arXiv:2602.01202v1 [cs.AI] 1 Feb 2026

Workflow-R1：用于多轮工作流构建的组子序列策略优化

Mingze Kong [1] **Zikun Qu** [1] **Zhongquan Zhou** [2] **Pengyu Liang** [3] **Xiang Li** [3] **Zhiwei Shang** [1] **Zhi Hong** [1] **Kaiyu Huang** [4] **Zhiyong Wang** [5] **Zhongxiang Dai** [1]

摘要

智能体工作流的快速发展展现了基于大语言模型的智能体在解决复杂推理任务方面的强大性能。然而，现有的工作流优化方法通常将工作流合成定义为静态的一次性代码中心生成问题。这一范式对模型的编码能力施加了过多限制，并限制了解决动态问题所需的灵活性。本文提出 Workflow-R1，这一框架将工作流构建重新定义为一个基于自然语言的多轮序贯决策过程。为解决此类多轮交互中固有的优化粒度不匹配问题，我们提出组子序列策略优化（GSsPO）。GSsPO 虽专为适配智能体推理中交错进行的思考-行动动态而设计，但从根本上说，它是一种具备结构感知能力的强化学习算法，可推广至广泛的多轮智能体序贯决策任务。通过将优化单元重新校准为复合子序列，具体而言即原子思考-行动周期，该方法使梯度更新与这些交互的语义边界相契合，从而确保在复杂的多轮推理任务中实现稳健学习。通过在多个问答基准上开展大量实验，Workflow-R1 的表现优于具有竞争力的基线方法，验证了 GSsPO 是一种通用的序贯推理解决方案，并确立了 Workflow-R1 作为自动化工作流优化的一种有前景的新范式。

[1]香港中文大学（深圳） [2]上海交通大学 [3]天津大学 [4]同济大学[5]爱丁堡大学。通信作者：Zhongxiang Dai< daizhongxiang@cuhk.edu.cn>。

预印本。*2026*年*2*月*3*日。

1. 引言

大语言模型（LLM）早已远远超越简单的对话界面。通过与外部工具和结构化推理路径相结合，它们能够充当自主智能体，解决复杂问题（Hong 等，2024年；Wu 等，2024年）。业界的关注点日益转向基于工作流的智能体，其被定义为能够将复杂查询分解为由搜索、答案、审查和修订等算子构成的可执行图的系统。这一范式已成为处理复杂任务的稳健标准。

认识到这一潜力后，越来越多的研究开始探索自动化工作流优化，旨在用能够自主发现最优交互模式的算法取代手动提示工程。许多近期方法会生成完整的工作流程序（通常是可执行代码），在任何算子执行并观察到其中间结果之前，通过单次处理来编排算子（Zhang 等，2025d；Hu 等，2024年；Nie 等，2025年；Zhang 等，2025f；Wang 等，2025a；Zhang 等，2025a；Gao 等，2025年；Zhang 等，2025b；Ye 等，2025年；Wang 等，2025b；Su 等，2025年；Zhou 等，2025年）。尽管这一范式对于定义明确的任务十分有效，但它存在一个根本性限制，我们将其称为静态执行陷阱。

在这些框架中，计算图在执行前便已完全确定，从而有效地将规划阶段与运行时执行解耦。这形成了一个开环系统：智能体在无法获取过程中产生的中间观测或算子执行结果的情况下，便提交了完整的算子序列。因此，工作流缺乏应对动态环境的适应能力，因为无论观测如何变化，控制流都保持僵化。更为严谨的方法是不将工作流构建视为静态程序合成，而是将其视为一个序贯决策过程，在此过程中，策略会持续根据观测轨迹进行调整。

为克服这种固有的僵化，我们从迭代推理领域的最新进展中汲取灵感（Li 等人，2025年；Zhang

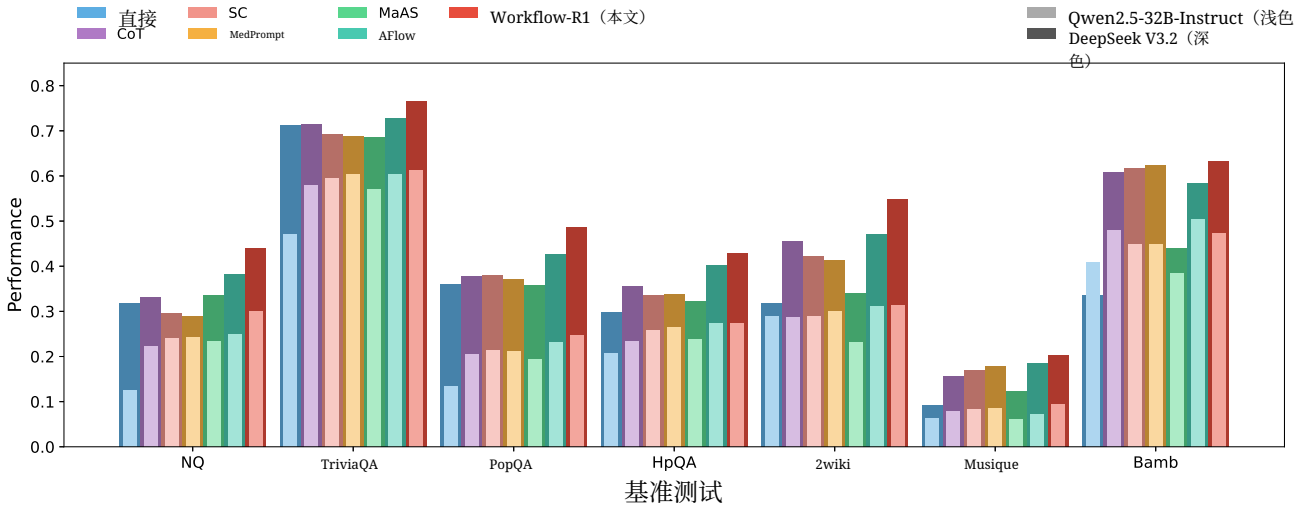


图 *1.* Workflow-R1 与 SOTA基线在七个基准测试上的性能比较。较浅和较深的色调分别表示使用 Qwen2.5-32B-Instruct 和 DeepSeek V3.2 作为骨干模型时的结果。与标准提示策略和先进的工作流优化方法相比，Workflow-R1 在所评估的基准测试中均展现出更优的性能。

等人，2025c；Jin 等人，2025年；Feng 等人，2025年；Zhang 等人，2025e），并提出了 `Workflow-R1` 。这一新颖框架推动范式从静态规划转向**动态交互**。具体而言，我们将工作流优化重新定义为智能体与环境之间的多轮对话。智能体不再生成单体脚本，而是通过交错进行的思考、行动与观察循环与环境互动。这种闭环设计确保每个决策都基于前序步骤的算子执行结果，使模型能够实时动态调整推理路径和算子选择。

此外，在 Workflow-R1 中，这种交互是通过自然语言而非可执行代码进行的。这一设计带来了关键的结构性优势：由于智能体不再受代码生成中固有的语法精确性和变量管理等严格约束，优化任务变得容易处理得多。这种抽象简化了决策过程，使模型能够专注于逻辑推理和算子协同。至关重要的是，这种简化赋予了模型冷启动能力：智能体无需专家代码示范或监督微调（SFT）预热，即可通过强化学习（RL）高效地从零开始学习优化工作流。

然而，将这种动态的多轮决策过程与强化学习对齐面临着一个独特的挑战：标准优化目标与智能体推理的内在语义结构之间存在**粒度不匹配** 。

当前方法所采用的粒度与问题求解的序列性质不匹配，陷入了两个有问题的极端。一端是词元-级方法，例如 GRPO（Shao 等人，2024年），在单个词元层面进行优化。这种方法实际上将复杂推理链视为彼此松散独立的词元集合。通过针对孤立词元而非连贯逻辑单元进行更新，它可能破坏思考-行动过程的语义完整性，同时忽视了将某个推理思维与后续行动连接起来的强因果依赖。另一端是序列级方法，例如 GSPO（Zheng 等人，2025年），将整个多轮交互序列视为一个单一整体。虽然这使优化与最终奖励信号保持一致，却掩盖了多轮工作流中固有的不同决策步骤。在长时域任务中，依赖一个针对整个序列的单一粗粒度信号，无法区分有效的推理步骤与错误步骤，从而导致学习效率低下和策略改进缓慢。

我们认为，最优对齐目标既不是单个词元，也不是整体序列，而是子序列，即从完整响应中解析出的连续片段所对应的决策单元。在我们的多轮设定中，每一轮都会产生一个思考-行动对，按构造即形成一个子序列。为此，我们提出**组子序列策略优化（GSsPO）**。这种结构感知强化学习算法将优化粒度与智能体的决策单元显式对齐。通过将每个子序列视为原子优化单元，GSsPO 弥合了细粒度反馈与全局一致性之间的差距，确保策略更新严格遵循智能体推理的决策边界。

总而言之，这项工作通过三项主要贡献，推动了自动化工作流优化的发展前沿：

1. 转向动态交互的范式变革。 我们提出了 Workflow-R1，将工作流构建从静态程序合成转变为动态序贯决策过程。

2. 结构感知策略优化。我们提出**组子序列策略优化（GSsPO）**，以解决多轮推理中的粒度不匹配。关键在于，GSsPO 是一种具有泛化能力的算法解决方案，能够适应并提升不同多轮智能体系统的序贯推理能力。

3. 实证达到最先进水平。我们证明，Workflow-R1 在复杂基准测试上实现了最先进的性能。这些结果通过实证验证了我们新颖的动态框架的有效性，并确立了其作为自动化工作流优化领域一种颇具前景的新范式的地位。

2. 预备知识

本文将参数化为 θ的大语言模型形式化为策略 π_θ。给定从数据集 $\mathcal{D}$中采样的输入提示x ，模型会生成响应 $y = [y_1, y_2, \ldots, y_T]$，其中 y_t 表示第 t步的词元。联合概率分解为 $\pi_\theta(y|x) = \prod_{t=1}^{T} \pi_\theta(y_t|y_{<t}, x)$。

2.1. 组相对策略优化（GRPO）

GRPO 通过组级采样估计优势，无需使用价值模型（Shao 等人，2024年）。对于每个输入，它会生成多个输出，并通过对组内奖励进行归一化来计算相对优势。优化目标定义在词元级（为简洁起见，我们省略 KL 正则化）：

$$\mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}_{x\sim\mathcal{D},\{y_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(\cdot|x)} \left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|y_i|}\sum_{t=1}^{|y_i|} \min\left(r_{i,t}(\theta)\widehat{A}_{i,t}, \text{clip}(r_{i,t}(\theta), 1-\varepsilon, 1+\varepsilon)\widehat{A}_{i,t}\right)\right], \tag{1}$$

其中，词元级重要性采样比率 $r_{i,\,t}(\theta)$ 定义如下：

$$r_{i,t}(\theta) = \frac{\pi_\theta(y_{i,t}|y_{i,<t}, x)}{\pi_{\theta_{\text{old}}}(y_{i,t}|y_{i,<t}, x)}. \tag{2}$$

与先前工作一致，本文中的所有方法均采用基于组的优势估计，即对每个采样组内的奖励进行标准化来计算优势：

$$\widehat{A}_{i,t} = \frac{r(x, y_i) - \text{mean}\left(\{r(x, y_j)\}_{j=1}^{G}\right)}{\text{std}\left(\{r(x, y_j)\}_{j=1}^{G}\right)}, \tag{3}$$

其中，$r(x, y_i)$ 表示在输入 x 条件下第 i 个采样响应的奖励。与结果监督强化学习一致，该公式将响应级优势均匀地传播到生成响应中的每个词元。

2.2. 组序列策略优化（GSPO）。

与 GRPO 的词元级优化不同，GSPO 将整个响应序列视为优化单位（Zheng 等人，2025年）。因此，它在序列级应用重要性采样比率，使梯度更新与整体生成质量保持一致：

$$\mathcal{J}_{\text{GSPO}}(\theta) = \mathbb{E}_{x\sim\mathcal{D},\{y_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(\cdot|x)} \left[\frac{1}{G}\sum_{i=1}^{G} \min\left(s_i(\theta)\widehat{A}_i, \text{clip}\left(s_i(\theta), 1-\varepsilon, 1+\varepsilon\right)\widehat{A}_i\right)\right], \tag{4}$$

其中，序列级重要性采样比率 $s_i(\theta)$ 是完整响应中各词元似然比的几何平均值：

$$s_i(\theta) = \left(\frac{\pi_\theta(y_i|x)}{\pi_{\theta_{\text{old}}}(y_i|x)}\right)^{\frac{1}{|y_i|}} = \exp\left(\frac{1}{|y_i|}\sum_{t=1}^{|y_i|}\log\frac{\pi_\theta(y_{i,t}|y_{i,<t}, x)}{\pi_{\theta_{\text{old}}}(y_{i,t}|y_{i,<t}, x)}\right). \tag{5}$$

$\widehat{}$优势 A_i 遵循与 GRPO 相同的组归一化公式，从而保持奖励归一化的一致性。

3. Workflow-R1

本节介绍 Workflow-R1 框架。我们首先介绍组子序列策略优化，这是我们的核心强化学习算法。随后，我们在强化学习设置下形式化多轮工作流构建任务。接着，我们定义多轮工作流构建的奖励函数。最后，我们明确提示模板、算子池和算子接口协议。

3.1. GSsPO：组子序列策略优化

动机。 现有的大语言模型强化学习范式采用的粒度与多轮决策的序列特性存在根本性错配。以 GRPO 为代表的词元级优化将复杂推理链视为松散组合、彼此独立的词元。通过针对孤立词元而非连贯的决策单元进行更新，这种方法可能会破坏决策过程的语义完整性。相反，GSPO 等序列级优化则将整个交互视为

一个决策单元。尽管这与最终奖励相符，却掩盖了不同的决策边界，无法区分有效与错误的决策单元。

为弥合这一差距，我们认为，优化粒度必须严格与智能体的决策边界保持一致。因此，我们提出了**组子序列策略优化（GSsPO）**。GSsPO 的核心设计是重新校准重要性采样边界，使其与解析得到的子序列边界一致；这些子序列被定义为对应于原子决策单元的连续片段。通过将优化建立在这一层级上，GSsPO 确保梯度更新尊重每个决策单元的逻辑连贯性。此外，为了减轻冗长偏差，我们明确消除了动作基数带来的混杂因素。这确保优化过程由决策单元的质量而非其长度所驱动，从而显著提升模型在复杂多轮工作流中的性能。

目标函数。 我们提出的方法 GSsPO 将优化粒度重新校准至子序列，以提升策略性能。GSsPO 不在词元或序列层级聚合梯度，而是将每个组成子序列视为一个离散的原子优化单元。形式化地，对于采样得到的响应 y_i，令 $\mathcal{S}_i$ 表示从响应 i中解析出的子序列集合。优化目标函数定义如下：

$$\mathcal{J}_{\text{GSsPO}}(\theta) = \mathbb{E}_{x\sim\mathcal{D},\{y_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(\cdot|x)}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|\mathcal{S}_i|}\sum_{s\in\mathcal{S}_i}\min\left(r_s(\theta)\widehat{A}_s, \text{clip}(r_s(\theta),1-\varepsilon,1+\varepsilon)\widehat{A}_s\right)\right], \tag{6}$$

其中，内层求和遍历从采样序列中解析出的所有子序列s。优势项 $\widehat{A}_s$ 继承序列级结果，这意味着同一响应中的所有子序列共享相同的组归一化优势值 A_i。子序列级重要性采样比率 $r_s(\theta)$定义为词元概率比的几何平均数：

$$\begin{aligned} r_s(\theta) &= \left(\frac{\prod_{t\in s}\pi_\theta(y_t|y_{<t},x)}{\prod_{t\in s}\pi_{\theta_{\text{old}}}(y_t|y_{<t},x)}\right)^{\frac{1}{|s|}} \\ &= \exp\left(\frac{1}{|s|}\sum_{t\in s}\log\frac{\pi_\theta(y_t|y_{<t},x)}{\pi_{\theta_{\text{old}}}(y_t|y_{<t},x)}\right). \end{aligned} \tag{7}$$

策略梯度。为优化策略参数 θ，我们计算优化目标 $\mathcal{J}_{\text{GSsPO}}$的梯度。策略梯度的推导如下（为简洁起见，省略了裁剪为简洁起见）：

$$\begin{aligned} \nabla_\theta\mathcal{J}_{\text{GSsPO}}(\theta) &= \nabla_\theta\mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|\mathcal{S}_i|}\sum_{s\in\mathcal{S}_i}r_s(\theta)\widehat{A}_s\right] \\ &= \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|\mathcal{S}_i|}\sum_{s\in\mathcal{S}_i}\widehat{A}_s\cdot\nabla_\theta r_s(\theta)\right] \\ &= \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|\mathcal{S}_i|}\sum_{s\in\mathcal{S}_i}r_s(\theta)\widehat{A}_s\cdot\nabla_\theta\log r_s(\theta)\right] \\ &= \mathbb{E}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|\mathcal{S}_i|}\sum_{s\in\mathcal{S}_i}\underbrace{\left(\frac{\pi_\theta(s|x)}{\pi_{\theta_{\text{old}}}(s|x)}\right)^{\frac{1}{|s|}}}_{r_s(\theta)}\widehat{A}_s \\ &\quad\cdot\underbrace{\frac{1}{|s|}\sum_{t\in s}\nabla_\theta\log\pi_\theta(y_t|y_{<t},x)}_{\text{Mean Token Gradient in Sub-sequence}}\Bigg]. \end{aligned} \tag{8}$$

为进行比较分析，附录 B 中给出了 GRPO 和 GSPO 的梯度推导。与现有强化学习方法不同，我们的公式在优化机制上引入了两项根本性变化。首先，在重要性采样方面，GSsPO 为同一子序列中的所有词元分配统一权重 $r_s(\theta)$。其次，梯度聚合采用分层的两阶段平均机制：先对每个子序列内的词元长度进行归一化，再按子序列总数 $|\mathcal{S}_i|$取平均。

3.2. 多轮思考-行动工作流构建

不同于合成固定执行图的静态范式，`Workflow-R1`将工作流构建重新表述为一个序贯决策过程。智能体通过递归认知循环自主探索工作流拓扑，动态发现最优算子组合，确保生成的拓扑符合特定查询的需求。交互遵循自然语言接口协议。收到查询后，智能体进入推理阶段，并在 <思考> 标签内制定策略性决策：通过 <答案> 结束，或通过 <工具>调用算子。我们将每轮生成的**思考-行动对**形式化定义为任务中的**子序列**。由于不同的推理过程与后续算子调用共同构成一个统一连贯的决策单元，我们将其确立为优化的原子单元，其分段边界明确无歧义（每一轮恰好构成一个子序列）。抽象的算子执行引擎拦截指令，并将算子执行结果封装在 <信息>标签中返回。这些反馈闭合了感知-行动循环，并触发后续推理阶段。通过消除僵化的结构

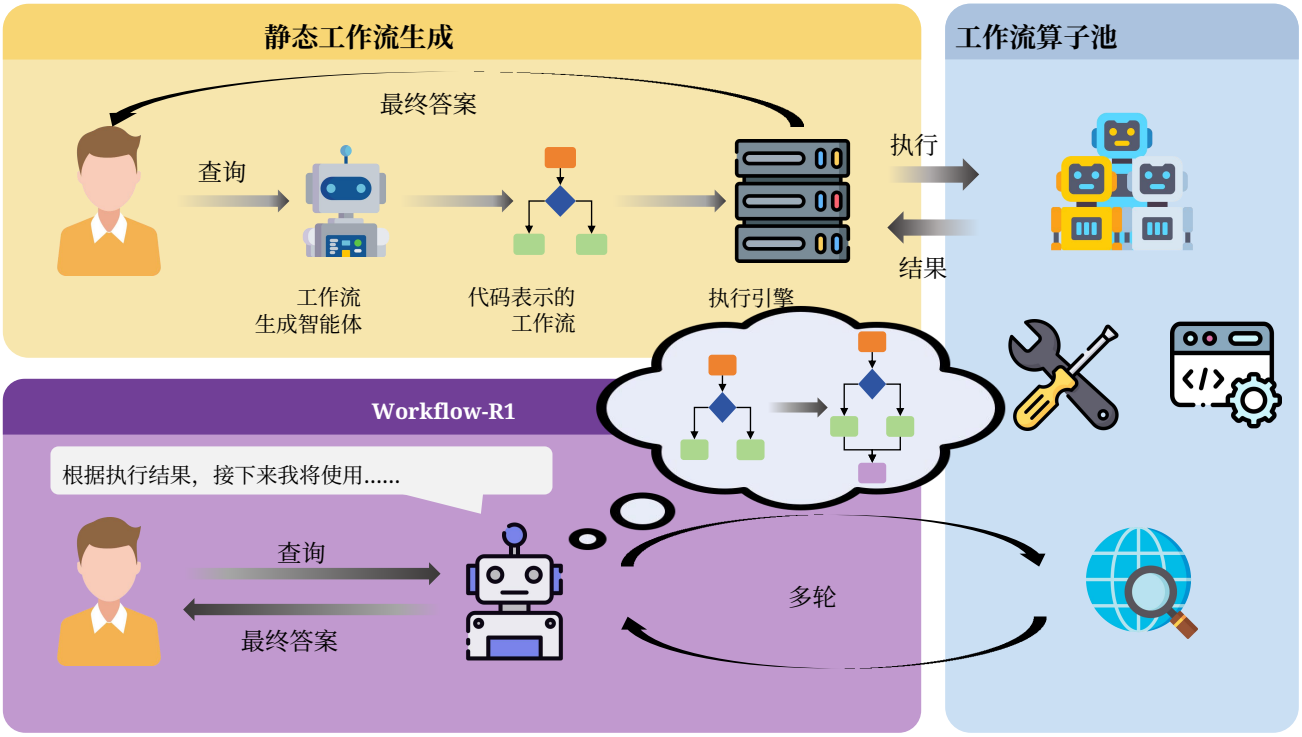


图 2. 上方展示了传统的静态工作流生成方式，智能体会生成完整的可执行代码表示的工作流。下方展示了 Workflow-R1 的多轮交互，智能体根据执行结果，通过思考-行动-观察循环逐步构建并调整工作流。

边界，该架构使模型能够积极探索算子协同效应，并自适应地编排资源，以进行深入的组合推理。

为实现自主工作流构建，我们采用强化学习框架。针对工作流算子池 $\mathcal{W}$ 的一般优化目标定义如下：

$$\max_{\pi_\theta} \mathbb{E}_{x\sim\mathcal{D}, y\sim\pi_\theta(\cdot|x;\mathcal{W})} \left[r_\phi(x,y) - \beta \log \frac{\pi_\theta(y|x;\mathcal{W})}{\pi_{\text{ref}}(y|x;\mathcal{W})} \right]. \tag{9}$$

该公式具有通用性，兼容多种强化学习范式，例如 GRPO 和 GSPO。在本工作中，我们采用提出的 GSsPO 作为默认算法，以优化工作流构建智能体的策略。

3.3. 奖励设计

我们采用基于规则的复合奖励函数，联合优化结构有效性和响应准确性。为支持复杂工作流的自主探索，同时确保格式正确，我们施加严格的结果监督和格式约束。

格式奖励。 为确保遵循接口协议，我们根据以下因素施加分级惩罚 $\mathcal{R}_{\text{Format}}$ 语法违规的严重程度：

$$\mathcal{R}_{\text{Format}} = \begin{cases} 0, & \text{if the format is completely correct} \\ -0.5, & \text{if the tag order is incorrect} \\ -1, & \text{if the <answer> tag is missing} \end{cases} \tag{10}$$

结果监督。 我们采用**精确匹配（EM）**作为结果奖励，以验证最终答案与真实答案是否一致：

$$\mathcal{R}_{\text{Outcome}}(y,g) = \text{EM}(y,g). \tag{11}$$

因此，最终奖励信号通过整合格式约束与结果监督构建而成。总奖励 $r(x,y)$ 的公式如下：

$$r_\phi(x,y) = \mathcal{R}_{\text{Format}} + \mathcal{R}_{\text{Outcome}}. \tag{12}$$

这种集成确保模型能够在严格的格式约束下构建工作流，同时保留探索多种算子组合的自由。

3.4. 算子与提示模板

我们使用一组通过自然语言接口定义实现的算子来构建工作流算子池。该设计使模型能够理解

方法	多轮	通用问答			多跳问答				Avg.
		NQ†	TriviaQA	PopQA	HpQA†	2wiki	Musique	Bamb	
Qwen2.5-32B-Instruct									
直接	✗	0.126	0.471	0.133	0.208	0.289	0.063	0.408	0.243
CoT	✗	0.222	0.579	0.205	0.234	0.286	0.079	0.480	0.298
SC（思维链×5）	✗	0.241	0.595	0.214	0.258	0.290	0.084	0.448	0.304
MedPrompt	✗	0.242	0.603	0.211	0.264	0.300	0.086	0.448	0.308
智能体工作流									
MaAS	✗	0.233	0.571	0.193	0.239	0.231	0.062	0.384	0.273
AFlow	✗	0.250	0.603	0.232	0.273	0.312	0.073	**0.504**	0.321
Workflow-R1	✓	**0.301**	**0.612**	**0.246**	**0.274**	**0.314**	**0.095**	0.472	**0.331**
搜索增强									
Search-R1（GRPO）	✓	0.429	0.623	0.427	0.386	0.346	0.162	0.400	0.396
Search-R1（PPO）	✓	0.393	0.610	0.397	0.370	0.414	0.146	0.368	0.385
Workflow-R1-Search	✓	**0.435**	**0.733**	**0.493**	**0.461**	**0.633**	**0.215**	**0.576**	**0.507**

表*1*。 基准测试中各基线方法的性能比较。**多轮** 列表示该方法是否涉及迭代交互。 † 表示域内数据集。

并通过自然语言指令直接执行算子调用。完整规范（包括详细的算子用法、接口协议以及完整的训练提示模板）见附录 D。

4. 实验

4.1. 数据集

我们在一套全面的七个数据集上评估 Workflow-R1 和其搜索增强版本 Workflow-R1-Search，这些数据集旨在测试智能体推理能力的极限。数据集分为两类：（1）通用问答，包括自然问题（Kwiatkowski et al., 2019）、TriviaQA（Joshi et al., 2017）和 PopQA（Mallen et al., 2023）；（2）多跳问答，包括 HotpotQA（Yang et al., 2018）、2WikiMultiHopQA（Ho et al., 2020）、Musique（Trivedi et al., 2022）和 Bamboogle（Press et al., 2023）。

数据划分。 在训练阶段，我们从自然问题和 HotpotQA 的训练划分中抽取一个子集，构建了一个数据集。如表 4 所示，最终共包含 10,000 个训练样本（自然问题提供 4,618 个，HotpotQA 提供 5,382 个）。这种特定的混合策略经过精心设计，旨在让智能体在强化学习探索阶段同时接触通用问答和复杂的多跳问答。

指标。我们采用全部七个基准测试的测试集。该综合评估方案使我们能够评估域内性能和域外泛化能力。我们使用精确匹配（EM）分数作为指标来衡量正确性。

我们在 Qwen2.5-32B-Instruct（Yang 等人，2024年）上使用**DirectInference**、**思维链**（CoT）（Wei 等人，2022年）、采用 5 条路径的自一致性（SC）（Wang 等人，2022年）以及 MedPrompt（Nori 等人，2023年）建立基线。对于工作流优化，我们将 Workflow-R1 与 AFlow（Zhang 等人，2025d）和 MaAS（Zhang 等人，2025a）进行比较，后两者分别代表当前最先进的任务级和查询级方法。这些基线的详细配置见附录 A.2。最后，我们将 Workflow-R1-Search 与 Search-R1（Jin 等人，2025年）的 PPO（Schulman 等人，2017年）和 GRPO（Shao 等人，2024年）变体进行基准测试，以评估我们的工作流优化框架在搜索增强场景中的有效性。

4.2. 实验详情

模型配置。我们采用 Qwen2.5-7B-Instruct（Yang 等人，2024年）作为工作流构建智能体的骨干模型。我们在所有智能体工作流方法（包括 AFlow 和 MaAS）中，统一使用固定的 Qwen2.5-32B-Instruct（Yang 等人，2024年）作为执行模型。此外，在第5节介绍的消融实验中，我们还使用 DeepSeek V3.2（Liu 等人，2025年）作为执行模型评估性能5。

训练设置。 我们采用全参数微调来训练两种策略变体，`Workflow-R1` 和`Workflow-R1-Search`，二者唯一的区别在于可用的算子池。除非另有说明，我们默认采用 GSsPO 作为强化学习算法。表1 报告了全面的基准测试结果，验证了我们的框架在标准场景和搜索增强场景下的表现；表3 则展示了消融实验，对比 GSsPO、GRPO 和 GSPO。有关详细的超参数和配置，请参阅附录A.3。

表 2. 采用**DeepSeek V3.2** 作为执行模型时，各方法在问答基准上的性能比较。此消融实验评估了工作流优化在不同骨干模型上的鲁棒性。† 表示域内数据集。

方法	通用问答			多跳问答				Avg.
	NQ†	TriviaQA	PopQA	HpQA†	2wiki	Musique	Bamb	
DeepSeek V3.2								
直接	0.318	0.713	0.359	0.299	0.318	0.091	0.336	0.347
CoT	0.332	0.714	0.377	0.355	0.455	0.156	0.608	0.428
SC（思维链×5）	0.295	0.693	0.380	0.336	0.421	0.170	0.616	0.416
MedPrompt	0.288	0.688	0.370	0.338	0.413	0.179	0.624	0.414
智能体工作流								
MaAS	0.335	0.685	0.358	0.323	0.339	0.123	0.440	0.372
AFlow	0.382	0.727	0.426	0.402	0.470	0.185	0.584	0.454
Workflow-R1（我们的方法）	0.440	**0.765**	0.487	0.428	0.548	0.202	**0.632**	0.500
搜索增强								
Search-R1（GRPO）	0.429	0.623	0.427	0.386	0.346	0.162	0.400	0.396
Search-R1（PPO）	0.393	0.610	0.397	0.370	0.414	0.146	0.368	0.385
Workflow-R1-Search（我们的）	**0.456**	0.749	**0.498**	**0.460**	**0.621**	**0.211**	0.584	**0.511**

方法	NQ†	TriviaQA	HpQA†	Musique
Workflow-R1				
GRPO	0.265	0.600	0.269	0.083
GSPO	0.283	0.609	0.272	0.091
GSsPO	**0.301**	**0.612**	**0.274**	**0.095**
Workflow-R1-Search				
GRPO	0.430	0.730	0.456	0.205
GSPO	**0.457**	0.731	0.450	0.204
GSsPO	0.435	**0.733**	**0.461**	**0.215**

表 3. 不同强化学习方法在各数据集上的性能比较。

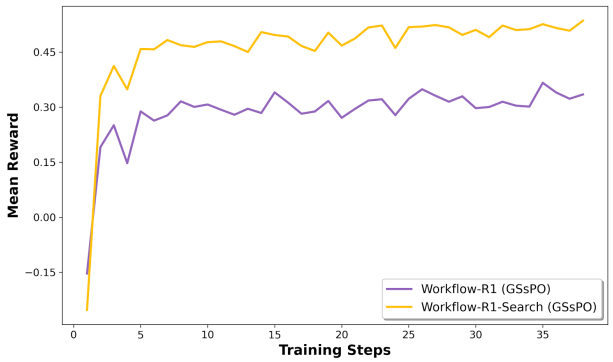


图 3. GSsPO 的平均奖励收敛。

4.3. 实验分析

智能体工作流生成方面的优势。Workflow-R1 在标准设置下达到了最先进的性能，显著优于推理时提示基线以及 AFlow 和 MaAS 等静态工作流优化方法。这证实了，与静态方法相比，我们的动态优化范式在推理任务上从根本上更为有效。此外，加入 Search 算子后，性能提升到了新的水平：
Workflow-R1-Search 在所有基线中得分最高，在所评估的基准测试上确立了最先进水平，平均精确匹配为 0.507。

案例研究。 图4 展示了构建出的工作流，揭示了**深度反思能力**的涌现：智能体通过持续监控算子执行结果，动态调整策略，从而超越僵化的执行序列。例如，智能体会重新考虑是否接受不确定的搜索结果，而非照单全收；它会自主调用审查操作符对结果进行验证。此外，遇到执行失败时，智能体会立即调整计划，尝试其他策略，例如利用 PromptOptimizer 重新表述查询表达式，随后调用其他算子再次尝试解决问题。这种理性的坚持使智能体能够自我纠正，并为复杂查询构建最优解决路径。如需查看原始响应的更多示例，请参阅附录 E。

5. 消融研究

高级骨干模型下的鲁棒性。 我们通过将直接问答和工作流执行所用的骨干模型升级为 DeepSeek V3.2，评估了鲁棒性。这一替换使所有智能体工作流方法的性能都大幅提升。值得注意的是，自动化工作流优化相较于基础提示基线的性能优势

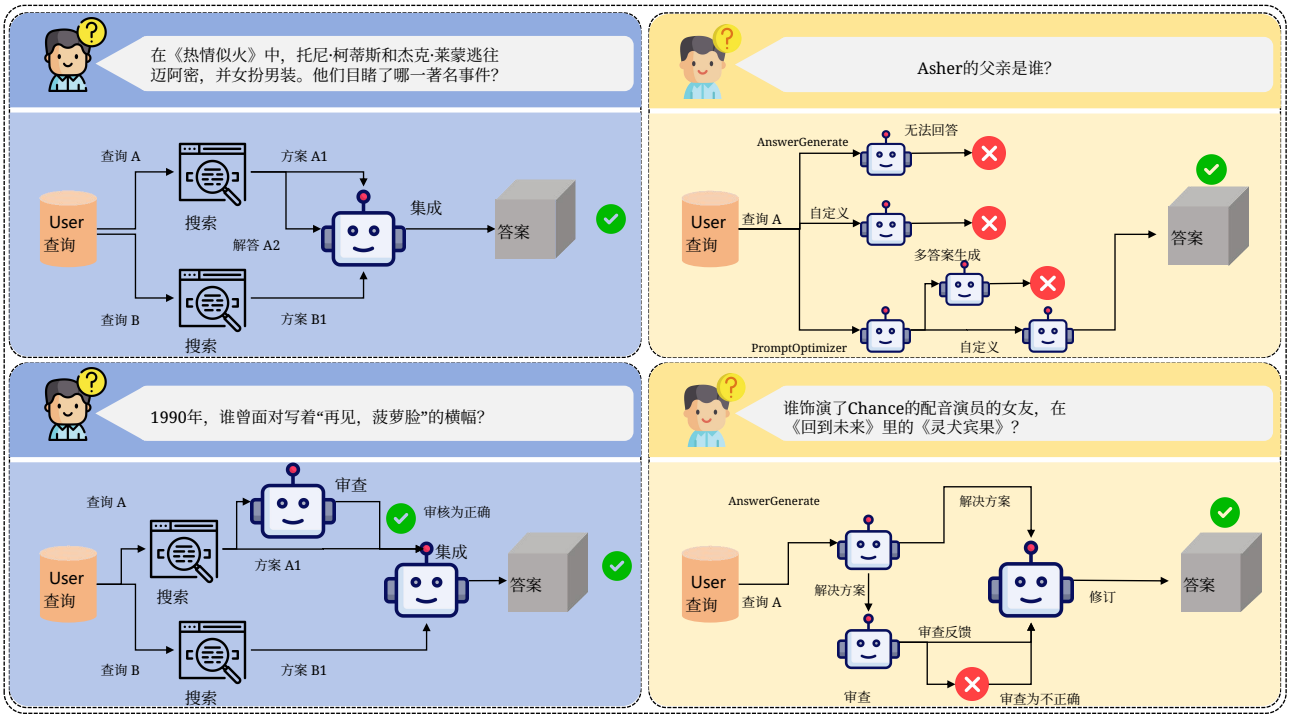


图 4. 可视化 Workflow-R1-Search（左，蓝色）和 Workflow-R1（右，黄色）的工作流构建过程).

（如自洽性和 MedPrompt）显著扩大。这些结果表明，优越的执行模型对于充分发挥复杂工作流的作用、实现性能的质的飞跃至关重要。

不同的强化学习算法。如表3所示，GSsPO 在所有数据集上的表现均持续优于 GRPO，并在大多数任务上超过 GSPO，验证了子序列是最优的优化单元。此外，图3 证实，GSsPO 保持了稳定的训练动态，并实现了高效收敛。

6. 相关工作

自动化工作流优化。智能体演化涵盖三种范式。DSPy（Khattab 等人，2023年）和 TextGrad（Yuksekgonul 等人，2024年）等模块化框架使用文本反馈，而 GPTSwarm（Zhuge 等人，2024年）则优化连接性。以代码为中心的 ADAS（Hu 等人，2024年）通过元智能体编写架构程序，AFlow（Zhang 等人，2025d）则使用蒙特卡罗树搜索对其进行了扩展。MAS-GPT（Ye 等人，2025年）、MaAS（Zhang 等人，2025a）和 FlowReasoner（Gao 等人，2025年）等查询自适应方法则会针对每个查询动态生成结构。然而，大多数方法仍受限于“静态执行陷阱”：预先固定的结构无法根据中间观测进行调整。

面向大语言模型的强化学习。强化学习通过多种方法对齐大语言模型，从基于人类反馈的强化学习（Ouyang 等人，2022年）（例如 PPO（Schulman 等人，2017年））到 DPO（Rafailov 等人，2023年）和 RLOO（Ahmadian 等人，2024年），但这类方法往往面临离策略推理挑战（Pang 等人，2024年）。近年来，DeepSeek-R1（Guo 等人，2025年）和 GRPO（Shao 等人，2024年）等“纯强化学习”范式通过组奖励消除了评论器，而 DAPO（Yu 等人，2025年）进一步提升了稳定性。然而，多轮场景中仍存在粒度不匹配问题：GRPO 的令牌级更新可能破坏语义完整性，而 GSPO（Zheng 等人，2025年）的序列级信号对于精确的信用分配而言仍过于粗略。

7. 结论

本文提出了 Workflow-R1，这一框架从根本上将工作流构建从静态合成转变为动态序贯决策过程。为实现这种适应查询的能力，我们提出了组子序列策略优化（GSsPO）。通过将思考-行动对设为决策单元，GSsPO 有效解决了传统方法固有的优化粒度不匹配问题。大量评估结果表明，Workflow-R1 在七个基准测试中显著优于先进基线。尤为重要的是，我们的分析揭示了深度反思能力的涌现，使智能体能够在执行过程中自主自我纠正。最终，Workflow-R1 为自主智能体奠定了坚实基础，展现了动态工作流优化在复杂推理领域的巨大潜力。

影响声明

本文介绍的研究旨在推动机器学习领域的发展。我们的工作可能会对社会产生许多影响，但我们认为没有哪项影响需要在此特别强调。

参考文献

Ahmadian,A., Cremer, C., Gall´e, M., Fadaee, M., Kreutzer,¨J., Pietquin，O.，Ust¨un, A., and Hooker, S. 回归基础：重新审视大语言模型中用于从人类反馈中学习的强化式优化。*arXiv* 预印本 *arXiv:2402.14740*，2024年。Anthropic。推出 Claude 3.5 Sonnet。
h
t
t
p
s://
www.anthropic.com/news/claude-3-5-sonnet，
2024年。Feng, J., Huang, S., Qu, X., Zhang, G., Qin, Y., Zhong, B., Jiang, C., Chi, J., and Zhong, W. Retool：大语言模型中用于策略性工具使用的强化学习。 *arXiv* 预印本 *arXiv:2504.11536*，2025年。Gao, H., Liu, Y., He, Y., Dou, L., Du, C., Deng, Z., Hooi, B., Lin, M., and Pang, T. FlowReasoner：强化查询级元智能体。 *arXiv* 预印本 *arXiv:2504.15257*，2025年。Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al. DeepSeek-R1：通过强化学习激励大语言模型的推理能力。 *arXiv* 预印本 *arXiv:2501.12948*，2025年。Ho, X., Nguyen, A.-K. D., Sugawara, S., and Aizawa, A. 构建多跳问答数据集，以全面评估推理步骤。 *arXiv* 预印本 *arXiv:2011.01060*，2020 年。Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Zhang, C., Wang, J., Wang, Z., Yau, S. K. S., Lin, Z., Zhou, L., Ran, C., Xiao, L., Wu, C., and Schmidhuber, J. MetaGPT：面向多智能体协作框架的元编程。pp. et al；Google Deepmind；Google Research；Meta；Microsoft –，Hybrid, Vienna，Austria，2024年。基于智能体；协作框架；复杂任务；语言模型；元编程；基于模型的OPC；多智能体；问题求解；智能体社会；工作流；。Hu, S., Lu, C., and Clune, J. 智能体系统的自动化设计。 *arXiv* 预印本 *arXiv:2408.08435*，2024年。Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A., et al. GPT-4o 系统卡。 *arXiv* 预印本 *arXiv:2410.21276*，2024 年。Jin, B., Zeng, H., Yue, Z., Yoon, J., Arik, S., Wang, D., Zamani, H. 和 Han, J. Search-r1：通过强化学习训练大语言模型进行推理并利用搜索引擎。*arXiv* 预印本 *arXiv:2503.09516*，2025年。Joshi, M., Choi, E., Weld, D. S. 和 Zettlemoyer, L. Triviaqa：用于阅读理解的大规模远程监督挑战数据集。*arXiv* 预印本 *arXiv:1705.03551*，2017年。Khattab, O., Singhvi, A., Maheshwari, P., Zhang, Z., San- thanam, K., Vardhamanan, S., Haq, S., Sharma, A., Joshi, T. T., Moazam, H. 等。Dspy：将声明式语言模型调用编译为自我改进流水线。*arXiv* 预印本 *arXiv:2310.03714*，2023年。Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., Epstein, D., Polosukhin, I., Devlin, J., Lee, K. 等。Natural questions：面向问答研究的基准测试。计算语言学协会会刊，7:453–466，2019年。Li, X., Dong, G., Jin, J., Zhang, Y., Zhou, Y., Zhu, Y., Zhang, P. 和 Dou, Z. Search-o1：智能体搜索增强型大型推理模型。*arXiv* 预印本 *arXiv:2501.05366*，2025年。Liu, A., Mei, A., Lin, B., Xue, B., Wang, B., Xu, B., Wu, B., Zhang, B., Lin, C., Dong, C. 等。Deepseek-v3. 2：推动开放式大语言模型的前沿。*arXiv* 预印本 *arXiv:2512.02556*，2025年。Mallen, A., Asai, A., Zhong, V., Das, R., Khashabi, D. 和 Hajishirzi, H. 何时不应信任语言模型：考察参数化记忆与非参数化记忆的有效性。载于第*61*届计算语言学协会年会论文集（第*1*卷：长篇论文），第9802–9822页，2023年。Nie, F., Feng, L., Ye, H., Liang, W., Lu, P., Yao, H., Alahi, A. 和 Zou, J. 弱助强：训练弱元智能体以驾驭强大执行器。*arXiv* 预印本 *arXiv:2504.04785*，2025年。Nori, H., Lee, Y. T., Zhang, S., Carignan, D., Edgar, R., Fusi, N., King, N., Larson, J., Li, Y., Liu, W. 等。通用基础模型能否胜过专用调优？医学领域的案例研究。*arXiv* 预印本 *arXiv:2311.16452*，2023年。Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. 等。训练语言模型遵循人类反馈下的指令。神经信息处理系统进展，35:27730–27744，2022年。

Pang, R. Y., Yuan, W., He, H., Cho, K., Sukhbaatar, S. 和 Weston, J. 迭代推理偏好优化。神经信息处理系统进展，37: 116617–116637，2024年。

Press, O., Zhang, M., Min, S., Schmidt, L., Smith, N. A., and Lewis, M. 测量并缩小语言模型中的组合性差距。见计算语言学协会发现：*EMNLP 2023*，第5687–5711页，2023年。

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., and Finn, C. 直接偏好优化：你的语言模型其实是一个奖励模型。神经信息处理系统进展，36：53728–53741，2023年。

Schulman, J.、Wolski, F.、Dhariwal, P.、Radford, A.与Klimov, O. 近端策略优化算法。*arXiv*预印本 *arXiv:1707.06347*，2017年。Shao, Z.、Wang, P.、Zhu, Q.、Xu, R.、Song, J.、Bi, X.、Zhang, H.、Zhang, M.、Li, Y.、Wu, Y.等。Deepseekmath：推动开放语言模型中数学推理的极限。*arXiv*预印本 *arXiv:2402.03300*，2024年。Sheng, G.、Zhang, C.、Ye, Z.、Wu, X.、Zhang, W.、Zhang, R.、Peng, Y.、Lin, H.与Wu, C. Hybridflow：一个灵活高效的基于人类反馈的强化学习框架。载于第二十届欧洲计算机系统会议论文集，第1279–1297页，2025年。Su, J.、Xia, Y.、Lan, Q.、Song, X.、Jingsong, Y.、He, L.与Shi, T. 大语言模型驱动工作流中的难度感知型智能体编排。*arXiv*电子预印本，第arXiv–2509页，2025年。Trivedi, H.、Balasubramanian, N.、Khot, T.与Sabharwal, A. Musique：通过单跳问题组合实现的多跳问题。计算语言学协会会刊，10:539–554，2022年。Wang, X.、Wei, J.、Schuurmans, D.、Le, Q.、Chi, E.、Narang, S.、Chowdhery, A.与Zhou, D. 自一致性提升语言模型中的思维链推理。*arXiv*预印本 *arXiv:2203.11171*，2022年。Wang, Y.、Liu, S.、Fang, J.与Meng, Z. Evoagentx：一种用于演化智能体工作流的自动化框架。载于*2025*年自然语言处理实证方法会议论文集：系统演示，第643–655页，2025a。Wang, Y.、Yang, L.、Li, G.、Wang, M.与Aragam, B. Score-flow：通过基于分数的偏好优化掌握大语言模型智能体工作流。*arXiv*预印本 *arXiv:2502.04306*，2025b。

Wei, J.、Wang, X.、Schuurmans, D.、Bosma, M.、Xia, F.、Chi, E.、Le, Q. V.、Zhou, D.等。思维链提示激发大语言模型中的推理。神经信息处理系统进展，35:24824–24837，2022年。Wu, Q.、Bansal, G.、Zhang, J.、Wu, Y.、Li, B.、Zhu, E.、Jiang, L.、Zhang, X.、Zhang, S.、Liu, J.等。Autogen：通过多智能体对话实现下一代大语言模型应用。载于第一届语言建模会议，2024年。Yang, A.、Yang, B.、Zhang, B.、Hui, B.、Zheng, B.、Yu, B.、Li, C.、Liu, D.、Huang, F.、Wei, H.、Lin, H.、Yang, J.、Tu, J.、Zhang, J.、Yang, J.、Yang, J.、Zhou, J.、Lin, J.、Dang, K.、Lu, K.、Bao, K.、Yang, K.、Yu, L.、Li, M.、Xue, M.、Zhang, P.、Zhu, Q.、Men, R.、Lin, R.、Li, T.、Tang, T.、Xia, T.、Ren, X.、Ren, X.、Fan, Y.、Su, Y.、Zhang, Y.、Wan, Y.、Liu, Y.、Cui, Z.、Zhang, Z.与Qiu, Z. Qwen2.5技术报告。常识；专家知识；高质量；知识能力；语言模型；多阶段；预训练；推理能力；强化学习；训练数据集；，2024年。URL
h
t
t
p://dx.doi.org/10.48550/arXiv.2412.15115。
Yang, Z.、Qi, P.、Zhang, S.、Bengio, Y.、Cohen, W.、Salakhutdinov, R.与Manning, C. D. Hotpotqa：一个用于多样化、可解释多跳问答的数据集。载于*2018*年自然语言处理实证方法会议论文集，第2369–2380页，2018年。Ye, R.、Tang, S.、Ge, R.、Du, Y.、Yin, Z.、Chen, S.与Shao, J. Mas-gpt：训练大语言模型以构建基于大语言模型的多智能体系统。第267卷，第72063–72090页，加拿大不列颠哥伦比亚省温哥华，2025年。自适应多智能体系统；数据构建；可执行代码；高质量；多智能体系统（MASs）；新颖任务；开源；重新构建；系统方法；用户查询；。Yu, Q.、Zhang, Z.、Zhu, R.、Yuan, Y.、Zuo, X.、Yue, Y.、Dai, W.、Fan, T.、Liu, G.、Liu, L.等。Dapo：一个大规模开源大语言模型强化学习系统。*arXiv*预印本 *arXiv:2503.14476*，2025年。Yuksekgonul, M.、Bianchi, F.、Boen, J.、Liu, S.、Huang, Z.、Guestrin, C.与Zou, J. Textgrad：通过文本实现自动“微分”。*arXiv*预印本 *arXiv:2406.07496*，2024年。Zhang, G.、Niu, L.、Fang, J.、Wang, K.、Bai, L.与Wang, X. 通过智能体超网络进行多智能体架构搜索。*arXiv*预印本 *arXiv:2502.04180*，2025a。

Zhang, G.、Yue, Y.、Sun, X.、Wan, G.、Yu, M.、Fang, J.、Wang, K.、Chen, T. 和 Cheng, D. G-设计器：通过图神经网络构建多智能体通信拓扑。第267卷，

第76678–76692页，加拿大不列颠哥伦比亚省温哥华，2025b年。最佳选择；集体智能；通信拓扑；图神经网络；个体智能体；智能体间通信；语言模型；基于模型的OPC；多智能体通信；特定任务；。

Zhang, H., Feng, T., and You, J. Router-r1：通过强化学习教授大语言模型多轮路由与聚合。载于第三十九届神经信息处理系统年会，2025c。

Zhang, J., Xiang, J., Yu, Z., Teng, F., Chen, X.-H., Chen, J., Zhuge, M., Cheng, X., Hong, S., Wang, J., Zheng, B., Liu, B., Luo, Y., and Wu, C. Aflow：自动化智能体工作流生成。第 100351–100388 页，新加坡，新加坡，2025d。

Zhang, S., Dong, Y., Zhang, J., Kautz, J., Catanzaro, B., Tao, A., Wu, Q., Yu, Z., and Liu, G. Nemotron-research-tool-n1：通过强化推理探索工具使用语言模型。*arXiv* 预印本 *arXiv:2505.00024*，2025e。

Zhang, Y.、Liu, X. 和 Xiao, C. Metaagent：基于有限状态机构建多智能体系统。第 267 卷，第 75667–75694 页，加拿大不列颠哥伦比亚省温哥华，2025f。’ current；自动化设计方法；通信结构；有限状态机；语言模型；元智能体；多智能体框架；多智能体系统（MASs）；工具集成；训练数据；。

Zheng, C., Liu, S., Li, M., Chen, X.-H., Yu, B., Gao, C., Dang, K., Liu, Y., Men, R., Yang, A., 等。组序列策略优化。*arXiv* 预印本 *arXiv:2507.18071*，2025年。Zhou, H., Wan, X., Sun, R., Palangi, H., Iqbal, S., Vuli´c, ¨I., Korhonen, A., and Arık, S. O. 多智能体设计：通过更优的提示词与拓扑优化智能体。*arXiv* 预印本 *arXiv:2502.02533*，2025年。Zhuge, M., Wang, W., Kirsch, L., Faccio, F., Khizbullin, D., and Schmidhuber, J. Gptswarm：将语言智能体表示为可优化的图。在第四十一届国际机器学习会议，2024年。

A. 实验设置

A.1. 数据集

我们从 NQ（4,618 个样本）和 HotpotQA（5,382 个样本）中构建了一个包含 10,000 个样本的混合训练集，以平衡单跳和多跳推理能力。评估时，我们使用涵盖两个领域的七个数据集：通用问答（NQ、TriviaQA、PopQA）和多跳问答（HotpotQA、2WikiMultiHopQA、Musique、Bamboogle）。详细统计数据见表 4。

类别	数据集	训练	Test
通用问答	NQ	4,618	3,610
	TriviaQA	–	11,313
	PopQA	–	14,267
多跳问答	HotpotQA	5,382	7,405
	2WikiMultiHopQA	–	12,576
	Musique	–	2,417
	Bamboogle	–	125
总计		**10,000**	**51,713**

表*4.* 数据集统计。

A.2. 工作流基线

对于智能体工作流基线，MaAS 和 AFlow 分别代表当前最先进的查询级和任务级工作流优化方法。在复现中，我们遵循其官方实现：MaAS 使用 GPT-4o- mini（Hurst等人，2024年），AFlow 使用 Claude-3.5-Sonnet（Anthropic，2024年）作为各自的工作流优化模型。为确保公平比较，在主要实验的训练和测试中，我们统一使用 Qwen2.5-32B-Instruct 作为工作流执行模型。

A.3. 训练配置

我们使用 **Qwen2.5-7B-Instruct** 作为策略模型（工作流优化模型），并使用 **Qwen2.5-32B-Instruct**作为 `Workflow-R1` 的工作流执行模型。所有实验均采用 GSsPO 算法，通过 VeRL 框架（Sheng等人，2025年）在 2 块 NVIDIA H20 GPU 上进行全参数微调。详细超参数见表 5。

超参数	值
训练轮数	2
训练样本	10,000
全局批次大小	512
展开	8
总训练步数	38
学习率	1e-6
KL 系数	0.001
裁剪比率	0.2

表 *5.* Workflow-R1 和 Workflow-R1-Search 的训练超参数。

关于策略变体，Workflow-R1-Search 将 MultiAnswerGenerate 算子替换为 Search 算子，以支持外部信息检索；而 Workflow-R1 保留标准算子池，用于自包含推理。

B. More RL 的策略梯度

本节将详细推导 GRPO 和 GSPO 的梯度，以便进行比较分析。

B.1. GRPO

组相对策略优化（GRPO）在**令牌级**上运行。

优化目标：目标函数定义如下（为简洁起见，省略裁剪机制）：

$$\mathcal{J}_{GRPO}(\theta) = \mathbb{E}_{x\sim\mathcal{D},\{y_i\}_{i=1}^{G}\sim\pi_{\theta_{old}}} \left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|y_i|}\sum_{t=1}^{|y_i|} w_{i,t}(\theta)\hat{A}_{i,t} \right] \tag{13}$$

重要性比率定义：令牌级重要性比率 $w_{i,\ t}(\theta)$ 的定义如下：

$$w_{i,t}(\theta) = \frac{\pi_\theta(y_{i,t}|x, y_{i,<t})}{\pi_{\theta_{old}}(y_{i,t}|x, y_{i,<t})} \tag{14}$$

梯度推导：策略梯度通过对目标函数求导得出：

$$\begin{aligned}
\nabla_\theta \mathcal{J}_{GRPO}(\theta) &= \nabla_\theta \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|y_i|}\sum_{t=1}^{|y_i|} \frac{\pi_\theta(y_{i,t}|x, y_{i,<t})}{\pi_{\theta_{old}}(y_{i,t}|x, y_{i,<t})}\hat{A}_{i,t} \right] \\
&= \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|y_i|}\sum_{t=1}^{|y_i|} \hat{A}_{i,t}\nabla_\theta \left(\frac{\pi_\theta(y_{i,t}|x, y_{i,<t})}{\pi_{\theta_{old}}(y_{i,t}|x, y_{i,<t})} \right) \right] \\
&= \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|y_i|}\sum_{t=1}^{|y_i|} \frac{\pi_\theta(y_{i,t}|x, y_{i,<t})}{\pi_{\theta_{old}}(y_{i,t}|x, y_{i,<t})}\hat{A}_{i,t}\nabla_\theta \log \pi_\theta(y_{i,t}|x, y_{i,<t}) \right]
\end{aligned} \tag{15}$$

这一推导证实，GRPO 根据单个词元的似然更新策略，并以各自特定的令牌级重要性比率作为权重。

B.2. GSPO

组序列策略优化（GSPO）将**整个响应**视为一个整体。

优化目标：目标函数表述如下（为简洁起见，省略裁剪机制）：

$$\mathcal{J}_{GSPO}(\theta) = \mathbb{E}_{x\sim\mathcal{D},\{y_i\}_{i=1}^{G}\sim\pi_{\theta_{old}}} \left[\frac{1}{G}\sum_{i=1}^{G} s_i(\theta)\hat{A}_i \right] \tag{16}$$

重要性比率定义：序列级重要性采样比率 $s_i(\theta)$ 基于似然比的几何平均数定义：

$$s_i(\theta) = \left(\frac{\pi_\theta(y_i|x)}{\pi_{\theta_{old}}(y_i|x)} \right)^{\frac{1}{|y_i|}} \tag{17}$$

梯度推导： 梯度推导过程如下：

$$\begin{aligned}\nabla_\theta \mathcal{J}_{GSPO}(\theta) &= \nabla_\theta \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\left(\frac{\pi_\theta(y_i|x)}{\pi_{\theta_{old}}(y_i|x)}\right)^{\frac{1}{|y_i|}}\hat{A}_i\right] \\ &= \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\hat{A}_i\nabla_\theta\left(\left(\frac{\pi_\theta(y_i|x)}{\pi_{\theta_{old}}(y_i|x)}\right)^{\frac{1}{|y_i|}}\right)\right] \\ &= \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\left(\frac{\pi_\theta(y_i|x)}{\pi_{\theta_{old}}(y_i|x)}\right)^{\frac{1}{|y_i|}}\hat{A}_i\nabla_\theta\log\left(\left(\frac{\pi_\theta(y_i|x)}{\pi_{\theta_{old}}(y_i|x)}\right)^{\frac{1}{|y_i|}}\right)\right] \\ &= \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|y_i|}\sum_{t=1}^{|y_i|}\left(\frac{\pi_\theta(y_i|x)}{\pi_{\theta_{old}}(y_i|x)}\right)^{\frac{1}{|y_i|}}\hat{A}_i\nabla_\theta\log\pi_\theta(y_{i,t}|x,y_{i,<t})\right]\end{aligned} \tag{18}$$

与 GRPO 不同，GSPO 对响应中的所有词元应用统一的序列级权重，从而确保梯度更新与序列的语义完整性一致。

C. 更多结果

除了使用 32B 模型获得的主要结果外，我们还评估了较小骨干模型 **Qwen2.5-7B-Instruct**在各项基准测试中的表现。表 6 展示了四种基线方法的表现。

方法	多轮	通用问答			多跳问答				Avg.
		NQ†	TriviaQA	PopQA	HpQA†	2wiki	Musique	Bamb	
Qwen2.5-7B-Instruct									
直接	✗	0.030	0.051	0.056	0.098	0.184	0.011	0.048	0.068
CoT	✗	0.122	0.102	0.121	0.173	0.202	0.043	0.312	0.154
自洽性（思维链×5）	✗	0.142	0.110	0.130	0.188	0.216	0.050	0.336	0.167
MedPrompt	✗	0.114	0.104	0.119	0.146	0.155	0.046	0.320	0.143

表6. 基准测试上基线方法的性能比较。**多轮**列表示该方法是否涉及迭代交互。† 表示域内数据集。

D. 算子描述和提示模板的详细信息

D.1. 提示模板

我们介绍训练提示模板，其中规定了多轮协议和严格的格式约束，以及完整的算子池和调用语法。值得注意的是，我们引入了WarningOp，以防止格式违规（例如错误的 Think-Action结构）。该算子会拦截无效输出，并将警告消息作为观察反馈返回，明确引导智能体在后续步骤中进行自我纠正。

查询：提示模板

回答给定的问题。每当你从算子处获得新信息时，你必须在采取任何进一步行动之前，于 <think> 和 </think> 内进行推理。这意味着你应该进行多次思考会话——每次从算子处获得信息后都进行一次，而不只是最初进行一次思考。

推理后，如果你发现可以利用某个算子来提高获得正确答案的可能性，则可以根据下面的算子描述调用该算子。而且在调用算子时，你必须遵循相应的格式。

可用算子： {可用算子池}

每次调用算子之前，你必须在 <think> 和 </think> 中明确推理，说明你为什么选择这个算子、将如何使用它，以及希望从中了解什么。调用算子后，响应将返回在 <info> 和 </info>之间。收到响应后，你必须再次在 <think> 和 </think>中思考这些信息意味着什么、如何帮助回答原始问题，以及下一步应该做什么。

关键指令：不要满足于简单的单步算子调用。你必须积极探索多样、复杂且非常规的算子组合。你的目标不仅是给出答案，还要全面探索不同算子如何协同发挥作用。你必须突破这些算子协同工作时所能达到的极限。被动、懒散或过于简单的策略不可接受；你必须通过丰富的多步算子编排展现深入推理，以揭示它们独特的能力和“神奇”的组合效果。

重要限制：算子调用最多为 20 次。请谨慎规划策略——你必须在此限制内收集足够的信息并得出有把握的答案。

始终遵循以下循环：先思考，然后在需要时选择使用算子；收到信息后，再次思考，然后决定下一步。重复此循环，直到掌握足够的信息，能够有把握地作答。收集到足够的信息，并经过多个思考周期的深入推理后，如果你确实确信自己能正确回答查询，请在 <答案> 和 </答案> 内仅用几个字或一个简短短语直接给出最终答案，以准确回答查询。如果你还不确定，就继续收集信息并推理，直到确信无误。

查询： {用户查询}

算子说明与用法

AnswerGenerate:
说明：调用高级大语言模型进行思考并回答问题。适用于特定领域的问题、知识问答、推理任务等。**用法**： <工具>AnswerGenerate：你的问题</工具>

自定义:
说明：调用大语言模型执行你定义的任意任务。只要提供清晰、详细的指令和任务内容，大语言模型几乎可以完成你需要的任何事情。**用法**：<工具>自定义：<任务>你的任务内容</任务> <指令>你的指令</指令> </工具>

PromptOptimizer:
说明：调用大语言模型，通过补充细节并制定逐步计划，完善和优化查询的澄清内容。当原始查询含糊不清或需要更多上下文时，此工具很有用。**用法**： <工具>PromptOptimizer：你的查询</工具>

审查:
说明：调用大语言模型，对解决方案的正确性进行严格审查并提供反馈。在最终确定答案之前，若想验证其是否正确，此工具很有用。**用法**： <工具>审查： <问题>原始问题</问题><解决方案>待审查的解决方案</解决方案> </工具>

修订:
说明：根据反馈调用大语言模型修订解决方案。当你收到指出解决方案存在问题的严格反馈时，此工具很有用。

用法： <工具>修订： <问题>原始问题</问题><解决方案>待修订的解决方案</解决方案> <反馈>收到的反馈</反馈> </工具>

多答案生成:
描述：调用高级大语言模型，为同一个问题生成多个独立答案（3 次）。当你希望探索多样化的推理路径或准备候选答案以进行集成选择时，这非常有用。**用法**： <工具>多答案生成：你的问题</工具>

ScEnsemble:
描述：调用大语言模型，通过自一致性集成选择从多个候选方案中选出最佳答案。你必须至少提供两个候选方案。**用法**： <tool>ScEnsemble： <query>原始问题</query><solution>候选 1</solution>...<solution>候选 N</solution> </tool>

搜索:
描述：调用搜索引擎在网络上搜索信息。当你需要查找尚未掌握的事实时非常有用。**用法**： <tool>搜索：你的查询</tool>

WarningOp:
描述：一种特殊算子，当模型未能遵循所需输出格式时返回警告消息。

输出：警告！！！我没有遵循所需格式。如果我想使用算子，我应该使用<tool>OperatorName: input</tool>。如果我想给出最终答案，我应该使用 <answer>...</answer>。

E. 更多案例研究

在本节中，我们展示了一组全面的定性示例，包括四个使用 Workflow-R1-Search 的案例和四个使用 Workflow-R1 的案例。这些案例研究凸显了我们框架固有的若干引人注目的涌现行为。我们观察到，当算子因内部错误而未能返回有效信息时，智能体会避免停滞，转而智能地转向替代策略以解决查询。此外，我们识别出稳健的自我纠正机制；在触发格式警告后，智能体会在其推理轨迹中（ <think> 标签内）明确反思该错误。总体而言，这些涌现行为证明了我们所提出范式的实用价值，标志着对自主多轮工作流构建前沿的一次大胆探索。

查询：如果 Kevin Scale 采用创始人的名字作为名称，它会叫什么？

标准答案：William Scale

[助手]
<思考>首先，我需要弄清楚 Kevin Scale 是什么，以及在此语境下创始人的名字是什么。不过，查询没有提供关于 Kevin Scale 或创始人的足够背景信息。我先搜索一下 Kevin Scale 的相关信息，以便更好地了解其背景。</思考><工具>搜索：Kevin Scale 是什么？它指的是什么？</工具>

[算子]
<信息>“Kevin Scale 是什么？它指的是什么？”的搜索结果：标题：开尔文

片段：开尔文温标是一种绝对温标，其起点是可能达到的最低温度（绝对零度），定义为 0K。根据定义，摄氏.....

标题：开尔文温标定义片段：科学家使用开尔文温标，是因为他们希望有一种温标，其零点代表热能的完全缺失。

标题：开尔文（K）——定义与事实
摘录：　……通过将摄氏温标平移 −273.15°，使绝对零度与 0 K 重合）……　　.

标题：开尔文温标概述 — ScienceDirect Topics 摘录：开尔文温标是一种温标，它将摄氏温标延伸至绝对零度（0 K），代表完全没有……

标题：什么是开尔文温标?
摘录：概述开尔文温标是什么，以及它在化学中的应用。开尔文与摄氏度的比较，以及绝对零度是……</信息>

[助手]
<思考>搜索结果并未提供任何关于 Kevin Scale 的直接信息。不过，这些结果大多与开尔文温标有关，这可能是对查询的误解。我需要澄清“Kevin Scale”的语境以及创始人的名字。我将使用 PromptOptimizer 来优化该查询，以便更好地理解。</思考>

<工具>PromptOptimizer：如果 Kevin Scale 采用了创始人的名字，它会被称为什么？</工具>

[算子]
<信息>PromptOptimizer 思考过程：该查询询问的是一种假设情景，即某个概念或¨　　¨名为 Kevin Scale 的理论会根据创始人的名字更改名称。然而，查询中似乎存在¨　　¨一种误解，因为 Kevin Scale 似乎并不是已知的科学或理论概念。它¨可能是虚构或名称有误的概念。要优化此查询，我们需要澄清 Kevin Scale¨指的是什么，以及¨创始人是谁。如果我们假设 Kevin Scale¨是一个地点……的名字。例如，如果该概念最初以全名为 ¨John Kevin，¨的人命名，且目前被称为 ¨Kevin Scale，¨那么该查询想知道，若改为以 John 命名，它会被称为什么。
　　¨PromptOptimizer 答案：假设 Kevin Scale¨是一个以某个人命名的真实概念的占位符，¨那么优化后的查询将是：如果目前称为 Kevin Scale 的概念¨改为以创始人的名字命名，它会被称为什么？</info>

[助手]
<思考>PromptOptimizer 提供了更具体、更清晰的查询版本。然而，它仍未明确说明 ¨Kevin Scale¨所指的是什么。我需要搜索关于 ¨Kevin Scale¨的更具体背景，以了解它的起源以及创始人的名字。现在我将搜索 ¨Kevin Scale¨，以 выяс выяс выяс弄清楚它是什么以及创始人是谁。</think>

<工具>搜索：Kevin Scale 是什么，创始人是谁？</tool>

[算子]
<信息>“Kevin Scale 是什么，创始人是谁？”的搜索结果：
标题：开尔文
摘要：19世纪英国科学家开尔文勋爵最早提出并制定了这一温标。20世纪初，它常被称为 ¨绝对摄氏¨温标……

标题：摄氏温标与开尔文温标 — Research Starters
摘要：开尔文温标是另一种常见的温标。英国发明家和科学家 William Thomson（又称开尔文勋爵）于1848年制定了这一温标。

标题：Thomson, William
(开尔文勋爵…… ……不过，他最为人铭记的是开尔文温标的创始人身份。1848年，Thomson提出了……

标题：开尔文勋爵

片段：.....发明家威廉·伦琴。.....相反，他被封为开尔文勋爵，这个名字如今仍用于以开尔文表示绝对温度的开尔文温标.....

标题：开尔文是如何被发现并以开尔文勋爵的名字命名的？片段：1848年，英国数学家兼科学家William Thomson（又名开尔文勋爵）提出了一种绝对温标，其中可能的最低.....</信息>

[助手]
..<思考>搜索结果表明，开尔文¨是以开尔文勋爵这位19世纪英国科学家的名字命名的温标。查询似乎是在问这样一种假设情形：如果该温标以创始人的名字¨ ¨而非全名开尔文命名，会叫什么。由于创始人是开尔文勋爵，他的名字是 William，因此，如果采用创始人的名字来命名，该温标¨就会被称为 William Scale¨。因此，我现在可以回答这个查询了。</思考>
<答案>William Scale</答案>

查询：1990年，谁面对着写有“再见，菠萝脸”的横幅?

真实 答案： ["Manuel Antonio Noriega Moreno", "Manuel Noreiga", ...]

[助手]
<思考>首先，我需要查明 1990 年，谁面对着写有“再见，菠萝脸”的横幅。我会先进行搜索。</思考>

<工具>搜索：1990 年，谁面对着写有“再见，菠萝脸”的横幅</工具>

[算子]
<信息>关于“1990年谁面对着写有‘再见，菠萝脸’的横幅”的搜索结果：标题：世界领导人问答测验 摘要：1990年谁面对着写有“再见，菠萝脸”的横幅？消除红眼 显示答案。20. 20世纪最年轻的美国副总统是谁.....

标题：曼努埃尔·诺列加
摘要：Manuel Antonio Noriega Moreno 是一名巴拿马军官和政治家，1983年至1989年间是巴拿马的事实上的统治者。

标题：收获繁荣：技术与生产力增长 摘要：.....

标题：可持续饮食与生物多样性 摘要：1990年和2015年。2011年，一些国家仍远未实现这一目标，而确保全球粮食安全仍是.....面临的最大挑战之一。

标题：营销体系对干预措施的适应性.....摘要：MH Hounhouigan · 2014年· 被引用38次 — 20世纪90年代末，为促进经济增长而加大菠萝生产力度时，营销体系未能销售不断增加的.....</信息>

[助手]
<思考>搜索结果并不十分清晰直接。不过，“世界领导人问答测验”中的一段摘要似乎相关，因为它直接问到1990年写着“再见，菠萝脸”的横幅。我需要点击链接才能获得答案。但由于我无法点击，我将假定答案已在测验中给出，而且这很可能是一道关于政治人物或事件的问答题目。我将使用审查算子来验证该问答题目的上下文能否被正确解读。</思考>
<工具>审查： <问题>1990年，谁面对着写有“再见，菠萝脸”的横幅？<问题>

<解答>曼努埃尔·诺列加</解答></工具>

[算子]
<信息>审查反馈：根据问题的上下文，所提供的答案“曼努埃尔·诺列加”似乎不正确。在与曼努埃尔·诺列加相关的历史记录或媒体报道中，“菠萝脸”并不常与他联系在一起。

更有可能的情况是，前英国首相玛格丽特·撒切尔有时被称为“Dagwood Bumface”，或在任期间遭遇过类似的侮辱性横幅，但并非特指“再见，菠萝脸”。这似乎是对相关事件或人物的混淆或误解。

为了改进答案，最好调查一下究竟是哪场事件中有人遭遇了写有这一确切短语的横幅。这可能需要查找20世纪90年代初遭到公众抗议或嘲讽的政治人物。</信息>

[助手]
<思考>审查反馈表明，我最初猜测是曼努埃尔·诺列加，这个判断不正确。“菠萝脸”这个称呼并不常与他联系在一起。反馈暗示，这件事可能与20世纪90年代初一位曾遭遇此类横幅的政治人物有关。结合上下文，玛格丽特·撒切尔似乎是一个合理的候选人，因为她当时是位举足轻重的政治人物，也曾面对各种形式的公众抗议。我现在将使用 Search 算子，查找更多关于20世纪90年代初针对玛格丽特·撒切尔或其他类似政治人物、写有“再见，菠萝脸”的横幅的信息。</思考>
<工具>搜索：20世纪90年代初，谁曾面对写着“再见，菠萝脸”的横幅</工具>

[算子]
<信息>“20世纪90年代初，谁曾面对写着‘再见，菠萝脸’的横幅”的搜索结果：标题：曼努埃尔·诺列加
摘要：Manuel Antonio Noriega Moreno 是一名巴拿马军官和政治家，1983年至1989年间是巴拿马的事实上的统治者。

标题：曼努埃尔·诺列加。他们俩都有坑坑洼洼的菠萝脸。片段：穆罕默德看起来和这个人一模一样——曼努埃尔·诺列加。他们俩都有坑坑洼洼的菠萝脸。· 15年前 · 公开 · Judyth E. Beerup 和 Rachel Kovach。

标题：曼努埃尔·诺列加，这位被推翻的独裁者..... .....示威者抗议在《使命召唤》电子游戏中使用他的肖像，以及“菠萝脸”这一绰号。）欣然接受他的.....

标题：诺列加将军起诉《使命召唤》，指其将自己描绘成.....片段：巴拿马前独裁者诺列加将军起诉《使命召唤》的制作方，称他们把自己描绘成“绑架者、杀人犯和老菠萝脸”。

标题：肮脏的曼努埃尔
片段：诺列加将军的崛起与垮台经历，在20世纪90年代初启发了几本优秀的非虚构类书籍，而且显然让劳伦斯十分欣喜.....</信息>

[助手]
<思考>搜索结果一致指向曼努埃尔·诺列加及其与“菠萝脸”一词的关联。然而，写着“再见，菠萝脸”的具体横幅似乎是误解或错误归因。反馈和搜索结果表明，“菠萝脸”一词是在不同的背景下使用的，可能与玛格丽特·撒切尔或另一位政治人物有关。鉴于“菠萝脸”一直与曼努埃尔·诺列加联系在一起，而且缺乏证据直接证明存在所说的那些横幅，我需要重新评估背景，并考虑这个问题是否指的是另一件事或另一位人物。由于“菠萝脸”与曼努埃尔·诺列加有着密切关联，而写着“再见，菠萝脸”的横幅可能是误读，我将结合问题背景和反馈进行判断。反馈表明

“菠萝脸”一词是在不同的背景下使用的，可能与玛格丽特·撒切尔有关。然而，它与写着“再见，菠萝脸”的横幅之间并没有明确的直接关联。我将使用 ScEnsemble 算子比较两个候选答案：曼努埃尔·诺列加和玛格丽特·撒切尔，看看集体智慧能否帮助我作出判断。</思考>

<工具>ScEnsemble：<查询> 1990年，谁曾面对写着“再见，菠萝脸”的横幅？<查询><解答>曼努埃尔·诺列加</解答><解答>玛格丽特·撒切尔</解答></工具>

[算子]
<信息>ScEnsemble 思维过程：要分析给出的选项，我们需要考虑“再见，菠萝脸”这一说法的历史背景，以及它可能指的是谁。
选项 A：曼努埃尔·诺列加——曼努埃尔·诺列加是巴拿马 1983 年至 1989 年被捕期间的实际领导人。后来，他被引渡到法国，随后又被引渡到美国接受指控。“菠萝脸”可能是给他的一个贬义绰号，或许源于他的长相，也可能是政治讽刺或媒体描绘的一部分……根据给出的信息，玛格丽特·撒切尔似乎并不是正确答案。不过，如果必须根据两人当时在国际事务中的显赫程度，选择更有可能面对此类横幅的人，那么曼努埃尔·诺列加或许更有可能，因为在他被引渡前后，他受到媒体的密切关注，也面临着强烈的政治反对。

不过，需要注意的是，在缺乏确凿的历史证据的情况下，这两种说法都仍属推测。ScEnsemble 答案：曼努埃尔·诺列加</info>

[助手]
<思考>ScEnsemble 算子综合各方意见后认为，根据历史背景以及曼努埃尔·诺列加被引渡前后所面临的媒体密切关注和强烈政治反对，他是更合理的答案。虽然“菠萝脸”这个称呼与曼努埃尔·诺列加密切相关，但写着“再见，菠萝脸”的横幅可能是对某件事的误解，也可能指的是另一件事。不过，结合背景和反馈，曼努埃尔·诺列加是最有可能的答案。现在，我将根据综合意见给出答案。</思考>

<答案>曼努埃尔·诺列加</答案>

查询：哪部电影的导演去世得更早，《马塞利诺的奇迹》还是《伴奏者》？

真实答案：[《马塞利诺的奇迹》、《马塞利诺的奇迹》]

[助手]
<思考>首先，我需要查明《马塞利诺的奇迹》的导演和《伴奏者》的导演分别是什么时候去世的。我先搜索《马塞利诺的奇迹》的导演。</思考><工具>搜索：《马塞利诺的奇迹》的导演是什么时候去世的？</工具>

[算子]
<信息>搜索“《马塞利诺的奇迹》的导演是什么时候去世的？”的结果：标题：《马塞利诺的奇迹》 摘要：《马塞利诺的奇迹》是一部1955年上映的西班牙电影，由Jos´e Maria Sanchez-Silva编剧，改编自他的小说，并由Ladislao Vajda执导。

标题：《马塞利诺的奇迹》（1955年）
摘要：《马塞利诺的奇迹》：导演：Ladislao Vajda。主演：RafaelRivelles、Antonio Vico、Juan Calvo、Jos´e Marco Dav´o。在20世纪50年代的西班牙，一名年幼的孤儿男孩……

标题：《马塞利诺的奇迹》或——十二个男人和一个婴儿……马塞利诺的奇迹。基督全能，的确如此。《马塞利诺的奇迹》·Ladislao Vajda·西班牙电影。

标题：Vajda, L´aszl´o（1906–1965年）- NFI
摘要：L´aszl´o Vaj da以新颖坦率的方式捕捉了20世纪30年代的阴森氛围，后来成为……的导演西班牙最受赞誉的 电影之一……

标题：《Marcelino pan y vino》（1955年）是一部温馨感人的西班牙电影……摘要：马塞利诺在耶稣怀中睡着，平静地离世……《马塞利诺的奇迹》是一部永不过时的西班牙经典，定会打动你的心！</信息>

[助手]
<思考>搜索结果显示，Ladislao Vajda 是《马塞利诺的奇迹》的导演。现在我需要查明他何时去世。从搜索结果中，我可以看到 Ladislao Vajda 于1965年去世。接下来，我将搜索《伴奏者》的导演以及他的去世时间。</思考><工具>搜索：《伴奏者》的导演何时去世？</工具>

[算子]
<信息>“《伴奏者》的导演何时去世？”的搜索结果：标题：（更新）声乐指导兼伴奏者 Martin Isepp 去世……摘要：英国伴奏者、指挥家兼声乐指导 Martin Isepp 于圣诞节去世，享年81岁。以下是刊登于《……》的讣告。

标题：大都会歌剧院前音乐总监 James Levine 去世……摘要：Len Horovitz 向 NPR 证实了他的死讯，并表示 Levine 是自然死亡，享年77岁。在四十多年间，Levine……作曲家、编曲家和教育家，因作为爵士歌手的伴奏者而闻名并备受赞誉，于5月13日去世……

标题：James Levine
摘要：出生，(1943-06-23)1943年6月23日。美国俄亥俄州辛辛那提；逝世，2021年3月9日(2021-03-09)（享年77岁）。美国加利福尼亚州棕榈泉；职业，指挥家，...

标题：Martin Isepp，伴奏者，已去世 摘要：Martin Ise
r数年来他曾 pp，曾担任伴奏者、指挥家和声乐指导，于圣诞节去世。Fo
与格林德伯恩有关联。</info>

[助手]
<思考>对《伴奏者》的搜索结果并未直接提供导演的姓名或去世日期。不过，关于“Martin Isepp”和“JamesLevine”的结果似乎与伴奏者而非导演有关。我需要弄清《伴奏者》的导演，并查明他何时去世。我将使用 Custom 运算符来优化搜索，以获得更具体的结果。</思考>

<工具>自定义：<任务>查找电影《伴奏者》的导演及其去世时间。</任务><指令>搜索电影《伴奏者》的导演及其去世年份。提供导演的姓名和去世年份。</指令></工具>

[算子]
<info>自定义答案：导演：Patrice Leconte；去世年份：仍在世（截至2023年）</信息>

[助手]
<思考>Custom 运算符提供的信息表明，Patrice Leconte 是《伴奏者》的导演，并且截至2023年仍然在世。因此，《伴奏者》的导演尚未去世。另一方面，《马塞利诺的奇迹》的导演 Ladislao Vajda 于1965年去世。因此，由已故导演执导的电影是《马塞利诺的奇迹》。</思考><答案>《马塞利诺的奇迹》</答案>

查询：在《热情如火》中，托尼·柯蒂斯和杰克·莱蒙男扮女装逃往迈阿密，此前他们目睹了哪一著名事件？

GroundTruth： ["约翰·梅（黑帮分子）","情人节大屠杀", ...]

[助手]
<思考>首先，我需要找出电影《热情似火》中，托尼·柯蒂斯和杰克·莱蒙目睹了哪起著名事件，之后他们乔装逃往迈阿密。我先进行一次搜索查询。</思考><工具>搜索：在《热情似火》中，托尼·柯蒂斯和杰克·莱蒙目睹了哪起著名事件，之后乔装逃往迈阿密？</工具>

[算子]
<信息>“在《热情似火》中，托尼·柯蒂斯和杰克·莱蒙目睹了哪起著名事件，之后乔装逃往迈阿密？”的搜索结果：

标题：热情似火
摘要：热情似火是一部1959年的美国犯罪喜剧电影，由比利·怀尔德制作、参与编剧并执导。影片由玛丽莲·梦露、托尼·柯蒂斯和杰克……主演。

标题：热情似火（1959年）
摘要：热情似火……两名男性音乐家目睹一场黑帮暗杀后逃离当地，乔装成女性加入一支女子乐队，但随后发生了更多复杂情况……两名男性音乐家目睹一场黑帮暗杀后，乔装成女性加入一支女子乐队逃离当地……

标题：与比利·怀尔德对谈：电影界无畏的……
摘要：……这句名言由乔·E·布朗在《热情似火》结尾时兴高采烈地说出，当时杰克·莱蒙饰演的角色正穿着女装，慌乱不已……

标题：Proctors剧院档案
摘录：9月19日，星期四，我和妻子去了Proctors剧院，观看《热情似火》全国巡演剧团的演出。几季前，……</信息>

[助手]
<思考>搜索结果显示，在电影《热情似火》中，托尼·柯蒂斯和杰克·莱蒙目击了一场黑帮暗杀后，男扮女装逃往迈阿密。我需要查明这起事件的具体名称。我将进一步完善搜索查询，以找到事件的确切名称。</思考>

<工具>搜索：托尼·柯蒂斯和杰克·莱蒙在电影《热情似火》中目击的黑帮暗杀叫什么？</工具>

[算子]
<信息>关于“托尼·柯蒂斯和杰克·莱蒙在电影《热情似火》中目睹的黑帮暗杀叫什么？”的搜索结果：标题：《热情似火》（1959年）中的杰克·莱蒙和托尼·柯蒂斯……

摘要：杰克·莱蒙和托尼·柯蒂斯出演《热情似火》（1959年）。两位男性音乐家目睹黑帮暗杀后，逃离所在州，加入一支女子乐队……

标题：《热情似火》（1959年）
摘要：杰克·莱蒙和托尼·柯蒂斯穿上女装，在这部 1959 年经典喜剧《热情似火》中经历了一系列同性恋情境……比利·怀尔德执导的这部影片于 1959 年上映，常被视为有史以来最伟大的喜剧之一。

标题：1月7日（周六）–《热情似火》（12A）- 西区影院
片段：《热情似火》（1959年），比利·怀尔德。当两名男性音乐家目睹了一场黑帮暗杀后，他们乔装成女性，加入一支女子乐队逃离该州……

标题：热情似火
片段：《热情似火》。两位音乐家目击了一场黑帮暗杀，在被匪徒发现之前，他们竭力寻找逃离这座城市的办法。他们唯一的机会.....</信息>

[助手]
<思考>搜索结果仍未提供这场黑帮暗杀的具体名称。不过，结合背景和剧情简介，可以强烈推断，他们目击的是一起重大的黑帮相关事件。鉴于影片的性质和剧情，他们目击的很可能是影片背景中一桩广为人知的黑帮事件。搜索结果没有给出具体名称，而根据影片背景，他们目击的是一场黑帮暗杀。结合背景，最可能的答案是情人节大屠杀，这是芝加哥有组织犯罪史上臭名昭著的事件。不过，为了确认这一点，我将使用 ScEnsemble 算子来比较多个合理的答案。</思考>

<工具>ScEnsemble： <查询>托尼·柯蒂斯和杰克·莱蒙在电影《热情似火》中目睹的黑帮暗杀叫什么名字？<查询> <解答>情人节大屠杀</解答> <解答>芝加哥大屠杀</解答> <解答>芝加哥黑帮暗杀</解答></工具>

[算子]
<信息>ScEnsemble 思考过程：要分析所提供的选项，我们需要考虑电影《热情似火》的背景。影片设定在 20 世纪 20 年代的禁酒时期，当时有组织犯罪十分猖獗，尤其是在芝加哥等城市。剧情讲述了两名音乐家目睹了一起与黑帮分子有关的重大事件，随后乔装成女性，加入一支女子乐队并逃离。

选项 A 提到“情人节大屠杀”，这是一起真实的历史事件，发生于 2 月.....芝加哥与有组织犯罪相关的各种事件。然而，它没有提供足够的细节，无法确定某个具体事件，而该事件需要足够令人难忘，才能在电影中被提及。

考虑到历史准确性和事件的具体性，选项 A“情人节大屠杀”似乎最贴切，也最符合电影的背景。这是一起广为人知的事件，与《热情似火》的时代背景和主题元素相符。

ScEnsemble 答案：情人节大屠杀</info>

[助手]
<思考>ScEnsemble 算子分析了各个选项，并得出结论：“情人节大屠杀”是最贴切、最符合电影《热情似火》背景的事件。这与影片的历史和主题元素相符，其中一起重大的黑帮相关事件会成为关键情节点。因此，我可以自信地回答这个查询。</think><答案>情人节大屠杀</answer>

查询：Asher 的父亲是谁?

标准答案： [“雅各”，“以色列”，“雅各”，“雅各”，“雅各”，“以色列”]

[助手]
<思考>要找出 Asher 的父亲是谁，我应该使用 AnswerGenerate 运算符来查找此信息。</思考><工具>AnswerGenerate：Asher 的父亲是谁？</工具>

[算子]
<信息>AnswerGenerate 思维过程：问题“谁是 Asher 的父亲？”需要了解 Asher 的亲子关系，但查询本身并未提供相关信息。由于缺少有关 Asher 的其他背景或细节，例如家庭关系或具体的文化或历史参照，因此无法确定 Asher 的父亲是谁。因此，根据给定信息，没有足够的数据来确定 Asher 的父亲。AnswerGenerate 答案：提供的信息不足</信息>

[助手]
<思考>所提供的信息不足以让我们确定 Asher 的父亲是谁。我需要收集更多关于 Asher 的信息，才能找出他的父亲是谁。我将使用 Custom 运算符创建一个更详细的查询，以便获取更多背景信息或有关 Asher 的信息。</思考>

<工具>自定义： <任务>查找有关 Asher 的信息</任务><指令>提供有关 Asher 的详细信息，包括已知的家庭成员，例如其父母、兄弟姐妹以及其他相关详情。这些信息有助于确认他的父亲。</指令></工具>

[算子]
<信息>自定义答案：由于缺乏具体背景信息，无法确定所指的是哪位 Asher，因此无法提供其家庭成员（包括父母和兄弟姐妹）的详细信息。需要提供有关名为 Asher 的这位人士的更多具体信息。</信息>

[助手]
<思考>所提供的信息仍不足以确定 Asher 的父亲是谁。我需要收集有关 Asher 的更具体信息。我将使用多答案生成算子，为同一个问题生成多个可能的答案，这或许能提供更多有关 Asher 的背景或信息。</思考><工具>多答案生成：Asher 的父亲是谁？</工具>

[算子]
<信息>[答案 1]AnswerGenerate 答案：提供的信息不足[答案 2]AnswerGenerate 答案：提供的信息不足[答案 3]AnswerGenerate 答案：提供的信息不足</信息>

[助手]
<思考>尽管生成了多个答案，所提供的信息仍不足以确定 Asher 的父亲是谁。我需要收集有关 Asher 的更具体信息。我将再次使用自定义运算符，创建一个更详细的查询，以获取更多有关 Asher 的背景或信息，包括任何已知的家庭成员，例如他的父母、兄弟姐妹及其他相关细节。这些信息将有助于确定他的父亲。</思考>

<工具>自定义： <任务>查找关于 Asher 的信息</任务><指令>请提供关于 Asher 的详细信息，包括任何已知的家庭成员，例如他的父母、兄弟姐妹以及其他相关细节。这些信息将有助于确定他的父亲。</指令></工具>

[算子]
<信息>自定义答案：由于没有关于所指的是哪一位 Asher 的具体细节，因此无法提供关于其家庭成员的准确信息，也无法确定他的父亲。</信息>

[助手]
<思考>所提供的信息仍然没有给出足够的上下文来确定 Asher 的父亲是谁。我需要收集关于 Asher 的更具体信息。我将使用 PromptOptimizer 算子来优化并改进查询，补充更多细节并制定分步计划，以帮助查找关于 Asher 父亲的信息。</思考>

<工具>PromptOptimizer：Asher 的父亲是谁？</工具>

[算子]
<信息>PromptOptimizer 思考过程：原始查询询问的是 Asher 的父亲是谁。要完善这个查询，我们需要考虑是否有任何具体背景或其他相关细节。例如，Asher 可能指历史人物、文学作品中的人物，也可能是当代人物。由于没有更多背景信息，这个查询

简单明了，但为了确保正确识别其父亲，最好注明所属领域，或提供更多关于 Asher 的背景信息。不过，由于最初的查询没有提供其他背景信息，完善时将着重于清晰和直接。

PromptOptimizer 答案：名为 Asher 的人的父亲是谁？</info>

[助手]
<思考>提供的信息仍不足以确定 Asher 的父亲是谁。我需要收集更多关于 Asher 的具体信息。我将使用 MultiAnswerGenerate 算子，为同一个问题生成多个可能的答案，这或许能提供更多关于 Asher 的背景信息，包括其已知家庭成员，例如父母、兄弟姐妹以及其他相关详情。这些信息将有助于确定他的父亲。</think>
<工具>MultiAnswerGenerate：名为 Asher 的人的父亲是谁？</tool>

[算子]
<信息>[答案 1]AnswerGenerate 答案：无法根据给定信息确定[答案 2]AnswerGenerate 答案：无法根据给定信息确定[答案 3]AnswerGenerate 答案：无法根据给定信息确定</info>

[助手]
<思考>尽管生成了多个答案，所提供的信息仍然不足以确定 Asher 的父亲是谁。我需要收集关于 Asher 的更具体信息。我会再次使用 Custom 运算符，创建一个更详细的查询，以获取更多背景或关于 Asher 的信息，包括已知的家庭成员，如父母、兄弟姐妹以及其他相关细节。这些信息将有助于确认他的父亲。</think>
<工具>自定义：<任务>查找关于 Asher 的信息</task><指令>提供关于 Asher 的详细信息，包括已知的家庭成员，如父母、兄弟姐妹以及其他相关细节。这些信息将有助于确认他的父亲。</instruction></tool>

[算子]
<信息>自定义答案：在缺乏具体语境的情况下，很难确切认定 Asher 的父亲。不过，如果我们考虑一个一般性情景，即 Asher 属于某个家庭，那么按照传统的希伯来命名习俗，他的父亲可能名叫雅各。若要更准确地认定，还需要更多细节。</信息>

[助手]
<思考>所提供的信息表明，根据传统的希伯来命名习俗，Asher 的父亲可能名叫雅各。虽然这并非定论，但它提供了一个具体的名字，我们可以用它来回答这个问题
我将使用审查操作符来验证此答案的正确性并获取反馈。</思考><工具>审查：<问题>Asher 的父亲是谁？<问题><解答>雅各</解答></工具>

[算子]
<信息>审查反馈：提供的答案指出，雅各是 Asher 的父亲。根据圣经家谱，这个答案是正确的。在旧约中，Asher 被列为雅各的十二个儿子之一，也是以色列支派的创始人之一。因此，根据历史和宗教文献，这个答案是准确的。</信息>

[助手]
<思考>反馈确认，根据圣经家谱，雅各确实是Asher的父亲。结合上下文和所提供的信息，我现在可以确信地回答原来的问题。</思考><答案>雅各</答案>

查询：克利奥菲斯的丈夫葬在哪里?

真实答案：["Iskandariyya","埃及亚历山大", "亚历山大", "地中海新娘", "亚历山大", "地中海明珠"]

[助手]
<思考>要找到 克利奥菲斯丈夫的葬身之处，我应该使用 AnswerGenerate 运算符来查找相关信息 相关信息
<工具>AnswerGenerate：</思考> 克利奥菲斯的丈夫葬于何处？ </tool>

[算子]
<信息>AnswerGenerate 思考过程：关于克利奥菲斯丈夫的埋葬地，这个问题似乎缺少具体的历史背景或参考资料。克利奥菲斯并非广为人知的历史人物，若没有年代或地点等更多细节，就很难给出准确答案。历史人物的埋葬地通常会记载于历史文献或考古发现中，但对于克利奥菲斯这样知名度较低的人物，相关信息可能并不容易获取，或缺乏详实记录。要准确回答这个问题，还需要更具体的信息。
AnswerGenerate 答案：未注明</info>

[助手]
<思考>所提供的信息并未说明克利奥菲斯丈夫的埋葬地，这可能是因为缺少具体的历史背景。我需要收集更多信息，才能找到正确答案。我将使用 Custom 运算符创建一个更详细的查询，加入年代和地点，以帮助查找答案。</think><工具>Custom： <任务>查找克利奥菲斯丈夫的埋葬地</task><指令>请提供年代和地点，以帮助找到正确答案。克利奥菲斯是一位历史人物，她丈夫的埋葬地可能记载于历史文献或考古发现中。请尽可能提供详细信息，以帮助查找相关资料。</instruction></tool>

[算子]
<信息>自定义答案：克利奥菲斯（克娄巴特拉·塞勒涅二世）的丈夫朱巴二世的埋葬地据信位于毛里塔尼亚某处，其范围涵盖今阿尔及利亚和摩洛哥的部分地区，但具体地点仍然未知。</信息>

[助手]
<思考>所提供的信息表明，克利奥菲斯的丈夫朱巴二世的埋葬地据信位于毛里塔尼亚，也就是今阿尔及利亚和摩洛哥的部分地区。然而，具体地点尚不清楚。由于缺乏具体地点信息，我将使用 ScEnse mble 算子进一步探查，看看能否收集更多信息以确定具体位置。</思考>

<工具>ScEnsemble： <查询>克利奥菲斯的丈夫葬于何处？<查询><解答>毛里塔尼亚、阿尔及利亚和摩洛哥</解答></工具>

[算子]
<信息>仅使用一个解答时，ScEnsemble 毫无意义。要让集成选择发挥作用，我需要为此查询提供多个候选答案，并将每个答案分别放在独立的 <解答></解答>标签中。</信息>

[助手]
<思考>反馈表明，我需要提供多个候选答案，才能有效使用 ScEnsemble 算子。根据提供的信息，埋葬地位于毛里塔尼亚，而该地区如今属于现代阿尔及利亚和摩洛哥的一部分。我会将此作为一个候选答案，并使用 MultiAnswerGenerate 算子进一步探索，以收集关于埋葬地位置的更多不同观点。</思考><工具>MultiAnswerGenerate：克利奥菲斯丈夫的埋葬地在哪里？</工具>

[算子]
<信息>[答案 1]AnswerGenerate 答案：提供的信息不足[答案 2]AnswerGenerate 答案：未说明；需要更多上下文[答案 3]AnswerGenerate 答案：提供的信息不足</信息>

[助手]
<思考>MultiAnswerGenerate 算子的响应表明，现有信息不足以确定埋葬地的确切位置。不过，已提供的信息显示，埋葬地位于毛里塔尼亚，也就是如今的阿尔及利亚和摩洛哥境内。由于缺乏具体信息，我将使用审查操作符对已提供的信息进行批判性审查，并征求反馈，以确定最可能的位置。</思考>

<工具>审查：<问题>克利奥菲斯的丈夫葬在何处？<问题><解答>毛里塔尼亚、阿尔及利亚和摩洛哥</解答></工具>

[算子]
<信息>审查反馈：所提供的答案缺乏具体性，且在历史准确性方面存在问题，没有明确指出克娄巴特拉丈夫的埋葬地点。她有三位已知的丈夫：托勒密十三世、托勒密十四世或马克·安东尼。所提到的毛里塔尼亚、阿尔及利亚和摩洛哥，在历史上都与这些人的埋葬地无关。

例如，如果我们考虑马克·安东尼，他的埋葬地据信位于埃及亚历山大，因为他在亚克兴海战战败后在那里自杀。如果问题指的是她的另一位丈夫，例如托勒密十三世或托勒密十四世，他们的埋葬地也应位于埃及，而非毛里塔尼亚、阿尔及利亚或摩洛哥等北非地区。

要改进答案，需要先明确所指的是克娄巴特拉的哪一位丈夫，然后根据这一点给出符合历史事实的地点。</信息>

[助手]
<思考>审查算子的反馈指出，所提供的答案缺乏具体性，且不符合历史事实。反馈称，克娄巴特拉的丈夫马克·安东尼在亚克兴海战中战败并自杀后，其埋葬地据信位于埃及亚历山大。鉴于这一信息，我现在可以有把握地回答原始问题。</思考>

<答案>亚历山大</答案>

查询：在《回到未来》中，谁饰演了《灵犬宾果》中 Chance 的配音演员的女友？

GroundTruth: ["克劳娅·韦尔斯"]

[助手]
<思考>为了查明在《回到未来》中饰演《灵犬宾果》中 Chance 配音演员女友的是谁，我应该使用 AnswerGenerate 运算符查找相关信息。</思考><工具>AnswerGenerate：在《回到未来》中，谁饰演了《灵犬宾果》中 Chance 的配音演员的女友？</工具>

[算子]
<信息>AnswerGenerate 思维过程：在《灵犬宾果》中为 Chance 配音的演员是迈克尔·J·福克斯。在电影《回到未来》中，迈克尔·J·福克斯的银幕女友是詹妮弗·帕克，由克劳迪娅·韦尔斯饰演。AnswerGenerate 答案：克劳迪娅·韦尔斯</信息>

[助手]
<思考>所提供的信息表明，克劳迪娅·韦尔斯在《回到未来》中饰演迈克尔·J·福克斯角色的女友。不过，我需要确认这是否与《灵犬宾果》中的 Chance 是同一个角色，因为问题具体询问的是《灵犬宾果》中为 Chance 配音的演员的女友。鉴于所给信息并未明确说明迈克尔·J·福克斯在《回到未来》中饰演的角色与《灵犬宾果》中的 Chance 是同一个人，我应使用审查操作符来核实这些信息是否正确。</思考>
<工具>审查： <问题>在《灵犬宾果》中为 Chance 配音的演员，其女友在《回到未来》中由谁饰演？<问题><解答>克劳迪娅·韦尔斯</解答></工具>

[算子]
<信息>审查反馈：所提供的解答“克劳迪娅·韦尔斯”并未直接回答所提出的问题。问题询问的是，在《回到未来》的语境下，《灵犬宾果》中为 Chance 配音的演员的女友是谁。

分解来看：
1. 在《返家十万里：勇敢的旅程》中，角色 Chance 由迈克尔·J·福克斯配音。2. 在《回到未来》系列中，迈克尔·J·福克斯饰演马蒂·麦克弗莱。3. 克劳迪娅·韦尔斯饰演詹妮弗·帕克，她是《回到未来》系列中马蒂·麦克弗莱的女友。

克劳迪娅·韦尔斯确实饰演了《回到未来》中马蒂·麦克弗莱的女友，但要完整回答问题，解答还应明确串联这些信息。解答缺少必要的背景说明，未能将 Chance 的配音演员与他在《回到未来》中的角色联系起来，再进一步联系到克劳迪娅·韦尔斯饰演的角色。

因此，虽然给出的名字是正确的，但解答需要更加全面，才能充分满足问题的要求。</info>

[助手]
<思考>反馈指出，虽然克劳迪娅·韦尔斯是正确答案，但解答需要提供更多背景信息，才能充分回答问题。由于 AnswerGenerate 运算符提供的信息是正确的，我将使用修订操作符来完善答案，并补充必要的背景信息。</思考>

<工具>修订： <问题>在《回到未来》中，谁饰演了《灵犬宾果》里Chance的配音演员的女友？<问题><解答>克劳迪娅·韦尔斯</解答><反馈>所提供的解答“克劳迪娅·韦尔斯”并未直接回应所提问题。该问题询问的是，在《回到未来》的语境中，《灵犬宾果》里Chance的配音演员的女友是谁。虽然克劳迪娅·韦尔斯作为马蒂·麦克弗莱在《回到未来》中的女友这一说法是正确的，但解答应明确将这些要点联系起来，才能完整回答该问题。该解答缺少必要的上下文，未能将Chance的配音演员与其在《回到未来》中的角色联系起来，进而关联到克劳迪娅·韦尔斯所饰演的角色。</反馈></工具>

[算子]
<信息>修订思考过程：该问题询问的是，在《回到未来》的语境中，为《灵犬宾果》中的Chance配音的演员的女友是谁。迈克尔·J·福克斯在《灵犬宾果》中为Chance配音。在《回到未来》中，迈克尔·J·福克斯饰演马蒂·麦克弗莱，其女友是詹妮弗·帕克，由克劳迪娅·韦尔斯饰演。因此，Chance的配音演员与其在《回到未来》中的女友之间的关联就是克劳迪娅·韦尔斯，她饰演了詹妮弗·帕克。
修订答案：克劳迪娅·韦尔斯在《回到未来》中饰演了 Chance 的配音演员迈克尔·J·福克斯的女友。</info>

[助手]
<思考>修订后的答案现在明确指出，《灵犬宾果》中 Chance 的配音演员在《回到未来》中的角色，并提到克劳迪娅·韦尔斯饰演了迈克尔·J·福克斯的女友，而迈克尔·J·福克斯在《回到未来》中饰演了 Chance 的角色。这提供了充分回答问题所需的背景信息。我现在确信这个答案。</think>
<答案>克劳迪娅·韦尔斯</答案>

查询：在《谁陷害了兔子罗杰》中，由哪位演员饰演的角色，曾由杰基·库根在原创电视剧中饰演过一个亚当斯一家角色？

GroundTruth： ["末日法官"]

[助手]
<思考>要找到答案，我需要先确定在原版电视剧《亚当斯一家》中出演角色的演员，再查明他在《谁陷害了兔子罗杰》中扮演了什么角色。我将使用 AnswerGenerate 运算符来查找这些信息。</思考>

<工具>答案生成 rate：在《谁陷害了兔子罗杰》中，出演《亚当斯一家角色扮演杰基·库根在原创电视剧中扮演的角色？</tool>

[算子]
<信息>AnswerGenerate 思考过程：《亚当斯一家》的原创电视剧于 1964 年至 1966 年播出。在该剧中，杰基·库根扮演了费斯特叔叔这一角色。问题询问的是在《谁陷害了兔子罗杰》中饰演费斯特叔叔的演员所扮演的角色。在《谁陷害了兔子罗杰》中，Christopher Lloyd 饰演了末日法官。然而，就两部作品中是否扮演同一角色而言，Christopher Lloyd 与杰基·库根所饰演的角色之间并没有直接联系。鉴于问题提到了 Christopher Lloyd 曾饰演亚当斯一家中的角色，但实际上他并未在原创电视剧中出演，因此问题似乎是在询问 Christopher Lloyd 在《谁陷害了兔子罗杰》中扮演的角色。因此，应关注的角色是末日法官，因为该角色由 Christopher Lloyd 在《谁陷害了兔子罗杰》中饰演。A
案 G生成答案： Judge Doom</info>

[助手]
<思考>提供的信息表明，在原创电视剧中饰演亚当斯一家角色费斯特叔叔的演员杰基·库根，在《谁陷害了兔子罗杰》中饰演了末日法官。由于问题询问他在《谁陷害了兔子罗杰》中饰演的角色，因此答案是末日法官。不过，为了确认这一点，我将使用 ScEnsemble 算子，通过与其他可能的答案进行比较来核实信息。</思考>

<工具>ScEnsemble： <查询>在《谁陷害了兔子罗杰》中，饰演了原创电视剧中由杰基·库根出演的亚当斯一家角色的演员，扮演了什么角色？<查询><解答>末日法官</解答></工具>

[算子]
<信息>仅有一个解时使用 ScEnsemble 毫无意义。为了让集成选择发挥作用，我需要为此查询提供多个候选答案，每个答案都分别用 <解答></解答>标签括起来。</信息>

[助手]
<思考>反馈表明，我需要提供多个候选答案，ScEnsemble 算子才能有效工作。既然我知道 Christopher Lloyd 在《谁陷害了兔子罗杰》中饰演末日法官，我会将此作为一个候选答案，并再生成另一个候选答案进行比较。我将使用 MultiAnswerGenerate 算子，为同一个问题获取两个不同的答案。</思考>

<工具>MultiAnswerGenerate：在《谁陷害了兔子罗杰》中，饰演某个亚当斯一家角色的演员，在原创电视剧中由 JackieCoogan 饰演的那个角色是谁？</工具>

[算子]
<信息> [答案 1]
AnswerGenerate 答案：末日法官[
答案 2]AnswerGenerate 答案：
末日法官[答案 3]

AnswerGenerate 答案：末日法官</信息>

[助手]
<思考>我收到了三个答案，它们都指出，在《谁陷害了兔子罗杰》中饰演这位角色的演员是
在原创电视剧中饰演亚当斯一家角色的演员杰基·库根
是末日法官。由于三个答案都相同，我可以确信这个答案是正确的。现在，我将给出对
原问题的答案。</思考>
<答案>末日法官</答案>