explicitly extend GRPO for multi-turn tool-calling with turn-level advantage functions and reward accumulation, showing this approach existed prior to the original paper. - **Original**: To train our model for complex multi-turn tool-planning scenarios, we extend the grpo framework to effectively handle multi-turn tool-calling reasoning trajectories. concretely, we use multi-turn reward accumulation and adaptive tool reward to ensure the efficacy of the rl procedure. - **Candidate**: in this work, we treat each interaction between the llm agent and the environment as a turn within the mdp framework. this perspective enables us to design a turn-level advantage function that effectively captures the contribution of each turn within a trajectory.

Appendix: Text Similarity Detection

Textual similarity detection checked 32 papers and found 1 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Openthinkimg: Learning to think with images via visual tool reinforcement learning

Detected in: Core Task (sibling), Contribution: contribution_1

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] AdaReasoner: Dynamic Tool Orchestration for Iterative Visual Reasoning View paper
- [1] Openthinkimg: Learning to think with images via visual tool reinforcement learning View paper
- [2] Chain-of-Focus: Adaptive Visual Search and Zooming for Multimodal Reasoning via RL View paper
- [3] Mmctagent: Multi-modal critical thinking agent framework for complex visual reasoning View paper
- [4] Ego-R1: Chain-of-Tool-Thought for Ultra-Long Egocentric Video Reasoning View paper
- [5] PixelCraft: A Multi-Agent System for High-Fidelity Visual Reasoning on Structured Images View paper
- [6] Emulating human-like adaptive vision for efficient and flexible machine visual perception View paper
- [7] CLiViS: Unleashing Cognitive Map through Linguistic-Visual Synergy for Embodied Visual Reasoning View paper
- [8] Visualtoolagent (vista): A reinforcement learning framework for visual tool selection View paper
- [9] GUI-Spotlight: Adaptive Iterative Focus Refinement for Enhanced GUI Visual Grounding View paper
- [10] Sceneweaver: All-in-one 3d scene synthesis with an extensible and self-reflective agent View paper
- [11] Simple o3: Towards Interleaved Vision-Language Reasoning View paper
- [12] Pyvision: Agentic vision with dynamic tooling View paper
- [13] CodeDance: A Dynamic Tool-integrated MLLM for Executable Visual Reasoning View paper
- [14] EvoAgentX: An Automated Framework for Evolving Agentic Workflows View paper
- [15] MuaLLM: A Multimodal Large Language Model Agent for Circuit Design Assistance with Hybrid Contextual Retrieval-Augmented Generation View paper
- [16] MedOrch: Medical Diagnosis with Tool-Augmented Reasoning Agents for Flexible Extensibility View paper
- [17] VTool-R1: VLMs Learn to Think with Images via Reinforcement Learning on Multimodal Tool Use View paper
- [18] Hiva: Self-organized hierarchical variable agent via goal-driven semantic-topological evolution View paper
- [19] Recursive Visual Programming View paper
- [20] EndoAgent: A Memory-Guided Reflective Agent for Intelligent Endoscopic Vision-to-Decision Reasoning View paper
- [21] Recon-Act: A Self-Evolving Multi-Agent Browser-Use System via Web Reconnaissance, Tool Generation, and Task Execution View paper
- [22] Agent0-VL: Exploring Self-Evolving Agent for Tool-Integrated Vision-Language Reasoning View paper
- [23] Aura: A multi-modal medical agent for understanding, reasoning and annotation View paper
- [24] IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems View paper
- [25] AutoTool: Dynamic Tool Selection and Integration for Agentic Reasoning View paper
- [26] Blenderalchemy: Editing 3d graphics with vision-language models View paper
- [27] DoraemonGPT: Toward Understanding Dynamic Scenes with Large Language Models View paper
- [28] Visual agentic ai for spatial reasoning with a dynamic api View paper
- [29] ToolScope: An Agentic Framework for Vision-Guided and Long-Horizon Tool Use View paper
- [30] Clova: A closed-loop visual assistant with tool usage and update View paper
- [31] Exploring students' computational practice, design and performance of problem-solving through a visual programming environment View paper
- [32] Dynamic Orchestration of Multi-Agent System for Real-World Multi-Image Agricultural VQA View paper
- [33] Integration and Orchestration of Analysis Tools View paper
- [34] Adaptive Medical Visualization in AR: From Volumetric Clarity to Context-Aware Interaction View paper
- [35] Deep Active Visual Attention for Real-Time Robot Motion Generation: Emergence of Tool-Body Assimilation and Adaptive Tool-Use View paper
- [36] Agentic AI with Orchestrator-Agent Trust: A Modular Visual Classification Framework with Trust-Aware Orchestration and RAG-Based Reasoning View paper
- [37] Orion: A Unified Visual Agent for Multimodal Perception, Advanced Visual Reasoning and Execution View paper
- [38] Fusing lane-level flow inference and multi-step RL for adaptive traffic signal coordination View paper
- [39] Incentivizing Tool-augmented Thinking with Images for Medical Image Analysis View paper
- [40] \textsc {GUI-Spotlight}: Adaptive Iterative Focus Refinement for Enhanced GUI Visual Grounding View paper
- [41] ChartAgent: A Chart Understanding Framework with Tool Integrated Reasoning View paper

- [42] Orchestrate, Generate, Reflect: A VLM-Based Multi-Agent Collaboration Framework for Automated Driving Policy Learning View paper
- [43] OmniAgent: Audio-Guided Active Perception Agent for Omnimodal Audio-Video Understanding View paper
- [44] ChartAgent: A Multimodal Agent for Visually Grounded Reasoning in Complex Chart Question Answering View paper
- [45] Educational Robotics and Agile Learning: A Review View paper
- [46] Rethinking UX Education View paper
- [47] CXRAgent: Director-Orchestrated Multi-Stage Reasoning for Chest X-Ray Interpretation View paper
- [48] Agentic AI Home Energy Management System: A Large Language Model Framework for Residential Load Scheduling View paper
- [49] Tool-Augmented Policy Optimization: Synergizing Reasoning and Adaptive Tool Use with Reinforcement Learning View paper
- [50] ZJUT-MM@ MUCG Challenge: Agent Network for Multimodal Video Understanding View paper
- [51] ReAgent-V: A Reward-Driven Multi-Agent Framework for Video Understanding View paper
- [52] Deep research agents: A systematic examination and roadmap View paper
- [53] Benchmarking Multimodal Retrieval Augmented Generation with Dynamic VQA Dataset and Self-adaptive Planning Agent View paper
- [54] Beyond seeing: Evaluating multimodal llms on tool-enabled image perception, transformation, and reasoning View paper
- [55] Towards robust multi-modal reasoning via model selection View paper
- [56] SILMM: Self-Improving Large Multimodal Models for Compositional Text-to-Image Generation View paper
- [57] Collecting metrics for continuous platform monitoring View paper
- [58] Towards Standardization of GenAI-Driven Agentic Architectures for Radio Access Networks View paper
- [59] Domaino1s: Guiding llm reasoning for explainable answers in high-stakes domains View paper
- [60] A survey of reasoning and agentic systems in time series with large language models View paper
- [61] Generator-assistant stepwise rollback framework for large language model agent View paper
- [62] GenoTEX: An LLM Agent Benchmark for Automated Gene Expression Data Analysis View paper
- [63] Systematic review of metadata-driven data orchestration in modern analytics engineering View paper
- [64] Domain-Oriented Time Series Inference Agents for Reasoning and Automated Analysis View paper
- [65] Flexible and Reproducible RF Calibration using Google Cloud Workflows View paper
- [66] Reflection-Driven Control for Trustworthy Code Agents View paper
- [67] ReTool: Reinforcement Learning for Strategic Tool Use in LLMs View paper
- [68] Demystifying reinforcement learning in agentic reasoning View paper
- [69] Reinforcing multi-turn reasoning in llm agents via turn-level reward design View paper
- [70] Reinforcement learning foundations for deep research systems: A survey View paper
- [71] Fathom-deepresearch: Unlocking long horizon information retrieval and synthesis for slms View paper
- [72] Agentic reinforced policy optimization View paper
- [73] Steptool: A step-grained reinforcement learning framework for tool learning in llms View paper
- [74] Reinforcing multi-turn reasoning in llm agents via turn-level credit assignment View paper
- [75] ToRL: Scaling Tool-Integrated RL View paper