

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



Deep Research Agents

Tasks, Benchmarks, and Methods

Akari Asai

Language Technologies Institute
Carnegie Mellon University

11-768: AI Agents Fall 2026

<https://akariasai.github.io/> | aasai@andrew.cmu.edu

深度研究智能体

任务、基准与方法

Akari Asai · 语言技术研究所 · 卡内基梅隆大学 · 11-768: AI 智能体 · 2026年秋季

<https://akariasai.github.io/> · aasai@andrew.cmu.edu

深度研究智能体：任务、基准与方法

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01

What are Deep Research Agents?

01

什么是深度研究智能体？

什么是深度研究智能体？

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01 / DEEP RESEARCH AGENTS

A simple question needs just one search

What is Akari Asai’s office number at CMU?

3

01 / 深度研究智能体

一个简单问题只需搜索一次。

Akari Asai 在 CMU 的办公室电话号码是多少？

一个简单问题只需搜索一次

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01 / DEEP RESEARCH AGENTS

A simple question needs just one search

What is Akari Asai’s office number at CMU?

Search

Carnegie Mellon University

Language Technologies Institute

School of Computer Science

AcademicsResearchPeopleNews and EventsAbout LTIApply

School of Computer Science > Language Technologies Institute > People > Faculty > Akari Asai



Akari Asai

Assistant Professor (Beginning Fall 2026),
Language Technologies Institute

5507 Gates & Hillman Centers

Address

5500 Forbes Avenue
Pittsburgh, PA 15213

Contact

THE RELEVANT EVIDENCE

5507 Gates & Hillman Centers

CMU LTI faculty profile, captured September 16, 2026.

01 / 深度研究智能体

一个简单问题只需搜索一次。

Akari Asai 在 CMU 的办公室电话号码是多少？

搜索

相关证据

来源：CMU LTI 教师个人资料，截取于2026年9月16日。

一个简单问题只需搜索一次

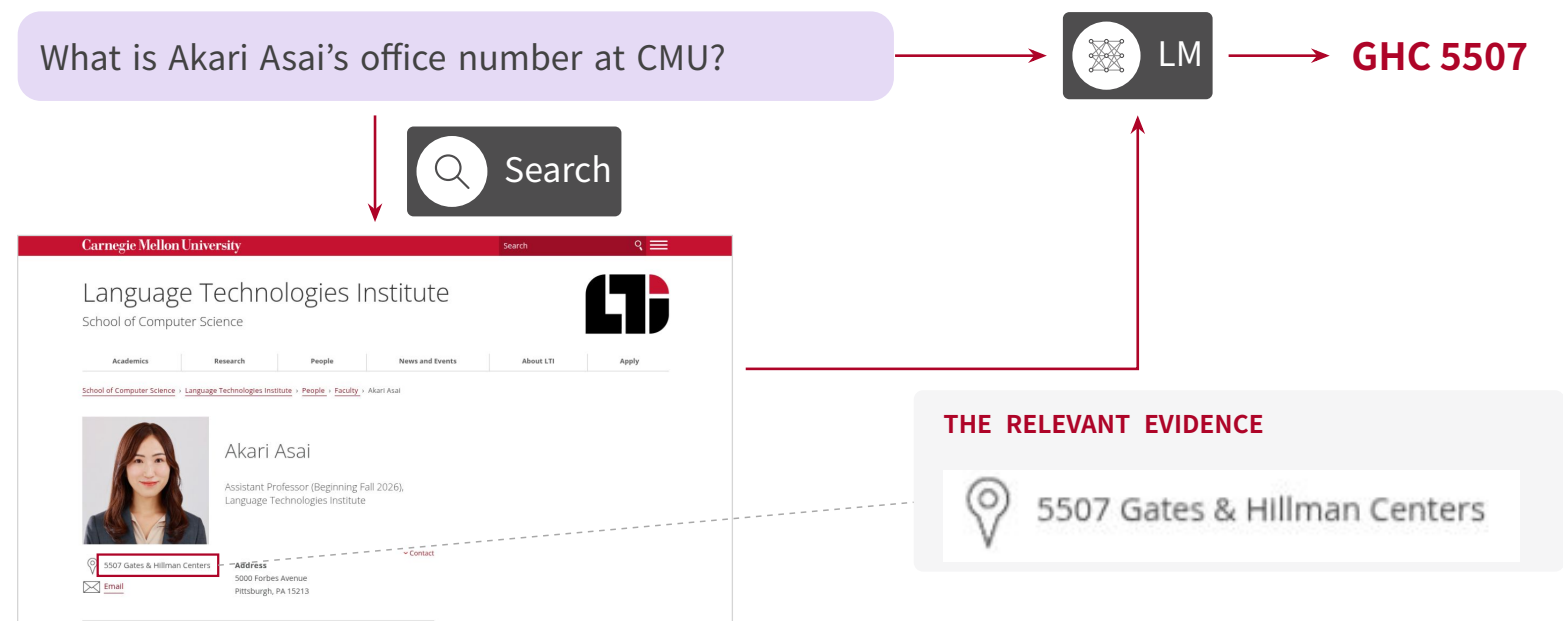
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01 / DEEP RESEARCH AGENTS

A simple question needs just one search



CMU LTI faculty profile, captured September 16, 2026.

3

一个简单问题只需搜索一次

01 / 深度研究智能体

一个简单问题只需搜索一次。

Akari Asai 在 CMU 的办公室电话号码是多少?

搜索

语言模型: GHC 5507

相关证据

来源：CMU LTI 教师个人资料，截取于2026年9月16日。

01 / DEEP RESEARCH AGENTS

Complex research needs many searches

Can AI agents synthesize scientific literature as well as human experts?



Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

复杂研究需要多次搜索

01 / 深度研究智能体

复杂研究需要多次搜索。

AI 智能体能否像人类专家一样综合科学文献？

左侧论文：《用检索增强语言模型合成科学文献》（OpenScholar），Nature 2026。

OpenScholar 摘要：科学进步取决于研究者综合不断增长的文献的能力。大型语言模型能否帮助科学家完成这项工作？OpenScholar 是一种专用的检索增强语言模型，它从 4,500 万篇开放获取论文中识别相关段落，并综合生成带引用的回答。

为评估 OpenScholar，作者构建 ScholarQABench：首个大规模多领域文献检索基准，包含 2,967 个专家撰写的问题和 208 篇长答案，覆盖计算机科学、物理学、神经科学和生物医学。

论文报告：在 ScholarQABench 上，OpenScholar-8B 的正确性比 GPT-4o 高 5%，比 PaperQA2 高 7%，尽管它是更小的开放模型。GPT-4o 有 78%–90% 的概率会捏造引文，而 OpenScholar 的引文准确度与人类专家相当。

OpenScholar 的数据存储、检索器和自我反馈推理循环也能提升现成语言模型；例如，OpenScholar-GPT4o 将 GPT-4o 的正确性提高了 12%。在人类评估中，专家分别有 51% 和 70% 的时候更偏好 OpenScholar-8B 与 OpenScholar-GPT4o 的回答，而 GPT-4o 为 32%。作者开放发布了全部代码、模型、数据存储、数据和公共演示。

OpenScholar 引言指出，综合科学文献有助于发现研究方向、改进方法并支持循证决策；但论文数量增长使研究者难以及时掌握文献。有效综合需要精准检索、准确归属和实时获取最新文献。大型语言模型虽有潜力，却面临幻觉、预训练数据过时和归因不透明等问题；作者的实验中，GPT-4 在计算机科学、生物医学等领域引用最新文献时，78%–90% 的引文系捏造。

右侧论文：《DR Tulu：用不断演进的评分标准进行强化学习，以训练深度研究》，ICML 2026。

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 2/2 部分

01 / DEEP RESEARCH AGENTS

Complex research needs many searches

Can AI agents synthesize scientific literature as well as human experts?



Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

复杂研究需要多次搜索

DR Tulu 摘要：深度研究智能体通过多步研究生成长篇且归属明确的回答。不过，多数开放深度研究智能体通过强化学习与可验证奖励，在容易验证的短答案问答任务上训练；这种做法无法扩展到真实的长篇任务。

该论文提出“不断演进评分标准强化学习”（RLER）：在训练期间构建并维护评分标准，使其与策略模型共同演进。评分标准可以纳入搜索中新发现的信息，并对照不同模型的回答，从而改进事实核查并提供更具区分力的 on-policy 反馈。

基于 RLER，作者开发了 DR Tulu-8B，这是首个直接针对开放式长篇深度研究训练的完全开放模型。在科学、医疗和通用领域的四个长篇深度研究基准上，DR Tulu 平均比 Tongyi DR 高 15.6%，平均与 OpenAI DR 持平或高 0.7%；模型规模更小，每次查询成本约低 1,000 倍。

图 1：深度研究模型的性能与成本。图中比较四个长篇深度研究基准（ScholarQA-CSv2、HealthBench、ResearchQA、DeepResearch-Bench）的平均性能，以及在 ScholarQA-CSv2 上每次查询的推理成本（美元）；图注称 DR Tulu-8B 位于 Pareto 前沿，优于更大的开放模型，并达到专有模型的水平。

DR Tulu 引言指出，深度研究智能体通过规划、搜索并综合多种来源，为复杂任务生成深入且引用充分的回答。现有开放智能体要么不经训练、使用手工提示和现成模型，要么在搜索密集但受约束的短答案任务上使用带可验证奖励的强化学习（RLVR）。开放式深度研究的强化学习依赖可靠奖励，但优质回答的标准往往定义不足，难以用静态的预定义标准完整刻画；准确评估还常需超出模型参数知识范围的、广泛且最新的外部信息。

论文题名与出处：Synthesizing scientific literature with retrieval-augmented language models. Nature 2026. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

两篇论文的首页面截图以很小字号嵌入幻灯片。此处据可辨认的论文标题、摘要、部分引言和图注翻译；截图中的完整作者/机构名单、链接文字、参考文献和其余微小正文无法可靠逐字核对，不能据此声称已转录论文全文。

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/2 部分

01 / DEEP RESEARCH AGENTS

Complex research needs many searches

Can AI agents synthesize scientific literature as well as human experts?



Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

复杂研究需要多次搜索

01 / 深度研究智能体

复杂研究需要多次搜索。

AI 智能体能否像人类专家一样综合科学文献？

没有任何单一结果能够解决这个问题。

左侧论文：《用检索增强语言模型合成科学文献》（OpenScholar）, Nature 2026。

OpenScholar 摘要：科学进步取决于研究者综合不断增长的文献的能力。大型语言模型能否帮助科学家完成这项工作？OpenScholar 是一种专用的检索增强语言模型，它从 4,500 万篇开放获取论文中识别相关段落，并综合生成带引用的回答。

为评估 OpenScholar，作者构建 ScholarQABench：首个大规模多领域文献检索基准，包含 2,967 个专家撰写的问题和 208 篇长答案，覆盖计算机科学、物理学、神经科学和生物医学。

论文报告：在 ScholarQABench 上，OpenScholar-8B 的正确性比 GPT-4o 高 5%，比 PaperQA2 高 7%，尽管它是更小的开放模型。GPT-4o 有 78%–90% 的概率会捏造引文，而 OpenScholar 的引文准确度与人类专家相当。

OpenScholar 的数据存储、检索器和自我反馈推理循环也能提升现成语言模型；例如，OpenScholar-GPT4o 将 GPT-4o 的正确性提高了 12%。在人类评估中，专家分别有 51% 和 70% 的时候更偏好 OpenScholar-8B 与 OpenScholar-GPT4o 的回答，而 GPT-4o 为 32%。作者开放发布了全部代码、模型、数据存储、数据和公共演示。

OpenScholar 引言指出，综合科学文献有助于发现研究方向、改进方法并支持循证决策；但论文数量增长使研究者难以及时掌握文献。有效综合需要精准检索、准确归属和实时获取最新文献。大型语言模型虽有潜力，却面临幻觉、预训练数据过时和归因不透明等问题；作者的实验中，GPT-4 在计算机科学、生物医学等领域引用最新文献时，78%–90% 的引文系捏造。

右侧论文：《DR Tulu：用不断演进的评分标准进行强化学习，以训练深度研究》，ICML 2026。

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 2/2 部分

01 / DEEP RESEARCH AGENTS

Complex research needs many searches

Can AI agents synthesize scientific literature as well as human experts?



Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

复杂研究需要多次搜索

DR Tulu 摘要：深度研究智能体通过多步研究生成长篇且归属明确的回答。不过，多数开放深度研究智能体通过强化学习与可验证奖励，在容易验证的短答案问答任务上训练；这种做法无法扩展到真实的长篇任务。

该论文提出“不断演进评分标准强化学习”（RLER）：在训练期间构建并维护评分标准，使其与策略模型共同演进。评分标准可以纳入搜索中新发现的信息，并对照不同模型的回答，从而改进事实核查并提供更具区分力的 on-policy 反馈。

基于 RLER，作者开发了 DR Tulu-8B，这是首个直接针对开放式长篇深度研究训练的完全开放模型。在科学、医疗和通用领域的四个长篇深度研究基准上，DR Tulu 平均比 Tongyi DR 高 15.6%，平均与 OpenAI DR 持平或高 0.7%；模型规模更小，每次查询成本约低 1,000 倍。

图 1：深度研究模型的性能与成本。图中比较四个长篇深度研究基准（ScholarQA-CSv2、HealthBench、ResearchQA、DeepResearch-Bench）的平均性能，以及在 ScholarQA-CSv2 上每次查询的推理成本（美元）；图注称 DR Tulu-8B 位于 Pareto 前沿，优于更大的开放模型，并达到专有模型的水平。

DR Tulu 引言指出，深度研究智能体通过规划、搜索并综合多种来源，为复杂任务生成深入且引用充分的回答。现有开放智能体要么不经训练、使用手工提示和现成模型，要么在搜索密集但受约束的短答案任务上使用带可验证奖励的强化学习（RLVR）。开放式深度研究的强化学习依赖可靠奖励，但优质回答的标准往往定义不足，难以用静态的预定义标准完整刻画；准确评估还常需超出模型参数知识范围的、广泛且最新的外部信息。

论文题名与出处：Synthesizing scientific literature with retrieval-augmented language models. Nature 2026. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

本页延续上一页的两篇论文截图与文字内容，另叠加提示语“没有任何单一结果能够解决这个问题”。论文截图字号很小，微小作者/机构名单、链接、参考文献和其余正文无法可靠逐字核对。

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01 / DEEP RESEARCH AGENTS

Complex research needs many searches

Can AI agents synthesize scientific literature as well as human experts?



Plan

- I need to separate strong benchmark results from direct comparisons with human experts.
- 1. Find studies of literature synthesis.
 - 2. Check how experts and agents were compared.
 - 3. Compare tasks and limits before drawing a conclusion.

Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

01 / 深度研究智能体

复杂研究需要多次搜索。

AI 智能体能否像人类专家一样综合科学文献？

计划：我需要区分基准上的强劲结果与直接和人类专家比较的结果。1. 查找有关文献综合的研究。2. 检查专家与智能体的比较方式。3. 得出结论前比较任务与限制。

Synthesizing scientific literature with retrieval-augmented language models. Nature 2026. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

复杂研究需要多次搜索

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01 / DEEP RESEARCH AGENTS

Complex research needs many searches

Can AI agents synthesize scientific literature as well as human experts?

🔍 Search
AI agent literature synthesis human expert evaluation

Result A: a research benchmark
Reports answer scores against other AI systems.

Result B: an expert evaluation
Compares agent answers with human-written answers.

Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

4

01 / 深度研究智能体

复杂研究需要多次搜索。

AI 智能体能否像人类专家一样综合科学文献？

搜索：AI 智能体 文献综合 人类专家 评测。

结果 A：研究基准——报告与其他 AI 系统相比的回答得分。

结果 B：专家评测——比较智能体回答与人类撰写的回答。

Synthesizing scientific literature with retrieval-augmented language models. Nature 2026. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

复杂研究需要多次搜索

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01 / DEEP RESEARCH AGENTS

Complex research needs many searches

Can AI agents synthesize scientific literature as well as human experts?



Reflect

These studies use different comparisons. A higher benchmark score does not yet answer my question about human experts. Which tasks and domains were tested?



Search again

literature synthesis expert comparison tasks domains evaluation protocol

Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

4

01 / 深度研究智能体

复杂研究需要多次搜索。

AI 智能体能否像人类专家一样综合科学文献？

反思： 这些研究采用了不同的比较方式。基准得分更高还不能回答我关于人类专家的问题。测试了哪些任务和领域？

再次搜索： 任务、领域、评测协议、文献综合、专家比较。

Synthesizing scientific literature with retrieval-augmented language models. Nature 2026. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

复杂研究需要多次搜索

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01 / DEEP RESEARCH AGENTS

Complex research needs many searches

Can AI agents synthesize scientific literature as well as human experts?

FINAL ANSWER

Promising on tested tasks; broader parity with experts remains unproven.

On evaluated literature-synthesis questions, experts preferred OpenScholar answers over expert-written ones; citation accuracy was comparable to human experts. [1]

DR Tulu reports strong results across four long-form research benchmarks. Comparisons with other AI agents do not directly establish parity with human experts. [2]

[1] Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

[2] DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

01 / 深度研究智能体

复杂研究需要多次搜索。

AI 智能体能否像人类专家一样综合科学文献？

最终回答：在已测试任务上表现有希望；更广泛地达到专家水平尚未得到证明。

在已评测的文献综合问题中，专家更偏好 OpenScholar 的回答，而不是专家撰写的回答；其引用准确度与人类专家相当。[1]

DR Tulu 在四个长篇研究基准上报告了强劲结果。与其他 AI 智能体的比较并不能直接证明它已达到人类专家水平。[2]

[1] Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

[2] DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

复杂研究需要多次搜索

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

01 / DEEP RESEARCH AGENTS

Today’s lecture



Evaluation

Benchmarks, rubrics, and citation support



Modeling

Learning to search, reason, and synthesize



Retrieval for deep research

Finding evidence for the agent’s next step

5

本讲内容

评测：基准、评分标准与引用支持。

建模：学习搜索、推理和综合。

深度研究中的检索：为智能体的下一步寻找证据。

本讲内容

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02

Benchmarks & Evaluation

02

基准与评测

基准与评测

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Limitations of Classical QA Benchmarks

QUESTION

who wrote the score for the force awakens

WIKIPEDIA / SOURCE DOCUMENT

Star Wars: The Force Awakens – Original Motion Picture Soundtrack is the [film score](#) to the 2015 [film of the same name](#) composed by [John Williams](#) with Williams and [William Ross](#) conducting, and [Gustavo Dudamel](#) appearing as a "special guest conductor". The album was released by [Walt Disney Records](#) on December 18, 2015 in both [digipak](#) CD and digital formats.^{[1][2]}

ANSWER

John Williams

Official Natural Questions example; linked Wikipedia revision (2018).

7

02 / 基准

经典问答基准的局限。

问题：谁为《原力觉醒》的配乐作曲？

维基百科 / 来源文档。

答案：John Williams。

Natural Questions 官方示例；链接至维基百科修订版本（2018年）。

经典问答基准的局限

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Limitations of Classical QA Benchmarks

QUESTION

who wrote the score for the force awakens

Simple lookup

WIKIPEDIA / SOURCE DOCUMENT

Star Wars: The Force Awakens – Original Motion Picture Soundtrack

is the film score to the 2015 film of the same name composed by John Williams with Williams and William Ross conducting, and Gustavo Dudamel appearing as a "special guest conductor". The album was released by Walt Disney Records on December 18, 2015 in both digipak CD and digital formats.^{[1][2]}

ANSWER

John Williams

Official Natural Questions example; linked Wikipedia revision (2018).

7

02 / 基准

经典问答基准的局限。

问题：谁为《原力觉醒》的配乐作曲？

简单查找。

维基百科 / 来源文档。

答案：John Williams。

Natural Questions 官方示例；链接至维基百科修订版本（2018年）。

经典问答基准的局限

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 17 | 原页 15

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Limitations of Classical QA Benchmarks

QUESTION

who wrote the score for the force awakens

WIKIPEDIA / SOURCE DOCUMENT

Star Wars: The Force Awakens – Original Motion Picture Soundtrack

is the film score to the 2015 film of the same name composed by John Williams with Williams and William Ross conducting, and Gustavo Dudamel appearing as a "special guest conductor". The album was released by Walt Disney Records on December 18, 2015 in both digipak CD and digital formats.^{[1][2]}

ANSWER

John Williams

Simple lookup

General domain

Official Natural Questions example; linked Wikipedia revision (2018).

7

02 / 基准

经典问答基准的局限。

问题：谁为《原力觉醒》的配乐作曲？

简单查找；通用领域。

维基百科 / 来源文档。

答案：John Williams。

Natural Questions 官方示例；链接至维基百科修订版本（2018年）。

经典问答基准的局限

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 18 | 原页 16

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Limitations of Classical QA Benchmarks

QUESTION

who wrote the score for the force awakens

WIKIPEDIA / SOURCE DOCUMENT

Star Wars: The Force Awakens – Original Motion Picture Soundtrack

is the film score to the 2015 film of the same name composed by John Williams with Williams and William Ross conducting, and Gustavo Dudamel appearing as a "special guest conductor". The album was released by Walt Disney Records on December 18, 2015 in both digipak CD and digital formats.^{[1][2]}

ANSWER

John Williams

Simple lookup

General domain

Short-form answer

Official Natural Questions example; linked Wikipedia revision (2018).

7

02 / 基准

经典问答基准的局限。

问题：谁为《原力觉醒》的配乐作曲？

简单查找；通用领域；简短答案。

维基百科 / 来源文档。

答案：John Williams。

Natural Questions 官方示例；链接至维基百科修订版本（2018年）。

经典问答基准的局限

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 19 | 原页 17

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Limitations of Classical QA Benchmarks

QUESTION

who wrote the score for the force awakens

WIKIPEDIA / SOURCE DOCUMENT

Star Wars: The Force Awakens – Original Motion Picture Soundtrack

is the film score to the 2015 film of the same name composed by John Williams with Williams and William Ross conducting, and Gustavo Dudamel appearing as a "special guest conductor". The album was released by Walt Disney Records on December 18, 2015 in both digipak CD and digital formats.^{[1][2]}

ANSWER

John Williams

Simple lookup

General domain

Evidence use unclear

Short-form answer

Official Natural Questions example; linked Wikipedia revision (2018).

7

02 / 基准

经典问答基准的局限。

问题：谁为《原力觉醒》的配乐作曲？

简单查找；通用领域；证据使用情况不清楚。

维基百科 / 来源文档。

答案：John Williams。

Natural Questions 官方示例；链接至维基百科修订版本（2018年）。

经典问答基准的局限

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 20 | 原页 18

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

Simple lookup

A familiar fact may be recalled or found in one search.

2 / DOMAIN EXPERTISE

General domain

Broad web questions may not test expert knowledge.

3 / ANSWER QUALITY

Short-form answer

Answer matching does not assess multi-document synthesis.

4 / EVIDENCE SUPPORT

Evidence use unclear

Correct answers do not establish evidence use or reliability.

8

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：简单查找。熟悉的事实可能凭记忆回答，也可能只需一次搜索。

2 / 领域专长：通用领域。宽泛的网络问题未必能检验专家知识。

3 / 答案质量：简短答案。答案匹配无法评估多文档综合。

4 / 证据支持：证据使用情况不清楚。答案正确不能证明其使用了证据或具有可靠性。

从问答转向深度研究基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

Simple lookup

A familiar fact may be recalled or found in one search.

2 / DOMAIN EXPERTISE

General domain

Broad web questions may not test expert knowledge.

3 / ANSWER QUALITY

Short-form answer

Answer matching does not assess multi-document synthesis.

4 / EVIDENCE SUPPORT

Evidence use unclear

Correct answers do not establish evidence use or reliability.

8

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：简单查找。熟悉的事实可能凭记忆回答，也可能只需一次搜索。

2 / 领域专长：通用领域。宽泛的网络问题未必能检验专家知识。

3 / 答案质量：简短答案。答案匹配无法评估多文档综合。

4 / 证据支持：证据使用情况不清楚。答案正确不能证明其使用了证据或具有可靠性。

从问答转向深度研究基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

BrowseComp: building many-search questions

1,266 human-written questions

CREATE THE QUESTION

Start with the answer

A known person, event, or artifact.

Write a complex question

Hide its name; combine factual clues.

VERIFY THE DIFFICULTY

LM + search checks

Models should fail;

5 Google searches should not suffice.

Human check

Still hard after 10 minutes?

Target checked on a subset.

BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025.

9

02 / 基准

BrowseComp：构造需要多次搜索的问题。

1,266个人类撰写的问题。

构造问题：从答案开始——一个已知人物、事件或作品；隐藏其名称，把事实线索组合成复杂问题。

验证难度：语言模型 + 搜索检查——模型应答不出来，5次 Google 搜索也不应足够；人工检查——测试子集是否仍需超过10分钟才能解决？

BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025.

BrowseComp：构造需要多次搜索的问题

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

BrowseComp: a question built from clues

ORIGINAL QUESTION / CLUES IN READING ORDER

Please identify the fictional character who

BEHAVIOR	occasionally <u>breaks the fourth wall with the audience</u> ,
BACKSTORY	has a backstory involving <u>help from selfless ascetics</u> ,
PERSONALITY	is <u>known for his humor</u> , and
TV SERIES	had a TV show that <u>aired between the 1960s and 1980s</u>
EPISODE COUNT	with <u>fewer than 50 episodes</u> .

Find one character who satisfies **all five clues**.

BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025.

10

02 / 基准

BrowseComp：由线索构成的问题。

原始问题 / 按阅读顺序给出的线索：请找出一位虚构人物：行为——偶尔打破第四面墙，与观众直接互动；背景——曾得到无私苦行者的帮助；性格——以幽默著称；电视剧——其电视节目在20世纪60至80年代播出；集数——少于50集。

找出一位符合全部五条线索的人物。

BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025.

BrowseComp：由线索构成的问题

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

BrowseComp: a question built from clues

ORIGINAL QUESTION / CLUES IN READING ORDER

Please identify the fictional character who

BEHAVIOR	occasionally <u>breaks the fourth wall with the audience</u> ,
BACKSTORY	has a backstory involving <u>help from selfless ascetics</u> ,
PERSONALITY	is <u>known for his humor</u> , and
TV SERIES	had a TV show that <u>aired between the 1960s and 1980s</u>
EPISODE COUNT	with <u>fewer than 50 episodes</u> .

Find one character who satisfies **all five clues**.

Reference answer: **Plastic Man**

BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025.

10

02 / 基准

BrowseComp：由线索构成的问题。

原始问题 / 按阅读顺序给出的线索：请找出一位虚构人物：偶尔打破第四面墙与观众互动；其背景故事涉及无私苦行者的帮助；以幽默著称；拥有一部在20世纪60至80年代播出的电视剧；集数少于50集。

找出一位符合全部五条线索的人物。

参考答案：Plastic Man。

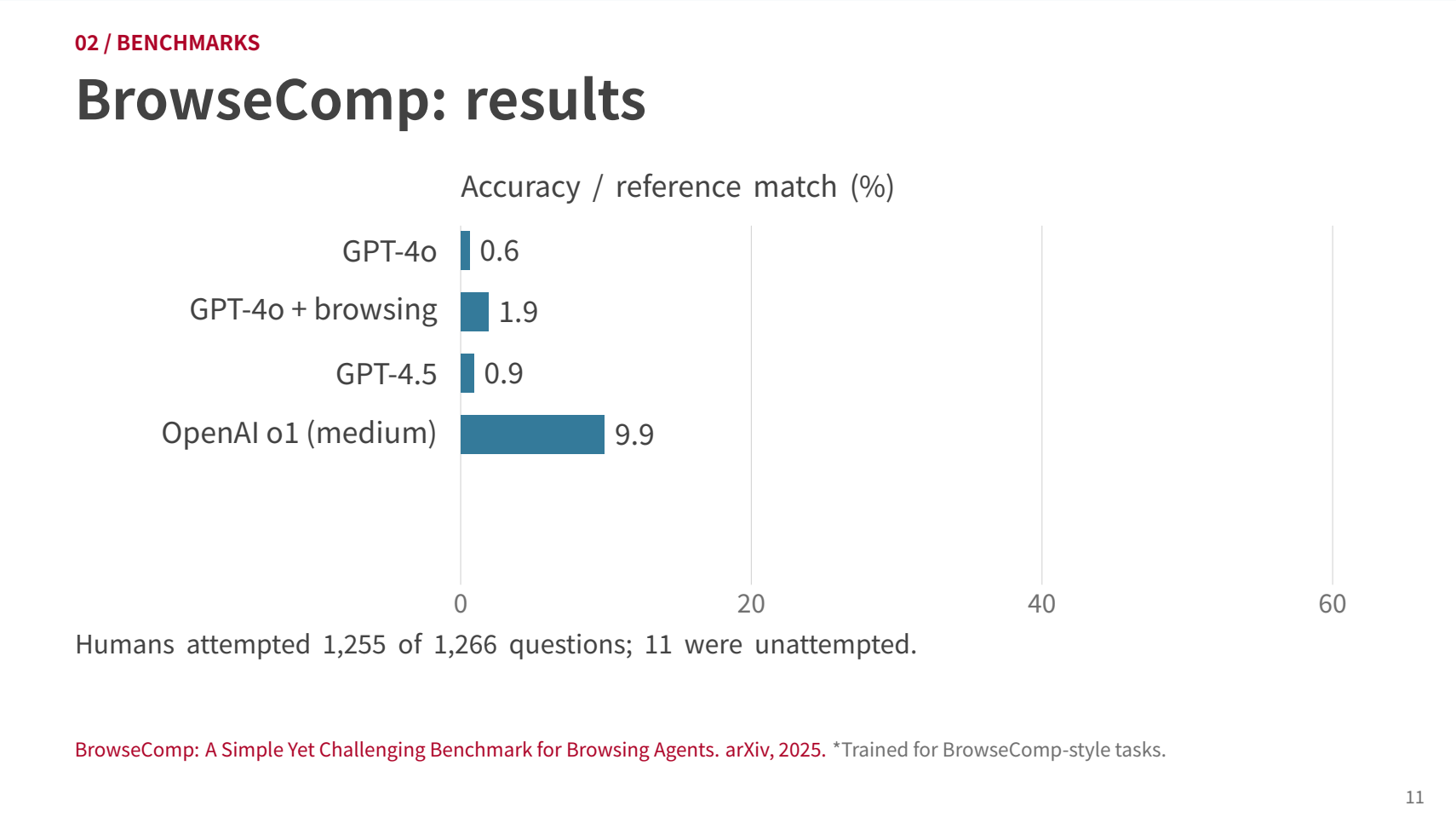
BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025.

BrowseComp：由线索构成的问题

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



02 / 基准

BrowseComp：结果。

准确率 / 与参考答案匹配率（%）：GPT-4o 0.6；GPT-4o + browsing 1.9；GPT-4.5 0.9；OpenAI o1（medium）9.9。

人类尝试回答了 1,266 个问题中的 1,255 个；11个未作答。

BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025.

BrowseComp：结果

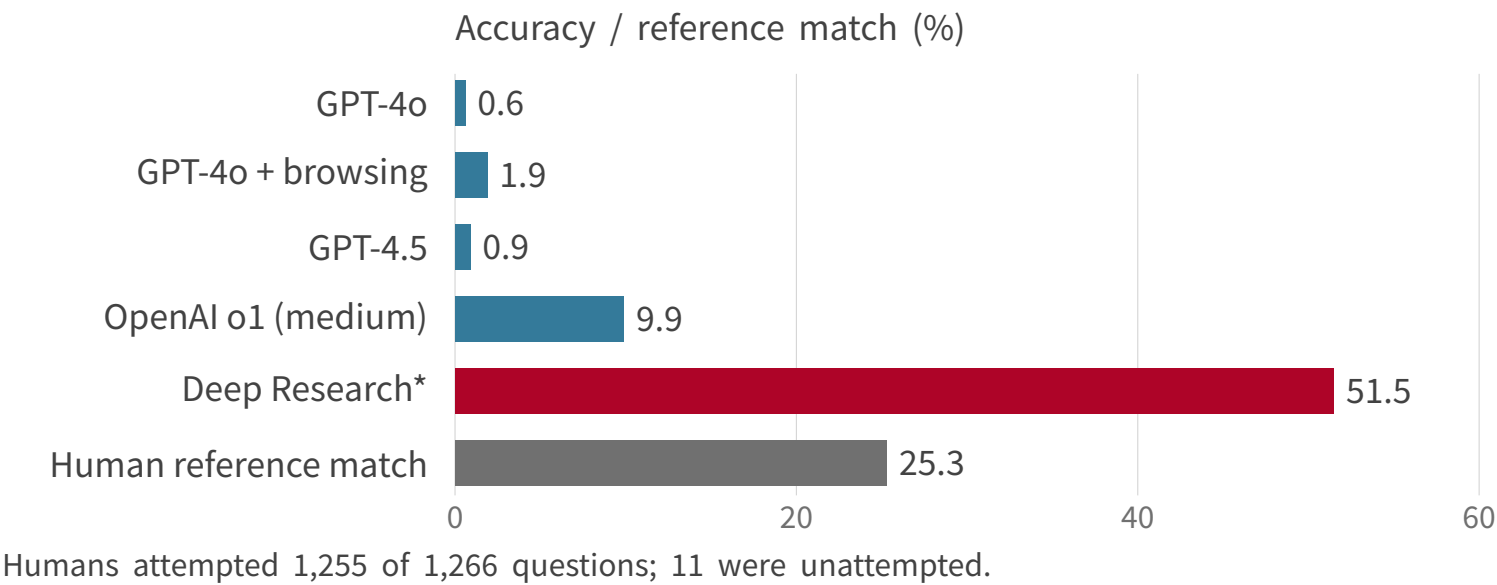
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

BrowseComp: results



BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025. *Trained for BrowseComp-style tasks.

11

02 / 基准

BrowseComp：结果。

准确率 / 与参考答案匹配率（%）：GPT-4o 0.6；GPT-4o + browsing 1.9；GPT-4.5 0.9；OpenAI o1（medium）9.9；Deep Research* 51.5；人类参考答案匹配率 25.3。

人类尝试回答了 1,266 个问题中的 1,255 个；11个未作答。

BrowseComp: A Simple Yet Challenging Benchmark for Browsing Agents. arXiv, 2025。*针对 BrowseComp 类任务训练。

BrowseComp：结果

第 10 讲 · 深度研究智能体：任务、基准与方法

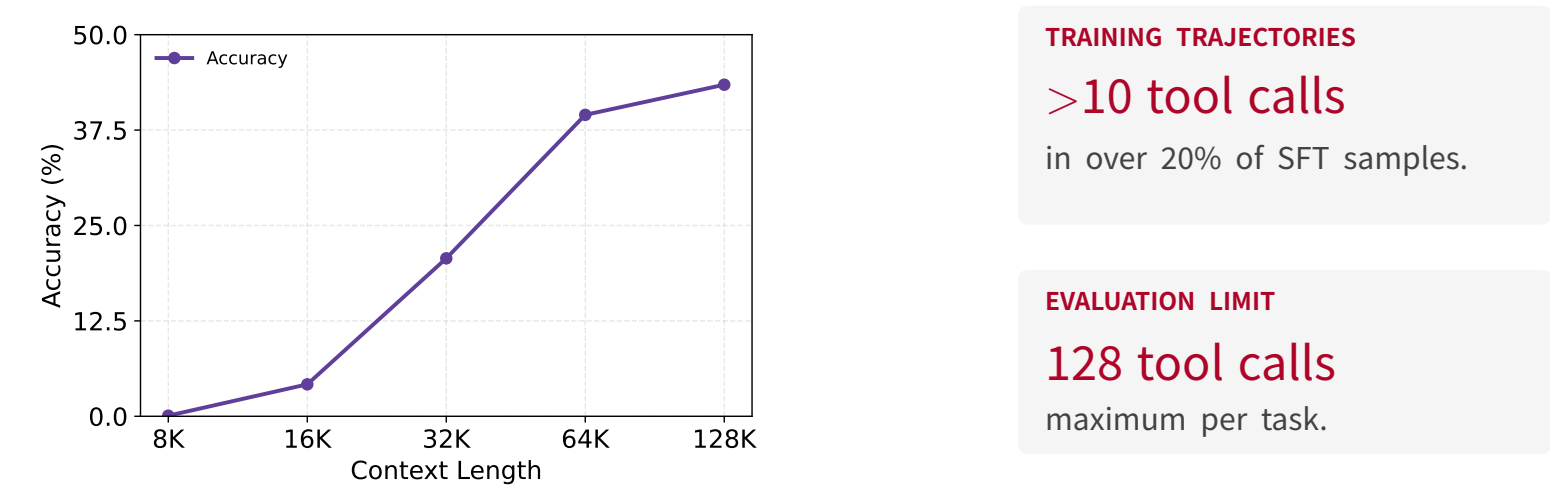
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

More room to search improves accuracy

Tongyi DeepResearch on BrowseComp



More context allows longer interaction histories; accuracy rises with the budget.

Tongyi DeepResearch Technical Report. arXiv, 2025.

12

02 / 基准

增加搜索预算可提升准确率。

Tongyi DeepResearch 在 BrowseComp 上的表现。

图(a) BrowseComp 的交互轮数扩展：横轴上下文长度 8K、16K、32K、64K、128K；纵轴准确率（%）， 约从0升至50。上下文越长，可保留的交互历史越长；预算增加时准确率上升。

图(b) 模拟环境中的奖励：横轴训练步数，纵轴奖励；曲线标注“模拟环境”和“EMA 平滑”。

训练轨迹：超过20%的 SFT 样本包含超过10次工具调用。评测限制：每个任务最多128次工具调用；模拟环境中最多50次。

Tongyi DeepResearch Technical Report. arXiv, 2025. 图10：交互扩展与模拟环境的详细分析。

译注 / 与原文翻译分开

左侧准确率曲线各点未在图上标出具体数值，未据图估算。

增加搜索预算可提升准确率

第 10 讲 · 深度研究智能体：任务、基准与方法

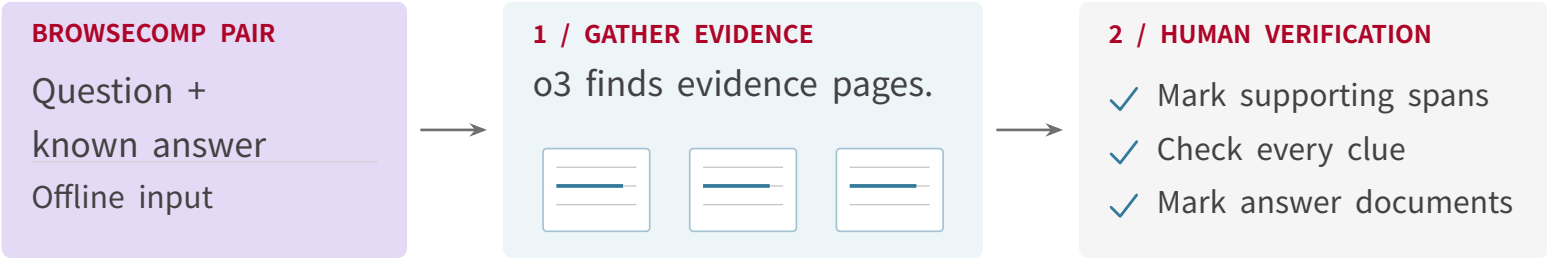
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

BrowseComp-Plus makes the corpus fixed

Collect and verify the evidence, then freeze the document collection.



BrowseComp-Plus: A Fair and Disentangled Evaluation Benchmark for Deep Search Agents. ACL 2026.

13

02 / 基准

BrowseComp-Plus 固定了语料库。

先收集并核验证据，再冻结文档集合。

BrowseComp 配对样本：问题 + 已知答案，作为离线输入。

1 / 收集证据：o3 找到证据页面。

2 / 人工核验：标记支持性段落；检查每条线索；标记答案文档。

BrowseComp-Plus: A Fair and Disentangled Evaluation Benchmark for Deep Search Agents. ACL 2026.

BrowseComp-Plus 固定了语料库

第 10 讲 · 深度研究智能体：任务、基准与方法

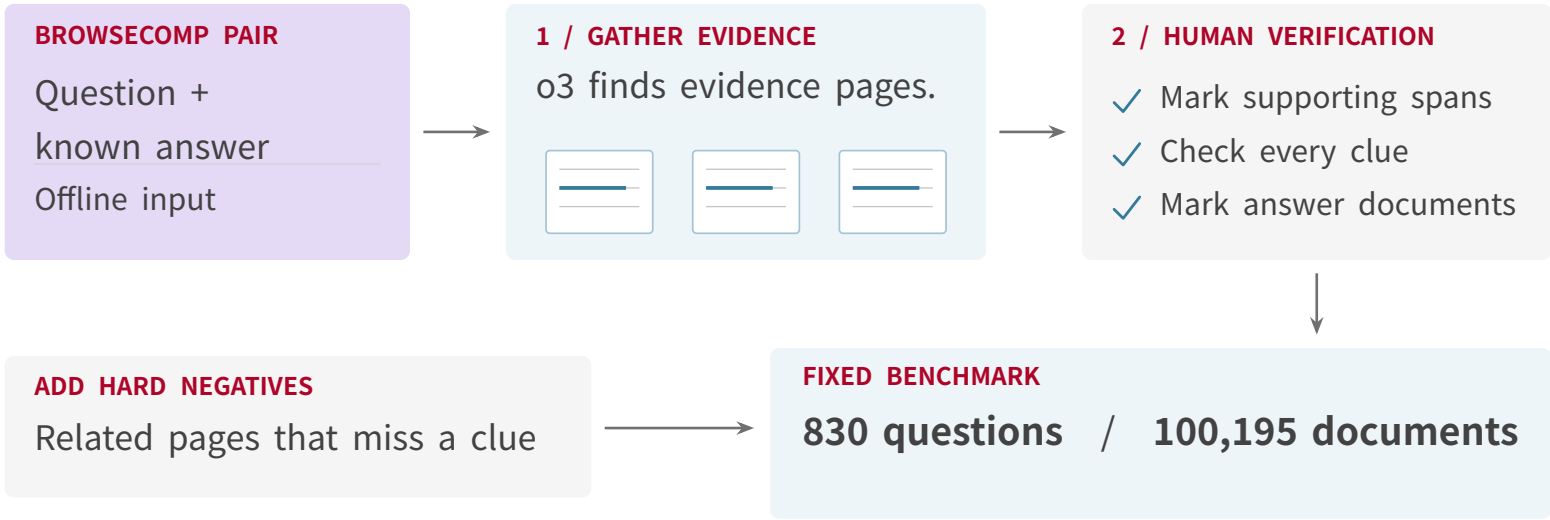
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

BrowseComp-Plus makes the corpus fixed

Collect and verify the evidence, then freeze the document collection.



BrowseComp-Plus: A Fair and Disentangled Evaluation Benchmark for Deep Search Agents. ACL 2026.

13

02 / 基准

BrowseComp-Plus 固定了语料库。

先收集并核验证据，再冻结文档集合。

BrowseComp 配对样本：问题 + 已知答案，作为离线输入。

1 / 收集证据：o3 找到证据页面。2 / 人工核验：标记支持性段落、检查每条线索、标记答案文档。

加入困难负例：与问题相关但缺少某条线索的页面。

固定基准：830个问题 / 100,195份文档。

BrowseComp-Plus: A Fair and Disentangled Evaluation Benchmark for Deep Search Agents. ACL 2026.

BrowseComp-Plus 固定了语料库

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

BrowseComp / BrowseComp-Plus

Hard-to-find answers; a fixed corpus for controlled comparisons.

2 / DOMAIN EXPERTISE

General domain

Broad web questions may not test expert knowledge.

3 / ANSWER QUALITY

Short-form answer

Answer matching does not assess multi-document synthesis.

4 / EVIDENCE SUPPORT

Evidence use unclear

Correct answers do not establish evidence use or reliability.

14

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：BrowseComp / BrowseComp-Plus——答案难以找到；固定语料库便于受控比较。

2 / 领域专长：通用领域——宽泛的网络问题未必能检验专家知识。

3 / 答案质量：简短答案——答案匹配无法评估多文档综合。

4 / 证据支持：证据使用情况不清楚；答案正确不能证明其使用了证据或具有可靠性。

从问答转向深度研究基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

BrowseComp / BrowseComp-Plus

Hard-to-find answers; a fixed corpus for controlled comparisons.

2 / DOMAIN EXPERTISE

General domain

Broad web questions may not test expert knowledge.

3 / ANSWER QUALITY

Short-form answer

Answer matching does not assess multi-document synthesis.

4 / EVIDENCE SUPPORT

Evidence use unclear

Correct answers do not establish evidence use or reliability.

14

从问答转向深度研究基准

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：BrowseComp / BrowseComp-Plus——答案难以找到；固定语料库便于受控比较。

2 / 领域专长：通用领域——宽泛的网络问题未必能检验专家知识。

3 / 答案质量：简短答案——答案匹配无法评估多文档综合。

4 / 证据支持：证据使用情况不清楚；答案正确不能证明其使用了证据或具有可靠性。

第 10 讲 · 深度研究智能体：任务、基准与方法

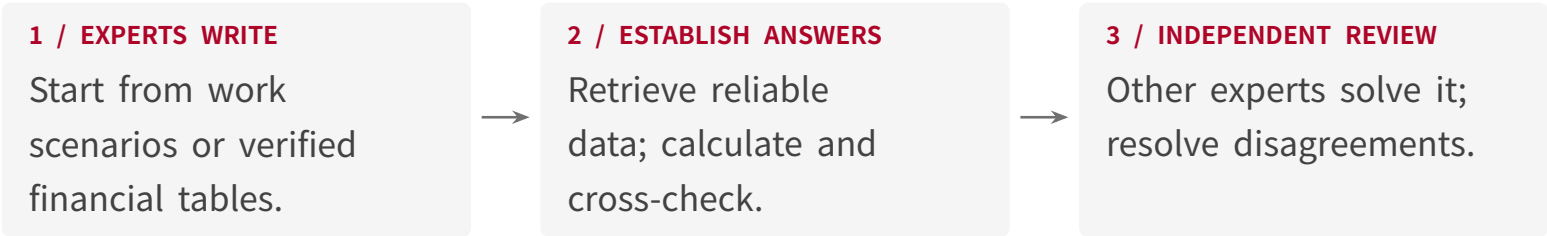
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

FinSearchComp: expert-written questions

Focus: complex historical investigation



EXAMPLE / PARAPHRASED
How did Johnson & Johnson’s international share of revenue change year over year during 2022–2024?

FinSearchComp: Towards a Realistic, Expert-Level Evaluation of Financial Search and Reasoning. ICLR 2026.

15

02 / 基准

FinSearchComp：由专家编写的问题。

重点：复杂的历史调查。

- 1 / 专家撰写：从工作场景或已核实的财务表格出发。
- 2 / 确立答案：检索可靠数据、进行计算并交叉核验。
- 3 / 独立审查：其他专家独立解题，并解决分歧。

示例 / 意译：强生公司2022–2024年国际收入占比逐年如何变化？

FinSearchComp: Towards a Realistic, Expert-Level Evaluation of Financial Search and Reasoning. ICLR 2026.

FinSearchComp：由专家编写的问题

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Search benchmarks with domain expertise

Domain knowledge + multi-step evidence gathering

MedBrowseComp Medicine: link trials, drugs, and regulatory facts.
MedBrowseComp: Benchmarking Medical Deep Research and Computer Use. GenAI4Health Workshop, NeurIPS 2025.

FinSearchComp Finance: investigate facts across financial sources.
FinSearchComp: Towards a Realistic, Expert-Level Evaluation of Financial Search and Reasoning. ICLR 2026.

Scientific deep research / paper retrieval
ScholarSearch: academic search. AutoResearchBench: one paper or all matches.
ScholarSearch: Benchmarking Scholar Searching Ability of LLMs. arXiv, 2025.
AutoResearchBench: Benchmarking AI Agents on Complex Scientific Literature Discovery. arXiv, 2026.

16

02 / 基准

领域知识 + 多步证据收集。

MedBrowseComp：医学领域，将临床试验、药物与监管事实联系起来。
MedBrowseComp: Benchmarking Medical Deep Research and Computer Use. GenAI4Health Workshop, NeurIPS 2025.

FinSearchComp：金融领域，跨金融来源调查事实。FinSearchComp: Towards a Realistic, Expert-Level Evaluation of Financial Search and Reasoning. ICLR 2026.

科学深度研究 / 论文检索：ScholarSearch 测试学术搜索；AutoResearchBench 测试单篇论文或找出全部匹配论文。

ScholarSearch: Benchmarking Scholar Searching Ability of LLMs. arXiv, 2025. · AutoResearchBench: Benchmarking AI Agents on Complex Scientific Literature Discovery. arXiv, 2026.

具备领域专长的搜索基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

BrowseComp / BrowseComp-Plus

Hard-to-find answers; a fixed corpus for controlled comparisons.

2 / DOMAIN EXPERTISE

MedBrowseComp
FinSearchComp

Expert knowledge guides medical and financial search.

3 / ANSWER QUALITY

Short-form answer

Answer matching does not assess multi-document synthesis.

4 / EVIDENCE SUPPORT

Evidence use unclear

Correct answers do not establish evidence use or reliability.

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：BrowseComp / BrowseComp-Plus——答案难以找到；固定语料库便于受控比较。

2 / 领域专长：MedBrowseComp、FinSearchComp——专家知识指导医学与金融搜索。

3 / 答案质量：简短答案——答案匹配无法评估多文档综合。

4 / 证据支持：证据使用情况不清楚；答案正确不能证明其使用了证据或具有可靠性。

从问答转向深度研究基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

BrowseComp / BrowseComp-Plus

Hard-to-find answers; a fixed corpus for controlled comparisons.

2 / DOMAIN EXPERTISE

MedBrowseComp
FinSearchComp

Expert knowledge guides medical and financial search.

3 / ANSWER QUALITY

Short-form answer

Answer matching does not assess multi-document synthesis.

4 / EVIDENCE SUPPORT

Evidence use unclear

Correct answers do not establish evidence use or reliability.

17

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：BrowseComp / BrowseComp-Plus——答案难以找到；固定语料库便于受控比较。

2 / 领域专长：MedBrowseComp、FinSearchComp——专家知识指导医学与金融搜索。

3 / 答案质量：简短答案——答案匹配无法评估多文档综合。

4 / 证据支持：证据使用情况不清楚；答案正确不能证明其使用了证据或具有可靠性。

从问答转向深度研究基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Research needs multi-dimensional evaluation

Evaluate **accuracy, coverage, qualifications, and evidence support.**

Can AI agents synthesize scientific literature as well as human experts?

ONE REFERENCE SUMMARY

Experts preferred OpenScholar in one study. Parity across domains is unproven.

18

02 / 基准

研究需要多维度评测。

评估准确度、覆盖度、限定说明和证据支持。

问题：AI 智能体能否像人类专家一样综合科学文献？

一份参考摘要：一项研究中专家更偏好 OpenScholar；跨领域是否达到同等水平尚未得到证明。

研究需要多维度评测

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Research needs multi-dimensional evaluation

Evaluate **accuracy, coverage, qualifications, and evidence support.**

Can AI agents synthesize scientific literature as well as human experts?

ONE REFERENCE SUMMARY
Experts preferred OpenScholar in one study. Parity across domains is unproven.

ANOTHER VALID SUMMARY
One study favors OpenScholar; it does not establish equivalence across domains.

⚠ Several valid answers

18

02 / 基准

研究需要多维度评测。

评估准确度、覆盖度、限定说明和证据支持。

问题：AI 智能体能否像人类专家一样综合科学文献？

一份参考摘要：一项研究中专家更偏好 OpenScholar；跨领域是否达到同等水平尚未得到证明。

另一种有效摘要：一项研究偏好 OpenScholar，但不能证明其在各领域都与专家等效。

存在多种有效答案。

研究需要多维度评测

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Research needs multi-dimensional evaluation

Evaluate **accuracy, coverage, qualifications, and evidence support.**

Can AI agents synthesize scientific literature as well as human experts?

ONE REFERENCE SUMMARY
Experts preferred OpenScholar in one study. Parity across domains is unproven.

ANOTHER VALID SUMMARY
One study favors OpenScholar; it does not establish equivalence across domains.

⚠ Several valid answers

SIMILAR, BUT INCORRECT
Experts preferred OpenScholar in one study. Parity across domains is **proven.**

⚠ Similarity to gold is not enough

18

02 / 基准

研究需要多维度评测。

评估准确度、覆盖度、限定说明和证据支持。

问题：AI 智能体能否像人类专家一样综合科学文献？

一份参考摘要：一项研究中专家更偏好 OpenScholar；跨领域是否达到同等水平尚未得到证明。

另一种有效摘要：一项研究偏好 OpenScholar，但不能证明其在各领域都与专家等效。

相似但错误的摘要：一项研究中专家更偏好 OpenScholar；跨领域已证明其与专家等效。

存在多种有效答案；与标准答案相似并不足够。

研究需要多维度评测

第 10 讲 · 深度研究智能体：任务、基准与方法

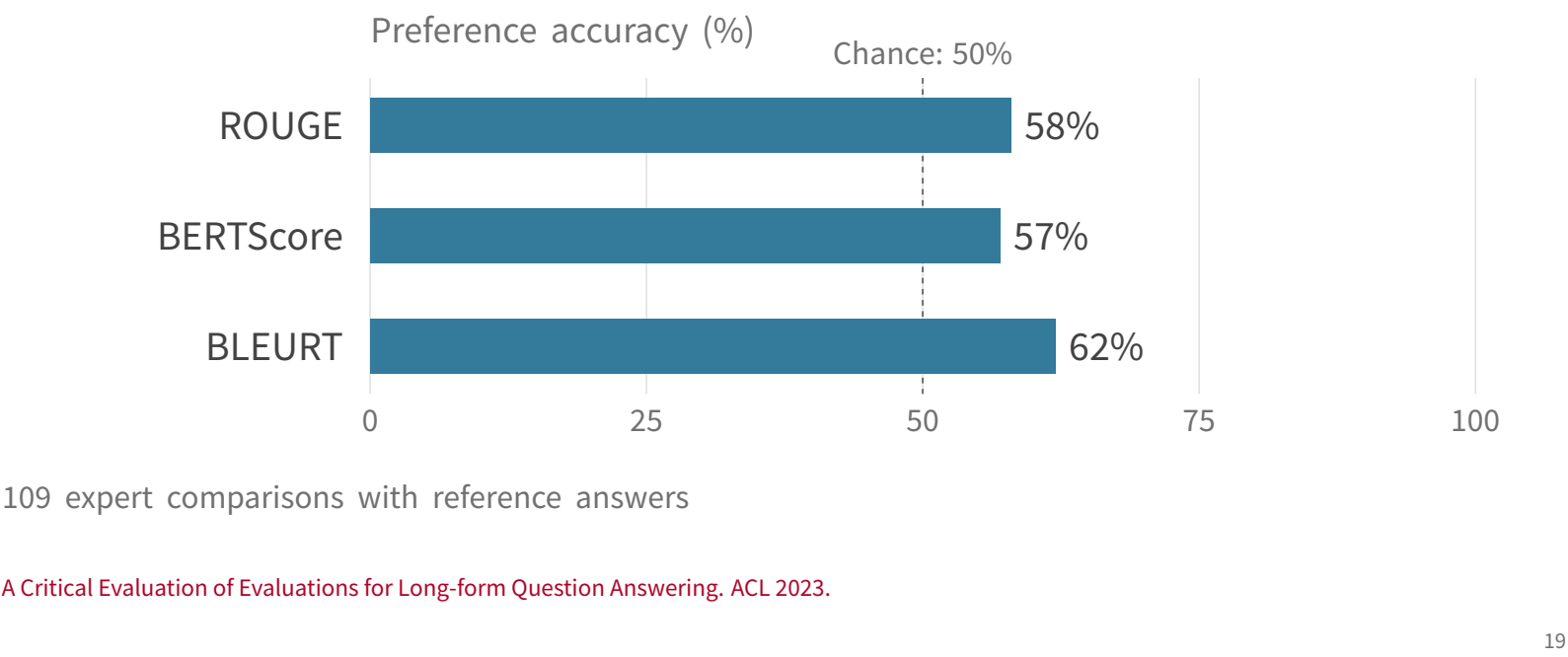
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Similarity metrics can misrank answers

Does a higher similarity score select the answer experts prefer?



02 / 基准

相似度得分更高，是否就能选出专家更偏好的答案？

偏好准确率（%）：ROUGE 58%；BERTScore 57%；BLEURT 62%；随机水平为50%。

109次专家与参考答案的比较。

A Critical Evaluation of Evaluations for Long-form Question Answering. ACL 2023.

相似度指标可能错误排序答案

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

ScholarQABench: evaluate with expert rubrics

Check each rubric item against the answer, then combine the scores.

QUESTION

Can AI agents synthesize scientific literature as well as human experts?

[1] Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

02 / 基准

ScholarQABench：用专家评分标准评测。

逐项对照回答检查评分标准，再合并各项得分。

问题：AI 智能体能否像人类专家一样综合科学文献？

[1] Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

ScholarQABench：用专家评分标准评测

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

ScholarQABench: evaluate with expert rubrics

Check each rubric item against the answer, then combine the scores.

QUESTION

Can AI agents synthesize scientific literature as well as human experts?



RUBRIC ITEM	Weight	Check
Reports a direct comparison with human experts.	2	
Qualifies the conclusion by task and domain.	1	

[1] Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

02 / 基准

ScholarQABench：用专家评分标准评测。

逐项对照回答检查评分标准，再合并各项得分。

问题：AI 智能体能否像人类专家一样综合科学文献？

评分项：报告与人类专家的直接比较，权重2；按任务和领域限定结论，权重1。

[1] Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

ScholarQABench：用专家评分标准评测

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

ScholarQABench: evaluate with expert rubrics

Check each rubric item against the answer, then combine the scores.

QUESTION

Can AI agents synthesize scientific literature as well as human experts?

CANDIDATE ANSWER

OpenScholar wins 70% of expert comparisons [1]. This holds across all fields.

RUBRIC ITEM	Weight	Check
Reports a direct comparison with human experts.	2	1
Qualifies the conclusion by task and domain.	1	0

$$S_c = \frac{\sum_i w_i b_i}{\sum_i w_i} = \frac{2 \times 1 + 1 \times 0}{3} = 0.67$$

[1] Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

02 / 基准

ScholarQABench：用专家评分标准评测。

逐项对照回答检查评分标准，再合并各项得分。

问题：AI 智能体能否像人类专家一样综合科学文献？

候选回答：“在专家比较中，OpenScholar 有70%的胜出率。[1] 这一结果适用于所有领域。”

评分项：报告与人类专家的直接比较（权重2，检查得分1）；按任务和领域限定结论（权重1，检查得分0）。

评分： $S_c = \sum_i w_i b_i / \sum_i w_i = (2 \times 1 + 1 \times 0) / 3 = 0.67$ 。

[1] Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

ScholarQABench：用专家评分标准评测

第 10 讲 · 深度研究智能体：任务、基准与方法

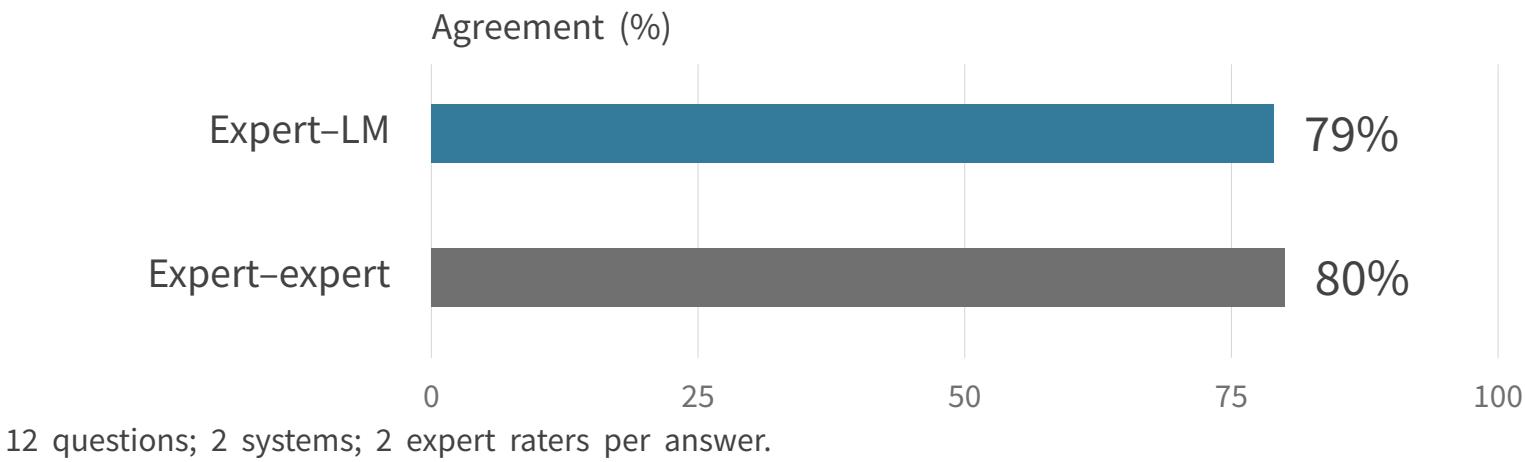
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Do rubric judgments agree with experts?

Compare judgments on the same rubric items.



Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

21

02 / 基准

评分标准判断与专家意见一致吗？

比较对相同评分项作出的判断。

一致率（%）：专家-语言模型 79%；专家-专家 80%。

12个问题；2个系统；每个答案由2位专家评分。

Synthesizing scientific literature with retrieval-augmented language models. Nature 2026.

评分标准判断与专家意见一致吗？

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

ResearchQA: generate rubrics from surveys

Use LMs to scale up rubric annotation.

21K questions / 75 fields / 160K rubric items

Survey-Mined Query

How does the frequency of terms in pre-training data influence numerical reasoning performance in few-shot settings? (Engineering)

Research System: The frequency of terms in pre-training data significantly influences a model's numerical reasoning performance, particularly in few-shot learning scenarios [1]. Models pre-trained [1] Scaling Laws and Data Frequency Effects in Large Language [1]

Survey-Mined Evaluation Rubric

Does the response reference the “performance gap” concept from the Razeghi et al. (2022) paper [1]? 0/4 Not at all covered

Does the response include examples of studies or experiments that investigate the impact of term frequency on numerical reasoning performance? 4/4 Completely covered

Does the response discuss the correlation between the frequency of terms in pre-training data and numerical reasoning performance? 1/4 Barely covered

Additional rubric items ...

Source Survey: The Mystery of In-Context Learning (Zhou et al., 2024)

Survey section

LM

Research question

LM-generated rubric

ResearchQA: Evaluating Scholarly Question Answering at Scale Across 75 Fields with Survey-Mined Questions and Rubrics. TACL 2026.

22

ResearchQA：从调查问卷生成评分标准

02 / 基准

ResearchQA：从调查问卷生成评分标准。

使用语言模型扩大评分标准标注规模。21,000个问题 / 75个领域 / 160,000条评分项。

示例流程：从调查问卷的研究问题中挖掘问题；研究系统生成长篇回答；评判器依据自动生成的评分标准打分。图中覆盖度示例包括 “完全未覆盖” “完全覆盖” “基本覆盖” 。

示例研究问题：预训练数据中的术语频率如何影响少样本设置下的数值推理表现？（工程领域。）评分标准涉及 Razeghi 等（2022）提出的 performance gap 概念、相关研究示例，以及术语频率与数值推理表现之间的相关性。

ResearchQA: Evaluating Scholarly Question Answering at Scale Across 75 Fields with Survey-Mined Questions and Rubrics. TACL 2026.

译注 / 与原文翻译分开

本页主体为论文图示，部分小字和评分项在原图中难以逐字辨识；以上只转录可辨文字与结构。

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 45 | 原页 43

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

Generated rubrics can encode the wrong target

Evaluate the rubric itself, before using it to evaluate an answer.

RESEARCHQA / EXPERT AUDIT

15%

of parametric rubric items were
vacuous, erroneous, or unclear.

WHAT CAN GO WRONG?

- A criterion cites a nonexistent paper.
- A criterion merely restates the question.
- A vague criterion lets plausible errors pass.

Ground criteria in sources, then audit their relevance and verifiability.

DeepResearch Bench II uses expert reports + manual revision + expert review.

ResearchQA: Evaluating Scholarly Question Answering at Scale Across 75 Fields with Survey-Mined Questions and Rubrics. TACL 2026.; DeepResearch Bench II: Diagnosing Deep Research Agents via Rubrics from Expert Reports. arXiv, 2026.

23

02 / 基准

使用评分标准评答案前，先评估评分标准本身。

ResearchQA / 专家审计：15%的参数化评分项空泛、错误或含糊。

可能的问题：标准引用了不存在的论文；标准只是重复问题本身；标准过于模糊，导致看似合理的错误也能通过。

让标准有来源依据，再审查其相关性与可验证性。

DeepResearch Bench II 使用专家报告 + 人工修订 + 专家审查。

ResearchQA: Evaluating Scholarly Question Answering at Scale Across 75 Fields with Survey-Mined Questions and Rubrics. TACL 2026. ·
DeepResearch Bench II: Diagnosing Deep Research Agents via Rubrics from Expert Reports. arXiv, 2026.

自动生成的评分标准可能设定错误目标

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

BrowseComp / BrowseComp-Plus

Hard-to-find answers; a fixed corpus for controlled comparisons.

2 / DOMAIN EXPERTISE

MedBrowseComp
FinSearchComp

Expert knowledge guides medical and financial search.

3 / ANSWER QUALITY

ScholarQABench / ResearchQA

Rubrics assess response quality.

4 / EVIDENCE SUPPORT

Evidence use unclear

Correct answers do not establish evidence use or reliability.

24

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：BrowseComp / BrowseComp-Plus——答案难以找到；固定语料库便于受控比较。

2 / 领域专长：MedBrowseComp、FinSearchComp——专家知识指导医学与金融搜索。

3 / 答案质量：ScholarQABench / ResearchQA——评分标准评估回答质量。

4 / 证据支持：证据使用情况不清楚；答案正确不能证明其使用了证据或具有可靠性。

从问答转向深度研究基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

BrowseComp / BrowseComp-Plus

Hard-to-find answers; a fixed corpus for controlled comparisons.

2 / DOMAIN EXPERTISE

MedBrowseComp
FinSearchComp

Expert knowledge guides medical and financial search.

3 / ANSWER QUALITY

ScholarQABench / ResearchQA

Rubrics assess response quality.

4 / EVIDENCE SUPPORT

Evidence use unclear

Correct answers do not establish evidence use or reliability.

24

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：BrowseComp / BrowseComp-Plus——答案难以找到；固定语料库便于受控比较。

2 / 领域专长：MedBrowseComp、FinSearchComp——专家知识指导医学与金融搜索。

3 / 答案质量：ScholarQABench / ResearchQA——评分标准评估回答质量。

4 / 证据支持：证据使用情况不清楚；答案正确不能证明其使用了证据或具有可靠性。

从问答转向深度研究基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

DeepResearch Bench: check the citations

FACT checks whether the cited page supports each claim.

1 / EXTRACT CLAIM + URL

OpenScholar was compared with human answers. [1]

OpenScholar matches experts in every field. [1]

DeepResearch Bench: A Comprehensive Benchmark for Deep Research Agents. ICLR 2026.

25

02 / 基准

DeepResearch Bench：检查引用。

FACT 检查引用页面是否支持每项主张。

1 / 提取主张 + URL：主张一：“OpenScholar 与人类答案进行了比较。” [1]；主张二：“OpenScholar 在每个领域都与专家相匹配。” [1]。

DeepResearch Bench: A Comprehensive Benchmark for Deep Research Agents. ICLR 2026.

DeepResearch Bench：检查引用

第 10 讲 · 深度研究智能体：任务、基准与方法

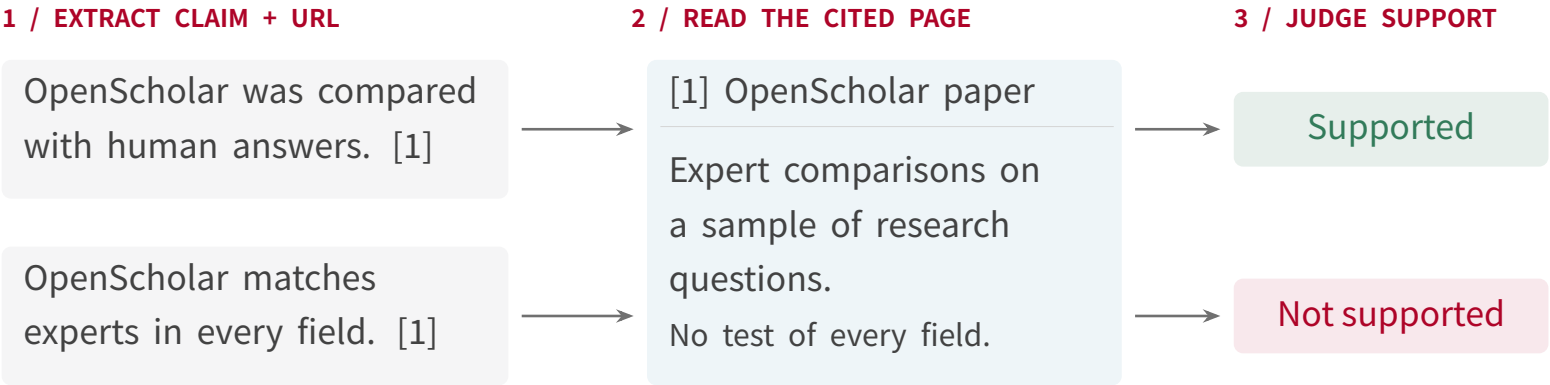
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

DeepResearch Bench: check the citations

FACT checks whether the cited page supports each claim.



DeepResearch Bench: A Comprehensive Benchmark for Deep Research Agents. ICLR 2026.

25

02 / 基准

DeepResearch Bench：检查引用。

FACT 检查引用页面是否支持每项主张。

1 / 提取主张 + URL：主张一：“OpenScholar 与人类答案进行了比较。” [1]；主张二：“OpenScholar 在每个领域都与专家相匹配。” [1]。

2 / 阅读被引用页面：[1] OpenScholar 论文提到对一组研究问题进行了专家比较。

3 / 判断支持情况：主张一——有支持；主张二——不支持，因为没有测试所有领域。

DeepResearch Bench: A Comprehensive Benchmark for Deep Research Agents. ICLR 2026.

DeepResearch Bench：检查引用

第 10 讲 · 深度研究智能体：任务、基准与方法

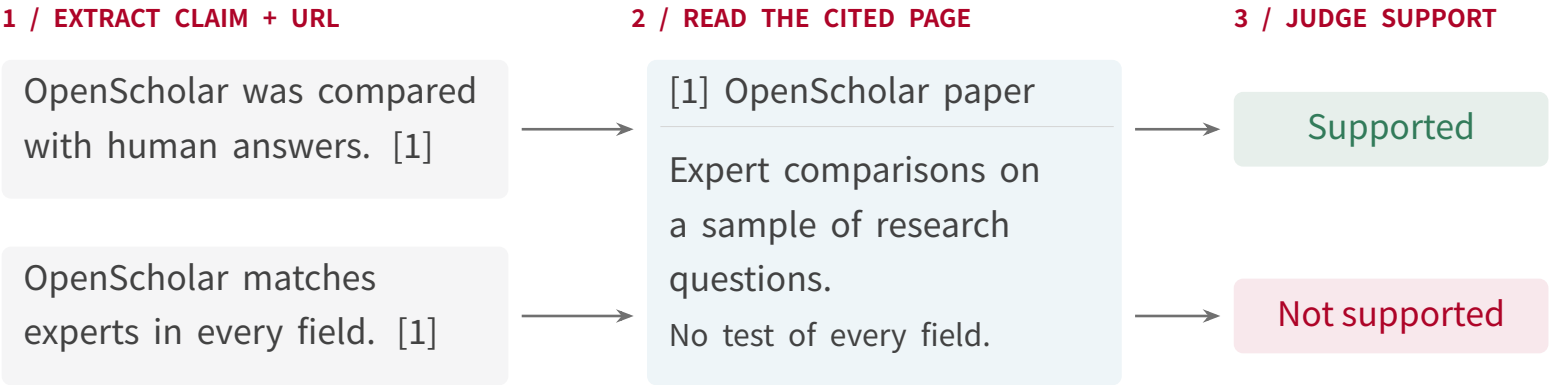
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

DeepResearch Bench: check the citations

FACT checks whether the cited page supports each claim.



Citation accuracy: 1/2 = 50% Effective citations: 1 supported pair

DeepResearch Bench: A Comprehensive Benchmark for Deep Research Agents. ICLR 2026.

25

02 / 基准

DeepResearch Bench：检查引用。

FACT 检查引用页面是否支持每项主张。

1 / 提取主张 + URL；2 / 阅读被引用页面；3 / 判断支持情况。

主张一：“OpenScholar 与人类答案进行了比较。” [1]——有支持：论文报告了对一组研究问题的专家比较。

主张二：“OpenScholar 在每个领域都与专家相匹配。” [1]——不支持：没有测试所有领域。

引用准确率：1/2 = 50%。有效引用：1组得到支持的主张—来源配对。

DeepResearch Bench: A Comprehensive Benchmark for Deep Research Agents. ICLR 2026.

DeepResearch Bench：检查引用

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

02 / BENCHMARKS

From QA to deep research benchmarks

1 / SEARCH COMPLEXITY

BrowseComp / BrowseComp-Plus

Hard-to-find answers; a fixed corpus for controlled comparisons.

2 / DOMAIN EXPERTISE

MedBrowseComp
FinSearchComp

Expert knowledge guides medical and financial search.

3 / ANSWER QUALITY

ScholarQABench / ResearchQA

Rubrics assess response quality.

4 / EVIDENCE SUPPORT

DeepResearch Bench

Check whether cited sources support the report’s claims.

26

02 / 基准

从问答转向深度研究基准。

1 / 搜索复杂度：BrowseComp / BrowseComp-Plus——答案难以找到；固定语料库便于受控比较。

2 / 领域专长：MedBrowseComp、FinSearchComp——专家知识指导医学与金融搜索。

3 / 答案质量：ScholarQABench / ResearchQA——评分标准评估回答质量。

4 / 证据支持：DeepResearch Bench——检查引用来源是否支持报告中的主张。

从问答转向深度研究基准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03

Agent Modeling

03

智能体建模

智能体建模

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Deep research agent inference

CONTEXT SO FAR

USER / QUESTION

Can agents match experts?

Language model agent

1 / GENERATE REASONING

<think>

Find

expert

comparisons

</think>

Blue: model-generated text Gray: user input and tool output

28

03 / 智能体建模

深度研究智能体推理过程。

当前上下文：用户问题—— “智能体能与专家相匹配吗？ ”

语言模型智能体。

1 / 生成推理： <think> 查找专家比较 </think>。

蓝色：模型生成文本；灰色：用户输入与工具输出。

深度研究智能体的推理过程

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Deep research agent inference

CONTEXT SO FAR

USER / QUESTION

Can agents match experts?

AGENT / REASONING

Find expert comparisons.

↓

Language model agent

↓

2 / GENERATE A TOOL CALL

<tool>

search(

"expert comparisons"

)

</tool>

Blue: model-generated text

Gray: user input and tool output

28

深度研究智能体的推理过程

03 / 智能体建模

深度研究智能体推理过程。

当前上下文包含用户问题 “智能体能与专家相匹配吗？” 及智能体推理 “查找专家比较” 。

语言模型智能体。

2 / 生成工具调用：<tool> search("expert comparisons") </tool>。

蓝色：模型生成文本；灰色：用户输入与工具输出。

译注 / 与原文翻译分开

工具调用代码与 API 字符串保留原样。

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 55 | 原页 53

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Deep research agent inference

CONTEXT SO FAR

USER / QUESTION
Can agents
match experts?

AGENT / REASONING
Find expert
comparisons.

AGENT / TOOL CALL
search("expert
comparisons")

Search

Language model agent

3 / EXECUTE THE SEARCH CALL

The search tool runs the generated query.

Blue: model-generated text Gray: user input and tool output

28

03 / 智能体建模

深度研究智能体推理过程。

当前上下文：用户问题、智能体推理，以及工具调用 `search("expert comparisons")`。

搜索。

3 / 执行搜索调用：搜索工具运行生成的查询。

蓝色：模型生成文本；灰色：用户输入与工具输出。

译注 / 与原文翻译分开

工具调用代码保留原样。

深度研究智能体的推理过程

第 10 讲 · 深度研究智能体：任务、基准与方法

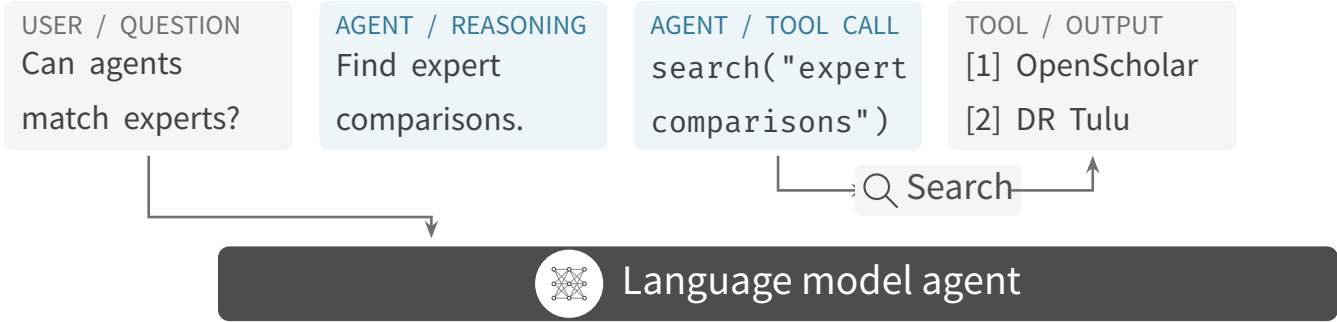
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Deep research agent inference

CONTEXT SO FAR



4 / APPEND THE TOOL OUTPUT

Retrieved text becomes part of the agent’s context.

Blue: model-generated text Gray: user input and tool output

03 / 智能体建模

深度研究智能体推理过程。

当前上下文：用户问题、智能体推理、工具调用，以及工具输出：[1] OpenScholar；[2] DR Tulu。

搜索。

4 / 附加工具输出：检索到的文本成为智能体上下文的一部分。

蓝色：模型生成文本；灰色：用户输入与工具输出。

深度研究智能体的推理过程

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Deep research agent inference

CONTEXT SO FAR

USER / QUESTION

Can agents match experts?

AGENT / REASONING

Find expert comparisons.

AGENT / TOOL CALL

search("expert comparisons")

TOOL / OUTPUT

[1] OpenScholar
[2] DR Tulu

Q Search

Language model agent

5 / CONTINUE REASONING WITH THE RETRIEVED TEXT

<think>

Check

the

tested

tasks

</think>

Blue: model-generated text

Gray: user input and tool output

28

03 / 智能体建模

深度研究智能体推理过程。

当前上下文包含：用户问题、智能体推理、工具调用、工具输出 [1] OpenScholar、[2] DR Tulu。

5 / 使用检索文本继续推理：<think> 检查已测试的任务 </think>。

蓝色：模型生成文本；灰色：用户输入与工具输出。

深度研究智能体的推理过程

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Deep research agent inference

CONTEXT SO FAR

USER / QUESTION

Can agents match experts?

AGENT / REASONING

Find expert comparisons.

AGENT / TOOL CALL

search("expert comparisons")

TOOL / OUTPUT

[1] OpenScholar
[2] DR Tulu

AGENT / REASONING

Check the tested tasks.

Search

Language model agent

6 / GENERATE AN ANSWER WITH CITATIONS

<answer>

Promising on tested tasks

[1,2]

...

</answer>

Blue: model-generated text

Gray: user input and tool output

28

深度研究智能体的推理过程

03 / 智能体建模

深度研究智能体推理过程。

当前上下文还包含智能体推理 “检查已测试的任务” 。

6 / 生成带引用的回答：<answer> 在已测试任务上表现有希望 [1,2] ... </answer>。

蓝色：模型生成文本；灰色：用户输入与工具输出。

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 59 | 原页 57

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

A training recipe for deep research

Learn from teacher demonstrations, then from your own attempts.

Mid-training

Learn agent behavior at scale.

TRAINING DATA

 Input + answer

 Teacher trajectories

Tongyi DeepResearch Technical Report. arXiv, 2025.

03 / 智能体建模

深度研究的训练方案。

先从教师示范中学习，再从自身尝试中学习。

中期训练：大规模学习智能体行为。训练数据：输入 + 答案、教师轨迹。

Tongyi DeepResearch Technical Report. arXiv, 2025.

深度研究的训练方案

第 10 讲 · 深度研究智能体：任务、基准与方法

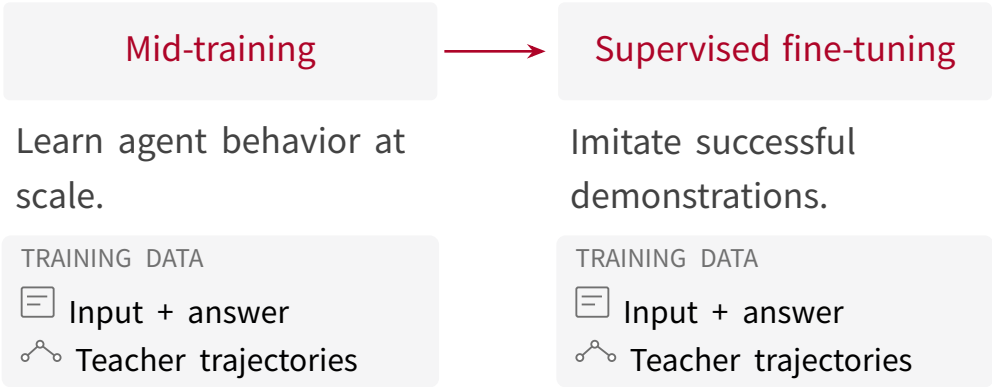
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

A training recipe for deep research

Learn from teacher demonstrations, then from your own attempts.



Tongyi DeepResearch Technical Report. arXiv, 2025.

03 / 智能体建模

深度研究的训练方案。

先从教师示范中学习，再从自身尝试中学习。

中期训练：大规模学习智能体行为。训练数据：输入 + 答案、教师轨迹。

监督微调：模仿成功示范。训练数据：输入 + 答案、教师轨迹。

Tongyi DeepResearch Technical Report. arXiv, 2025.

深度研究的训练方案

第 10 讲 · 深度研究智能体：任务、基准与方法

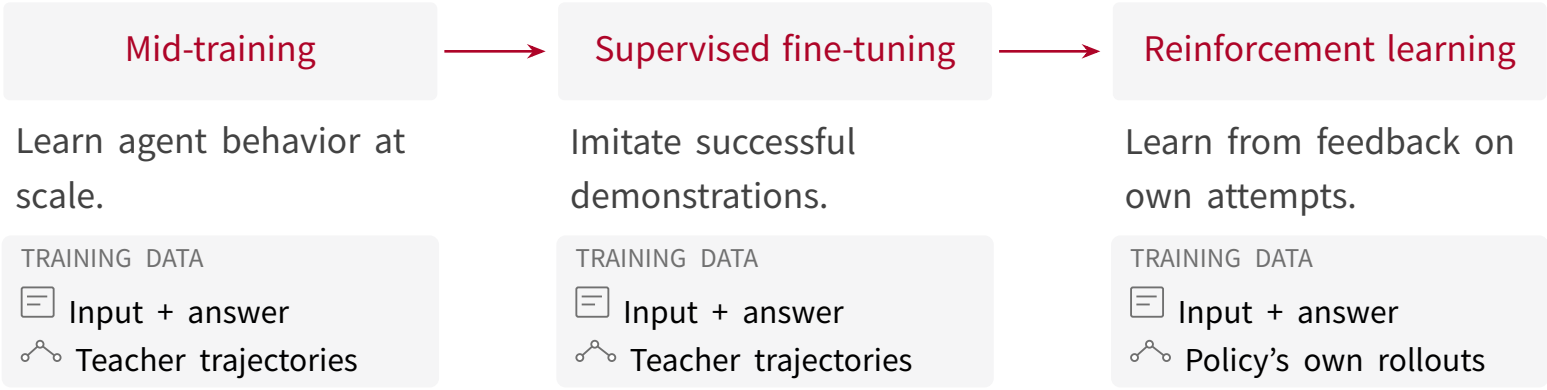
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

A training recipe for deep research

Learn from teacher demonstrations, then from your own attempts.



Next: where do the tasks and trajectories come from?

Tongyi DeepResearch Technical Report. arXiv, 2025.

03 / 智能体建模

深度研究的训练方案。

先从教师示范中学习，再从自身尝试中学习。

中期训练：大规模学习智能体行为。训练数据：输入 + 答案、教师轨迹。

监督微调：模仿成功示范。训练数据：输入 + 答案、教师轨迹。

强化学习：从自身尝试获得反馈并学习。训练数据：输入 + 答案、策略自身的采样轨迹。

接下来：任务与轨迹从哪里来？

Tongyi DeepResearch Technical Report. arXiv, 2025.

深度研究的训练方案

第 10 讲 · 深度研究智能体：任务、基准与方法

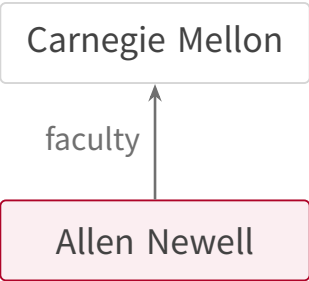
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Knowledge graphs: entities and relationships

A node is an **entity**; a labeled arrow is a **relationship**.



Subject: Allen Newell
Relationship: faculty member of
Object: Carnegie Mellon

CMU Archives; Nobel Prize (1978).

03 / 智能体建模

知识图谱：实体与关系。

节点是实体；带标签的箭头表示关系。

主体：Allen Newell；关系：Carnegie Mellon 的教员；宾语：Carnegie Mellon。

CMU 档案；诺贝尔奖（1978年）。

知识图谱：实体与关系

第 10 讲 · 深度研究智能体：任务、基准与方法

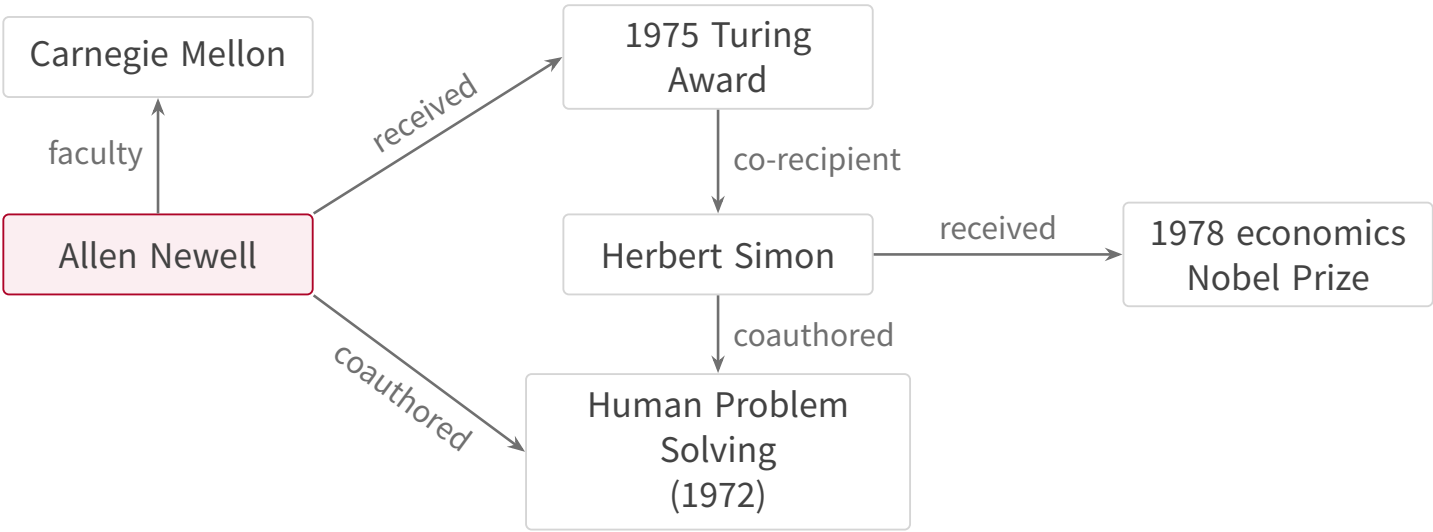
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Knowledge graphs: entities and relationships

A node is an **entity**; a labeled arrow is a **relationship**.



CMU Archives; Nobel Prize (1978).

30

03 / 智能体建模

知识图谱：实体与关系。

节点是实体；带标签的箭头表示关系。

Allen Newell 是 Carnegie Mellon 的教员；他于1975年获得图灵奖；与 Herbert Simon 共同获得该奖。Simon 于1978年获得诺贝尔经济学奖。Newell 与 Simon 合著了《Human Problem Solving》（1972）。

CMU 档案；诺贝尔奖（1978年）。

知识图谱：实体与关系

第 10 讲 · 深度研究智能体：任务、基准与方法

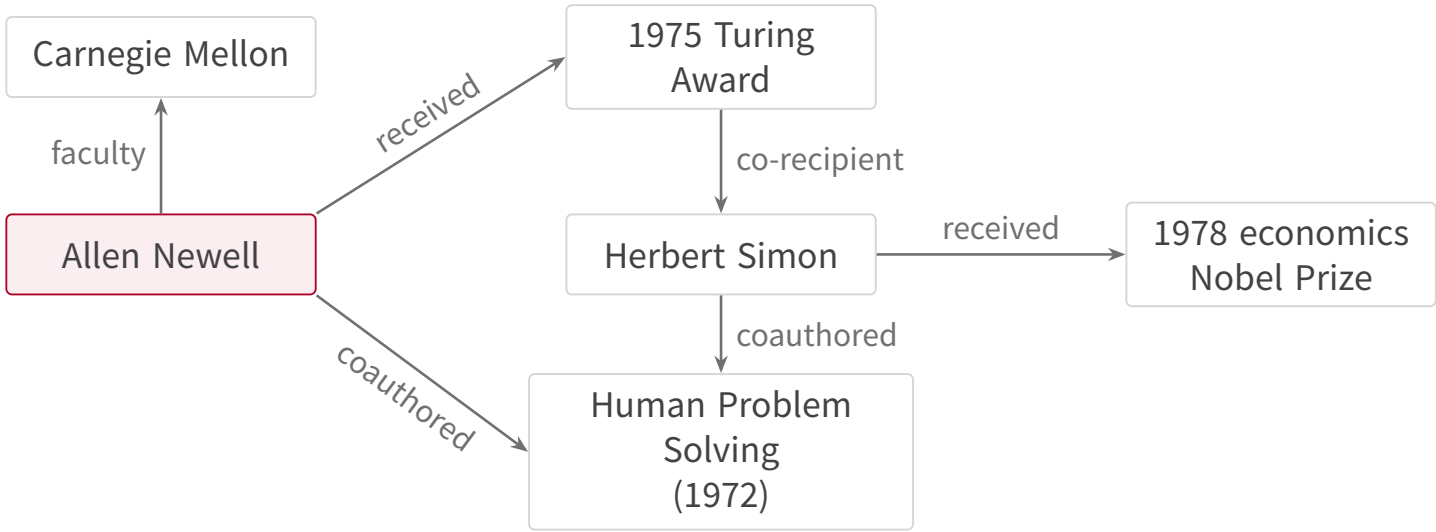
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Data creation: sample a knowledge graph

Start with sourced facts; sample a connected subgraph to create a question.



Method: WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; facts: CMU Archives, Nobel Prize (1978).

31

03 / 智能体建模

数据创建：从知识图谱采样。

从有来源的事实开始，采样一个连通子图来构造问题。

图中事实：Allen Newell 是 Carnegie Mellon 的教员；Newell 与 Herbert Simon 共同获得1975年图灵奖并合著《Human Problem Solving》（1972）；Simon 于1978年获得诺贝尔经济学奖。

方法：WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; 事实来源：CMU 档案、诺贝尔奖（1978年）。

数据创建：从知识图谱采样

第 10 讲 · 深度研究智能体：任务、基准与方法

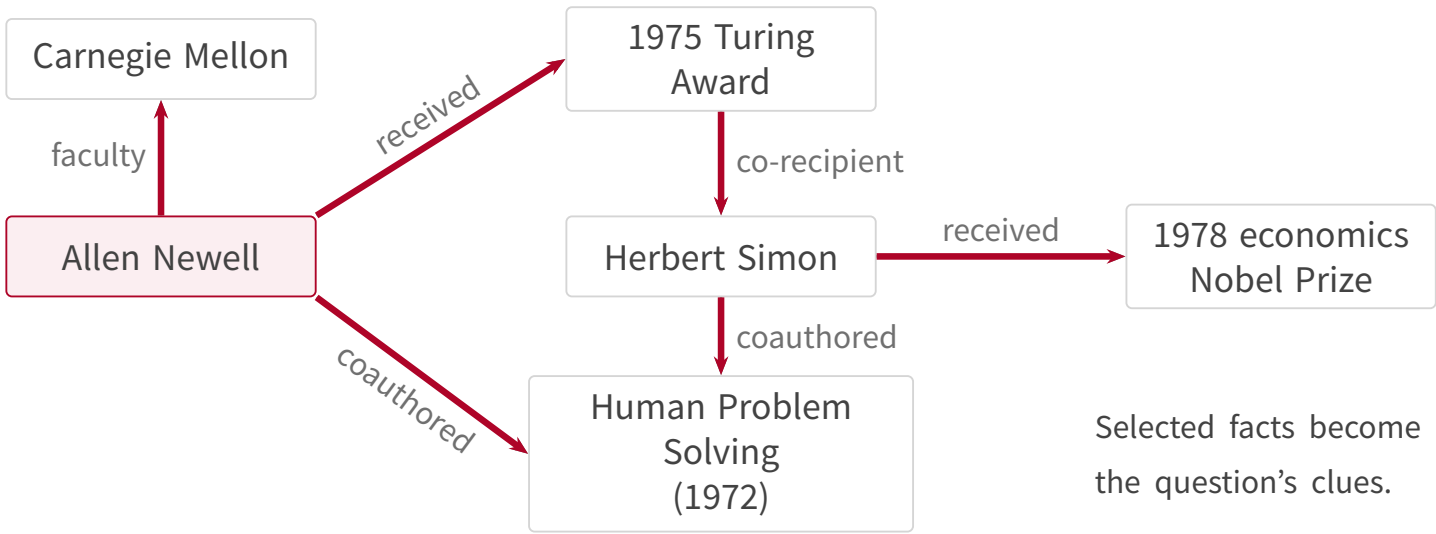
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Data creation: sample a knowledge graph

Start with sourced facts; sample a connected subgraph to create a question.



Method: WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; facts: CMU Archives, Nobel Prize (1978).

31

03 / 智能体建模

数据创建：从有来源的事实开始，采样一个连通子图来构造问题。

图中事实：Allen Newell 是 Carnegie Mellon 的教员；Newell 与 Herbert Simon 共同获得1975年图灵奖并合著《Human Problem Solving》（1972）；Simon 于1978年获得诺贝尔经济学奖。

选定的事实将成为问题的线索。

方法：WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; 事实来源：CMU 档案、诺贝尔奖（1978年）。

数据创建：从知识图谱采样

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Data creation: turn facts into a question

Hide the target name and turn its relations into clues.

FACTS THE ANSWER MUST SATISFY

- Newell: faculty at CMU.
- Newell + Simon: coauthors of Human Problem Solving (1972).
- Newell + Simon: shared the 1975 Turing Award.
- Simon: 1978 economics Nobel Prize.

Method: WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; facts: CMU Archives, Nobel Prize (1978).

03 / 智能体建模

数据创建：将事实转化为问题。

隐藏目标人物的名字，将人物关系转为线索。

答案必须满足的事实：Newell 是 CMU 教员；Newell 与 Simon 合著《Human Problem Solving》（1972）；Newell 与 Simon 共同获得1975年图灵奖；Simon 于1978年获得诺贝尔经济学奖。

方法：WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.；事实来源：CMU 档案、诺贝尔奖（1978年）。

数据创建：将事实转化为问题

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

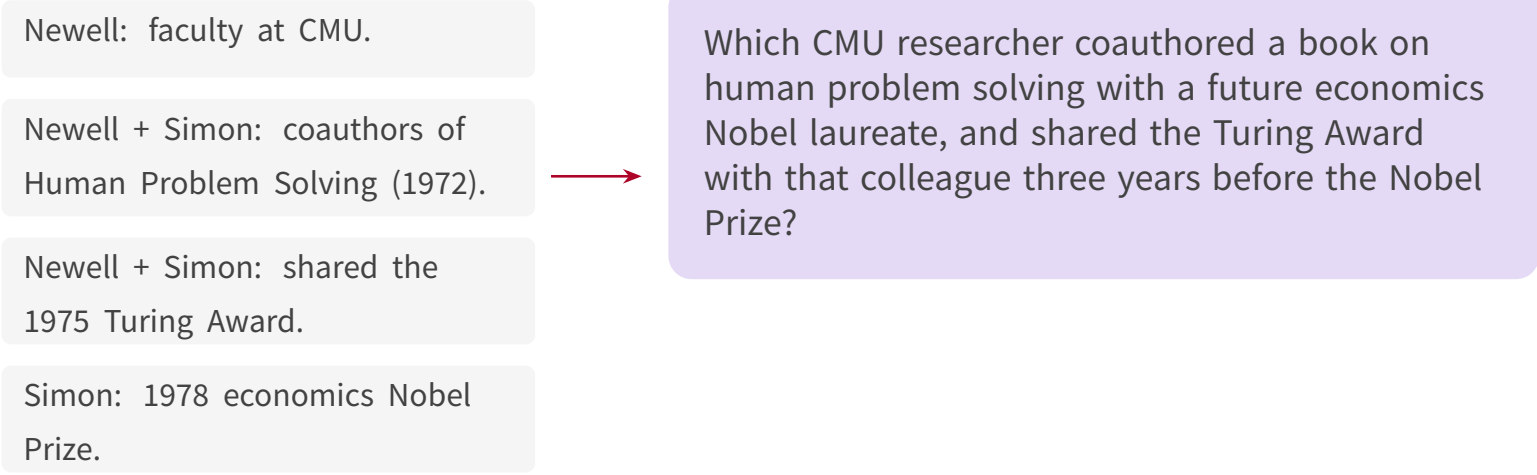
ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Data creation: turn facts into a question

Hide the target name and turn its relations into clues.

FACTS THE ANSWER MUST SATISFY



Method: WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; facts: CMU Archives, Nobel Prize (1978).

32

03 / 智能体建模

隐藏目标人物的名字，将人物关系转为线索。

答案必须满足的事实：Newell 是 CMU 教员；Newell 与 Simon 合著《Human Problem Solving》（1972）；两人共同获得1975年图灵奖；Simon 于1978年获得诺贝尔经济学奖。

问题：哪位 CMU 研究人员与一位未来的诺贝尔经济学奖得主合著了一本关于人类问题求解的书，并在该同事获得诺贝尔奖前三年与其共同获得图灵奖？

方法：WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; 事实来源：CMU 档案、诺贝尔奖（1978年）。

数据创建：将事实转化为问题

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

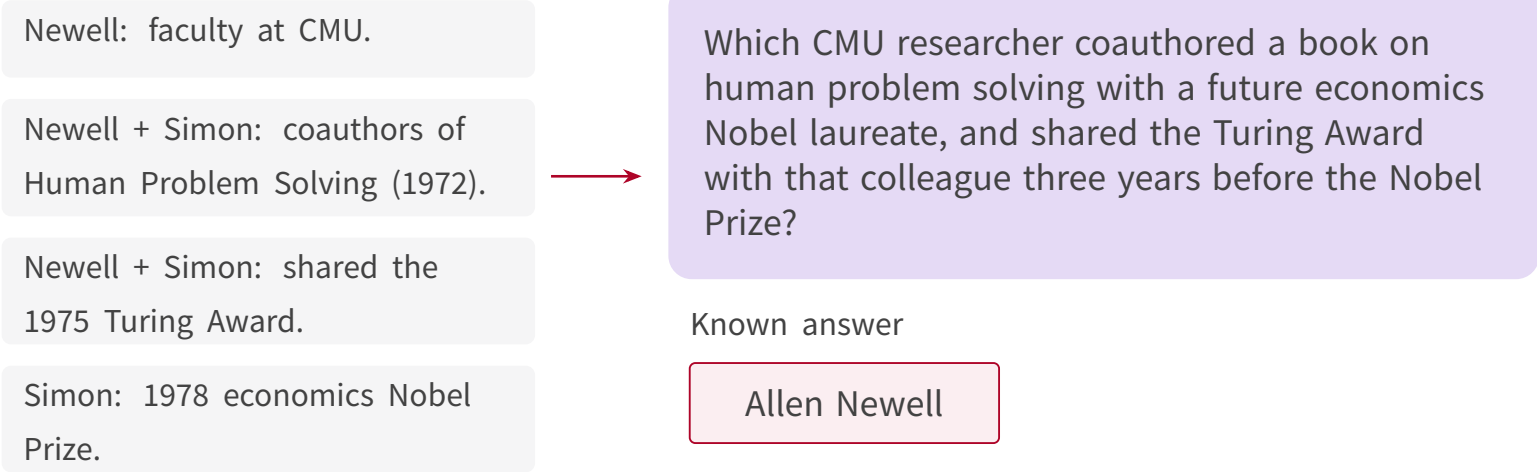
ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Data creation: turn facts into a question

Hide the target name and turn its relations into clues.

FACTS THE ANSWER MUST SATISFY



Method: WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; facts: CMU Archives, Nobel Prize (1978).

32

03 / 智能体建模

隐藏目标人物的名字，将人物关系转为线索。

答案必须满足的事实：Newell 是 CMU 教员；Newell 与 Simon 合著《Human Problem Solving》（1972）；两人共同获得1975年图灵奖；Simon 于1978年获得诺贝尔经济学奖。

问题：哪位 CMU 研究人员与一位未来的诺贝尔经济学奖得主合著了一本关于人类问题求解的书，并在该同事获得诺贝尔奖前三年与其共同获得图灵奖？

已知答案：Allen Newell。

方法：WebSailor-V2: Bridging the Chasm to Proprietary Agents via Synthetic Data and Scalable Reinforcement Learning. ICLR 2026.; 事实来源：CMU 档案、诺贝尔奖（1978年）。

数据创建：将事实转化为问题

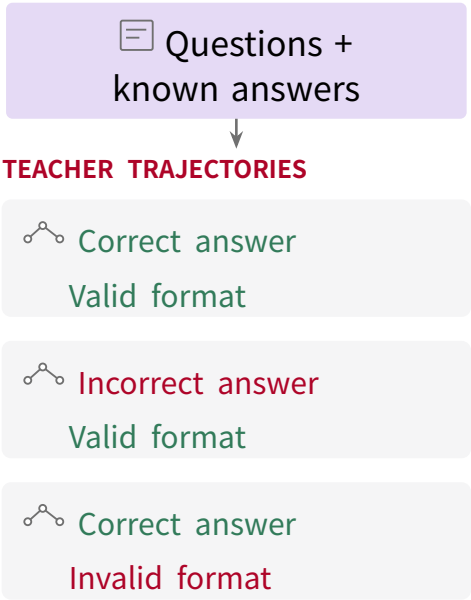
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

SFT: collect, filter, then imitate trajectories



WebDancer: Towards Autonomous Information Seeking Agency. NeurIPS 2025.; Tongyi DeepResearch Technical Report. arXiv, 2025.

03 / 智能体建模

监督微调：收集、筛选，然后模仿轨迹。

问题 + 已知答案。

教师轨迹：答案正确、格式有效；答案错误、格式有效；答案正确、格式无效。

WebDancer: Towards Autonomous Information Seeking Agency. NeurIPS 2025. · Tongyi DeepResearch Technical Report. arXiv, 2025.

监督微调：收集、筛选，然后模仿轨迹

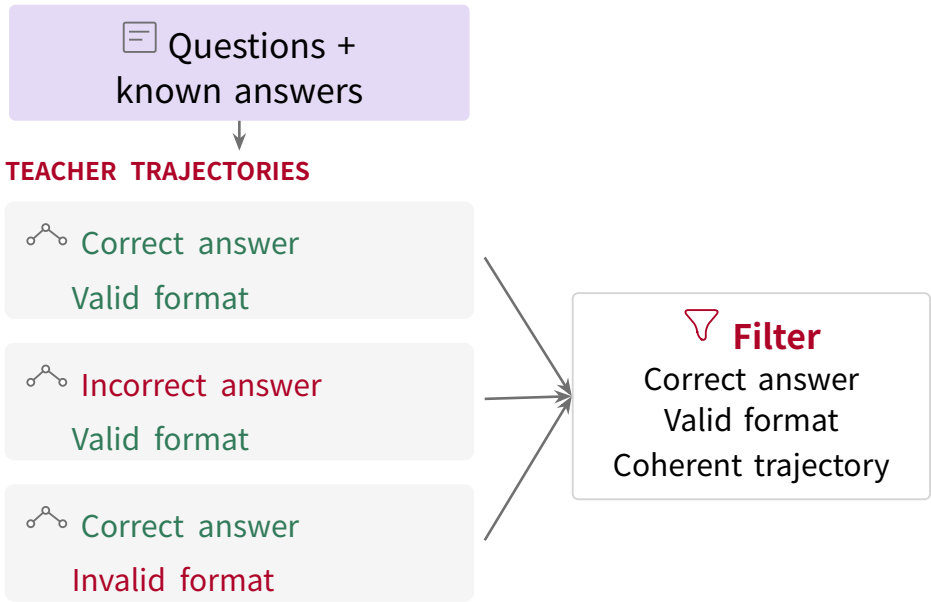
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

SFT: collect, filter, then imitate trajectories



WebDancer: Towards Autonomous Information Seeking Agency. NeurIPS 2025.; Tongyi DeepResearch Technical Report. arXiv, 2025.

03 / 智能体建模

监督微调：收集、筛选，然后模仿轨迹。

问题 + 已知答案 → 教师轨迹。

教师轨迹示例：答案正确、格式有效；答案错误、格式有效；答案正确、格式无效。

筛选后保留：答案正确、格式有效；轨迹连贯。

WebDancer: Towards Autonomous Information Seeking Agency. NeurIPS 2025. · Tongyi DeepResearch Technical Report. arXiv, 2025.

监督微调：收集、筛选，然后模仿轨迹

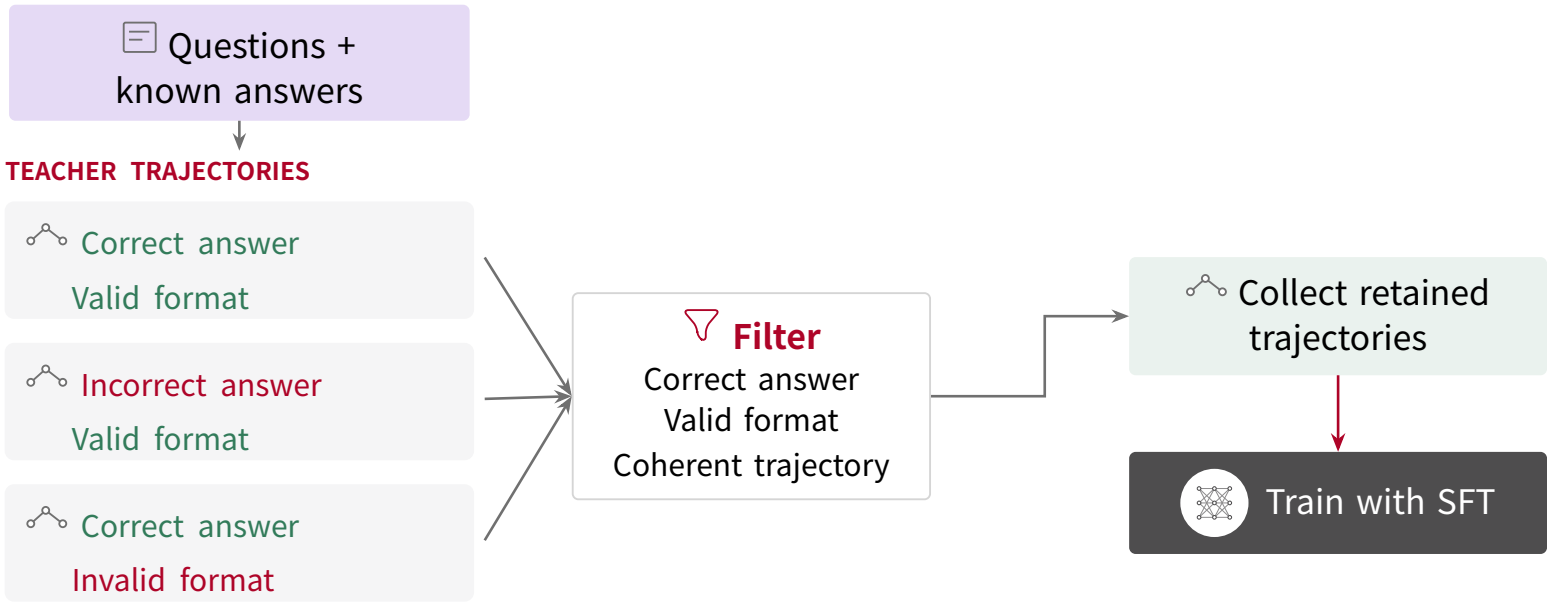
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

SFT: collect, filter, then imitate trajectories



WebDancer: Towards Autonomous Information Seeking Agency. NeurIPS 2025.; Tongyi DeepResearch Technical Report. arXiv, 2025.

33

03 / 智能体建模

监督微调：收集、筛选，然后模仿轨迹。

问题 + 已知答案 → 教师轨迹 → 筛选。

教师轨迹：答案正确、格式有效；答案错误、格式有效；答案正确、格式无效。

筛选后保留：答案正确、格式有效；轨迹连贯。

收集保留的轨迹；使用监督微调（SFT）进行训练。

WebDancer: Towards Autonomous Information Seeking Agency. NeurIPS 2025. · Tongyi DeepResearch Technical Report. arXiv, 2025.

监督微调：收集、筛选，然后模仿轨迹

第 10 讲 · 深度研究智能体：任务、基准与方法

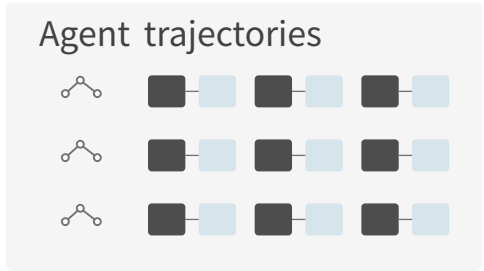
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Agentic mid-training

Tongyi uses continual pretraining to prepare the base model for agent workflows.



Scaling Agents via Continual Pre-training. ICLR 2026.

34

03 / 智能体建模

智能体中期训练。

Tongyi 使用持续预训练，让基础模型为智能体 workflow 做好准备。

智能体轨迹；通用预训练数据。

Scaling Agents via Continual Pre-training. ICLR 2026.

智能体中期训练

第 10 讲 · 深度研究智能体：任务、基准与方法

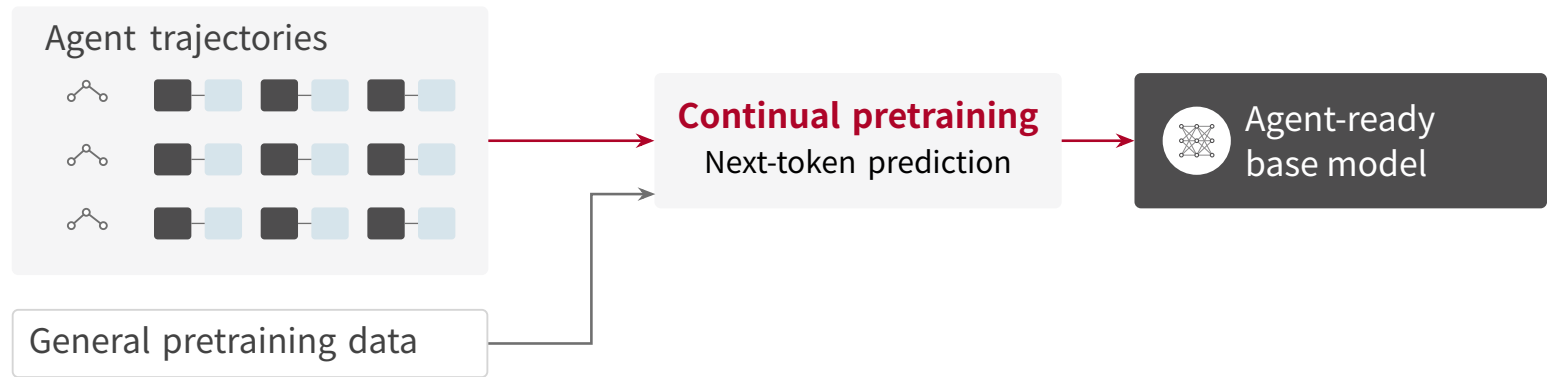
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Agentic mid-training

Tongyi uses continual pretraining to prepare the base model for agent workflows.



Scaling Agents via Continual Pre-training. ICLR 2026.

34

03 / 智能体建模

智能体中期训练。

Tongyi 使用持续预训练，让基础模型为智能体 workflow 做好准备。

智能体轨迹 + 通用预训练数据 → 持续预训练（下一词元预测）→ 适用于智能体的基础模型。

Scaling Agents via Continual Pre-training. ICLR 2026.

智能体中期训练

第 10 讲 · 深度研究智能体：任务、基准与方法

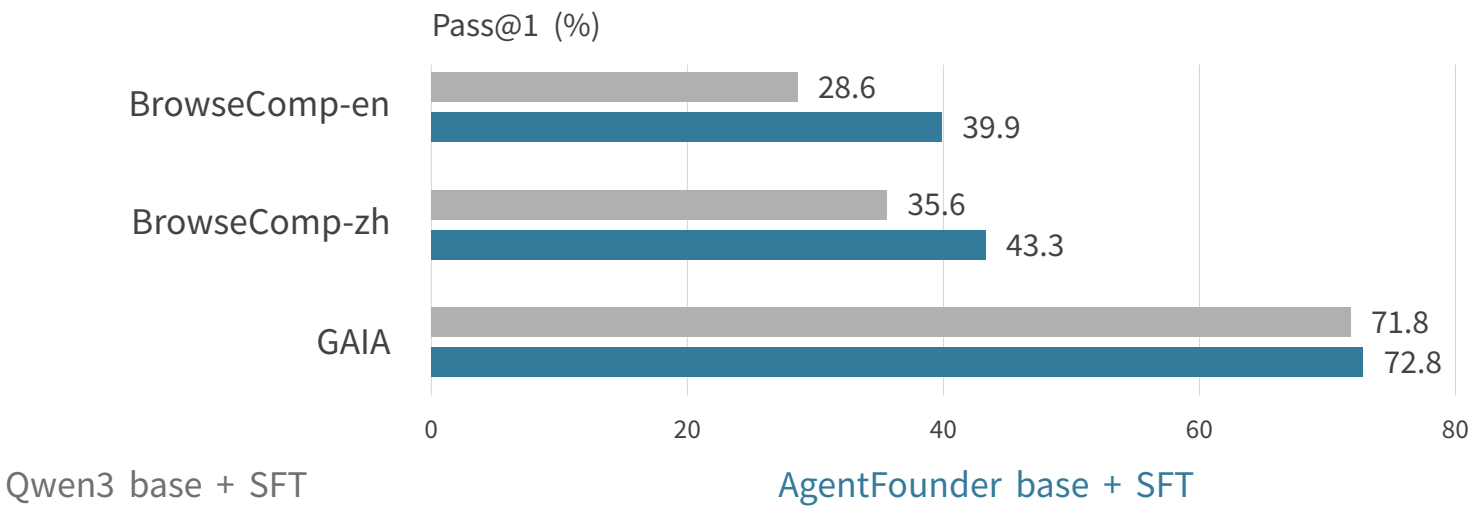
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Mid-training improves the starting point for SFT

Same downstream SFT-B data; different base-model initialization.



Scaling Agents via Continual Pre-training. ICLR 2026.

35

03 / 智能体建模

中期训练改善监督微调的起点。

下游 SFT-B 数据相同；基础模型初始化不同。

Pass@1 (%) : BrowseComp-en, Qwen3 base + SFT 28.6、AgentFounder base + SFT 39.9; BrowseComp-zh, 35.6、43.3; GAIA, 71.8、72.8。

Scaling Agents via Continual Pre-training. ICLR 2026.

中期训练改善监督微调的起点

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

RLVR: learning from answer correctness



Search-R1: Training LLMs to Reason and Leverage Search Engines with Reinforcement Learning. COLM 2025.: final-answer exact-match reward. CMU examples and rewards are illustrative.

36

03 / 智能体建模

RLVR：从答案正确性中学习。

问题 x ：CMU 线索问题 $\rightarrow$ 策略 π_θ （搜索并推理） $\rightarrow$ 采样搜索轨迹 y_1 、 y_2 、 y_3 。

轨迹答案：Herbert Simon；Allen Newell；Allen Newell。

Search-R1: Training LLMs to Reason and Leverage Search Engines with Reinforcement Learning. COLM 2025.: 最终答案精确匹配奖励。CMU 示例和奖励仅用于说明。

RLVR：从答案正确性中学习

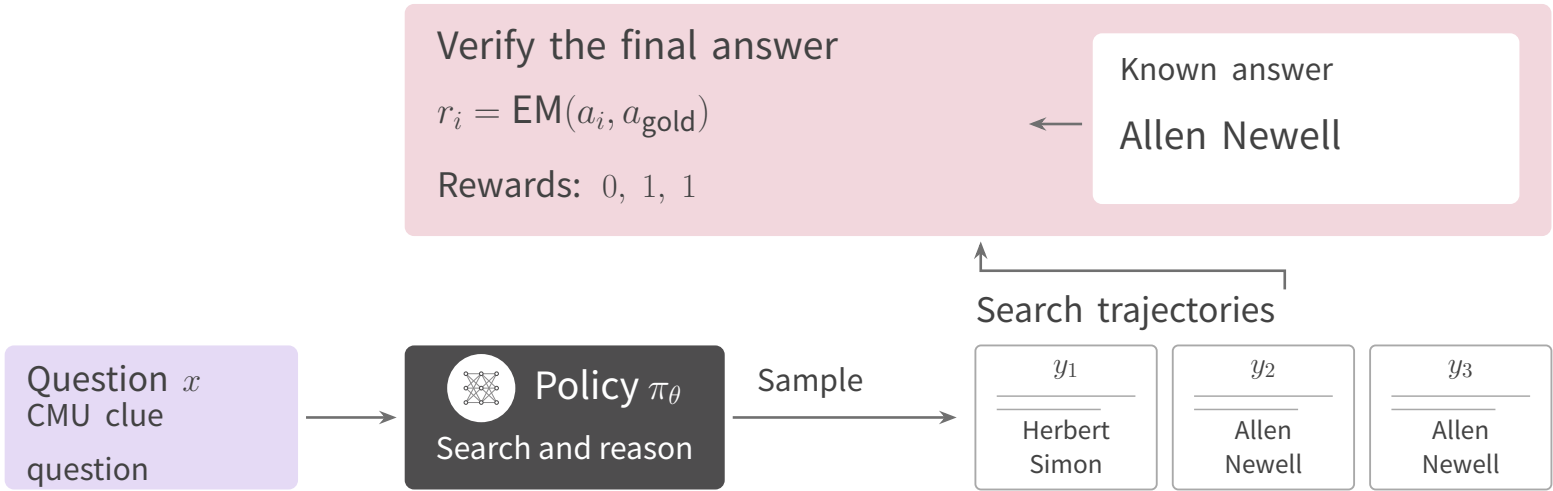
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

RLVR: learning from answer correctness



Search-R1: Training LLMs to Reason and Leverage Search Engines with Reinforcement Learning. COLM 2025.: final-answer exact-match reward. CMU examples and rewards are illustrative.

36

03 / 智能体建模

RLVR：从答案正确性中学习。

验证最终答案：已知答案 Allen Newell； $r_i = \text{EM}(a_i, a_{\text{gold}})$ 。奖励：0、1、1。

问题 x ：CMU 线索问题 → 策略 π_θ （搜索并推理）→ 采样搜索轨迹 y_1 、 y_2 、 y_3 ；答案分别为 Herbert Simon、Allen Newell、Allen Newell。

Search-R1: Training LLMs to Reason and Leverage Search Engines with Reinforcement Learning. COLM 2025.: 最终答案精确匹配奖励。CMU 示例和奖励仅用于说明。

译注 / 与原文翻译分开

公式和变量保留原样。

RLVR：从答案正确性中学习

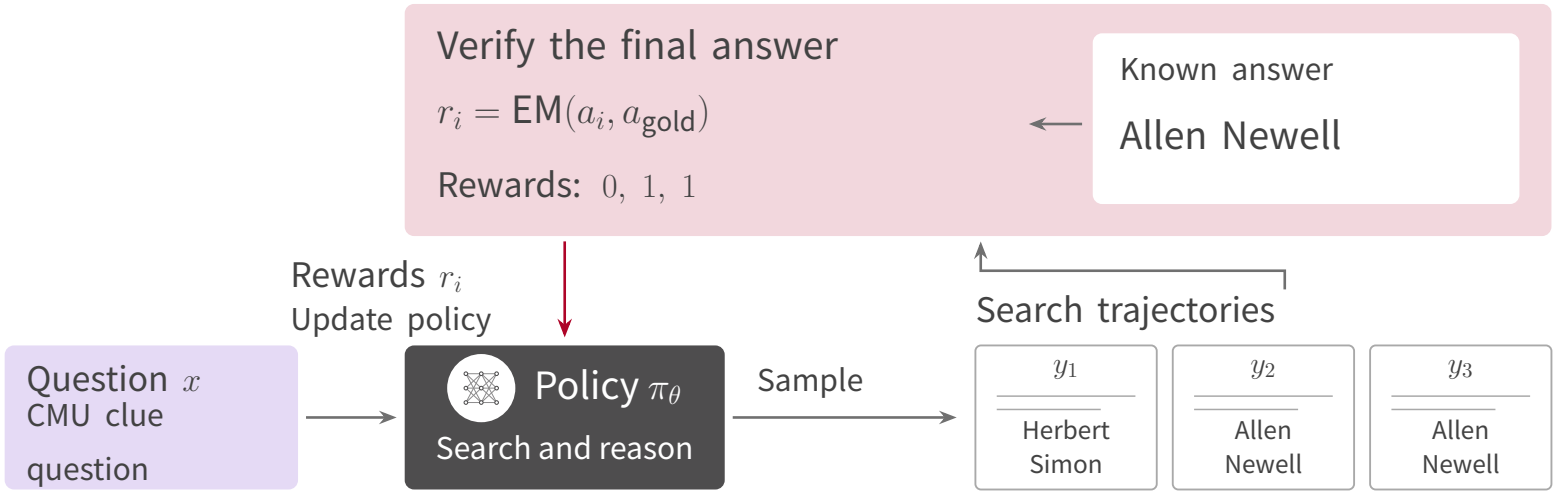
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

RLVR: learning from answer correctness



Search-R1: Training LLMs to Reason and Leverage Search Engines with Reinforcement Learning. COLM 2025.: final-answer exact-match reward. CMU examples and rewards are illustrative.

36

03 / 智能体建模

RLVR：从答案正确性中学习。

验证最终答案：已知答案 Allen Newell； $r_i = \text{EM}(a_i, a_{\text{gold}})$ 。奖励：0、1、1。

将奖励 r_i 用于更新策略。

问题 x ：CMU 线索问题 → 策略 π_θ （搜索并推理）→ 采样搜索轨迹 y_1 、 y_2 、 y_3 ；答案分别为 Herbert Simon、Allen Newell、Allen Newell。

Search-R1: Training LLMs to Reason and Leverage Search Engines with Reinforcement Learning. COLM 2025.: 最终答案精确匹配奖励。CMU 示例和奖励仅用于说明。

译注 / 与原文翻译分开

公式和变量保留原样。

RLVR：从答案正确性中学习

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

GRPO: learn from better attempts

One question x ; sample a group of attempts $y_1, \dots, y_G$

Herbert Simon
Reward: 0

Allen Newell
Reward: 1

Allen Newell
Reward: 1

DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv, 2024.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

37

03 / 智能体建模

GRPO：从更好的尝试中学习。

对同一个问题 x 采样一组尝试 $y_1, \dots, y_G$ 。

Herbert Simon：奖励0； Allen Newell：奖励1； Allen Newell：奖励1。

DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv, 2024. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

GRPO：从更好的尝试中学习

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

GRPO: learn from better attempts

One question x ; sample a group of attempts $y_1, \dots, y_G$

Herbert Simon Reward: 0	Allen Newell Reward: 1	Allen Newell Reward: 1
Below group average	Above group average	Above group average
Lower its probability	Raise its probability	Raise its probability

DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv, 2024.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

37

03 / 智能体建模

GRPO：从更好的尝试中学习。

对同一个问题 x 采样一组尝试 $y_1, \dots, y_G$ 。

Herbert Simon：奖励0， 低于组平均值——降低其概率。

Allen Newell：奖励1， 高于组平均值——提高其概率。

Allen Newell：奖励1， 高于组平均值——提高其概率。

DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv, 2024. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

GRPO：从更好的尝试中学习

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

GRPO: learn from better attempts

One question x ; sample a group of attempts $y_1, \dots, y_G$

Herbert Simon Reward: 0	Allen Newell Reward: 1	Allen Newell Reward: 1
Below group average Lower its probability	Above group average Raise its probability	Above group average Raise its probability
$\hat{A}$: relative reward Compare within the group	Clipped update Limit the update size	KL penalty Stay near the reference

DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv, 2024.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026..

37

03 / 智能体建模

对同一个问题 x 采样一组尝试 $y_1, \dots, y_G$ 。

Herbert Simon，奖励0：低于组平均值，降低其概率。Allen Newell，奖励1：高于组平均值，提高其概率（两次尝试）。

GRPO 目标（对模型词元取平均）： $J(\theta)=E[\min(p\hat{A}, \text{clip}(p,1-\epsilon,1+\epsilon)\hat{A}) - \beta D_{\text{KL}}]$ 。

$\hat{A}$ ：相对奖励，在组内比较。裁剪更新：限制更新幅度。KL 惩罚：保持接近参考模型。

DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv, 2024. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

公式按原图转写；期望下标为模型词元。

GRPO：学习更好的尝试

第 10 讲 · 深度研究智能体：任务、基准与方法

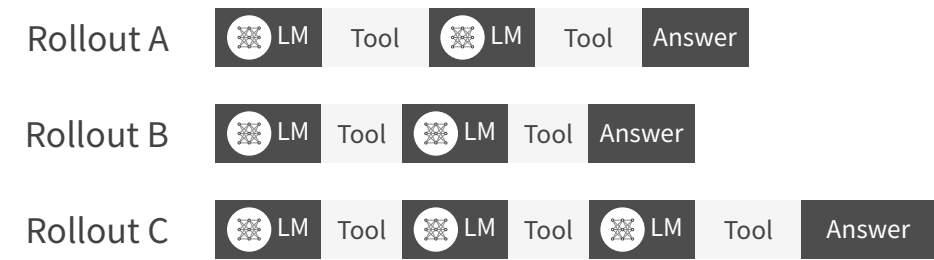
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Asynchronous rollouts: finish, score, then train

Score completed trajectories; assemble a batch for the policy update.



Tongyi DeepResearch Technical Report. arXiv, 2025.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

03 / 智能体建模

异步采样轨迹：完成、评分，然后训练。

对已完成的轨迹评分，并组装策略更新所需的批次。

轨迹 A：LM → 工具 → LM → 工具 → 答案。轨迹 B：LM → 工具 → LM → 工具 → 答案。轨迹 C：LM → 工具 → LM → 工具 → LM → 工具 → 答案。

Tongyi DeepResearch Technical Report. arXiv, 2025. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

异步采样轨迹：完成、评分，然后训练

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Asynchronous rollouts: finish, score, then train

Score completed trajectories; assemble a batch for the policy update.

Rollout A

LM

Tool

LM

Tool

Answer

Reward

Rollout B

LM

Tool

LM

Tool

Answer

Reward

Rollout C

LM

Tool

LM

Tool

LM

Tool

Answer

Reward

Completed batch
+ rewards
Policy update

Tongyi DeepResearch Technical Report. arXiv, 2025.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

38

03 / 智能体建模

异步采样轨迹：完成、评分，然后训练。

对已完成的轨迹评分，并组装策略更新所需的批次。

轨迹 A：LM → 工具 → LM → 工具 → 答案 → 奖励。轨迹 B：LM → 工具 → LM → 工具 → 答案 → 奖励。轨迹 C：LM → 工具 → LM → 工具 → LM → 工具 → 答案 → 奖励。

已完成批次 + 奖励 → 策略更新。

Tongyi DeepResearch Technical Report. arXiv, 2025. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

异步采样轨迹：完成、评分，然后训练

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 83 | 原页 81

第 10 讲 · 深度研究智能体：任务、基准与方法

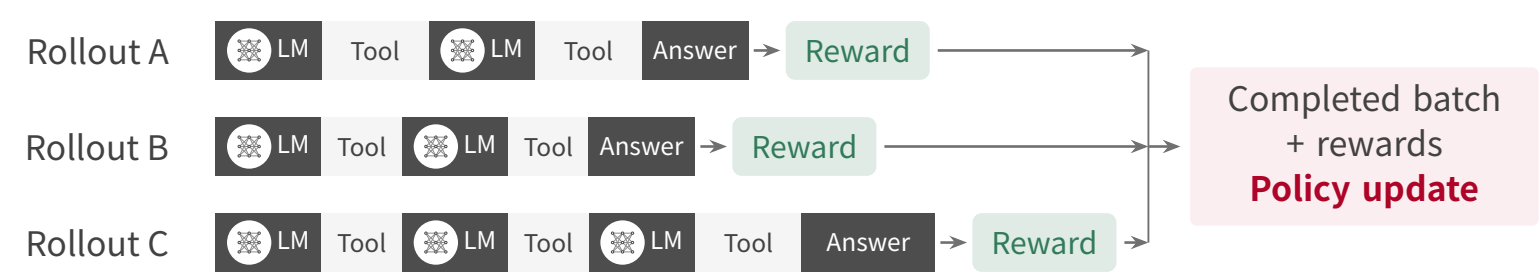
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Asynchronous rollouts: finish, score, then train

Score completed trajectories; assemble a batch for the policy update.



Additional efficiency: cached tool results, robust API handling, and background task curation

More on scheduling and training efficiency in the RL systems lecture

Tongyi DeepResearch Technical Report. arXiv, 2025.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

03 / 智能体建模

异步采样轨迹：完成、评分，然后训练。

对已完成的轨迹评分，并组装策略更新所需的批次。

轨迹 A、B、C 各自经过语言模型与工具交互、生成答案并获得奖励；完成批次及奖励用于更新策略。

其他效率改进：缓存工具结果、稳健处理 API，以及在后台整理任务。

更多有关调度和训练效率的内容见强化学习系统一讲。

Tongyi DeepResearch Technical Report. arXiv, 2025. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

异步采样轨迹：完成、评分，然后训练

第 10 讲 · 深度研究智能体：任务、基准与方法

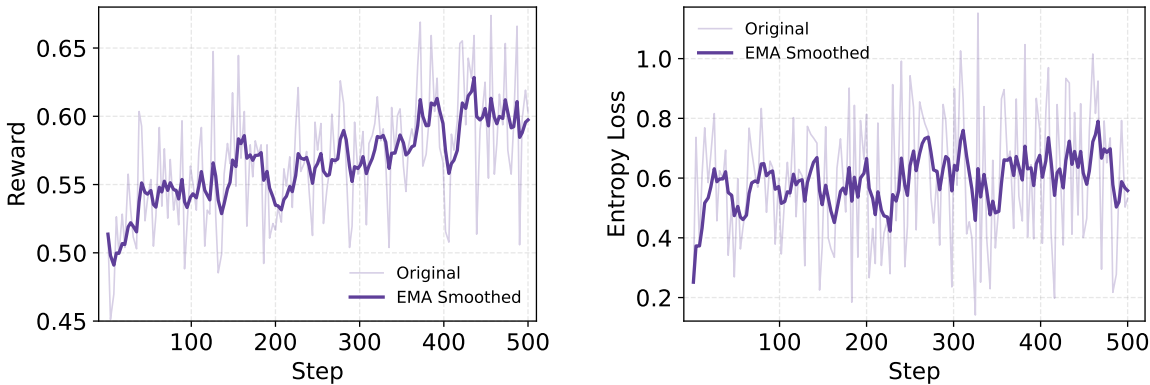
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Tongyi: RL improves with training

Training reward increases while policy entropy stays relatively stable.



These are RL training curves; the report does not isolate each training stage.

Tongyi DeepResearch Technical Report. arXiv, 2025..

03 / 智能体建模

Tongyi：强化学习训练带来提升。

训练奖励不断上升，而策略熵相对稳定。

左图：奖励；右图：熵损失。两图横轴均为训练步数（0–500）；显示 Original 与 EMA Smoothed 两条曲线。奖励约从0.50升至0.60；熵损失总体在约0.5–0.7 间波动。

这些是 RL 训练曲线；报告没有单独区分每个训练阶段。

Tongyi DeepResearch Technical Report. arXiv, 2025.

译注 / 与原文翻译分开

曲线未标注各点数值，文中范围为按坐标轴估读。

Tongyi：强化学习训练带来提升

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

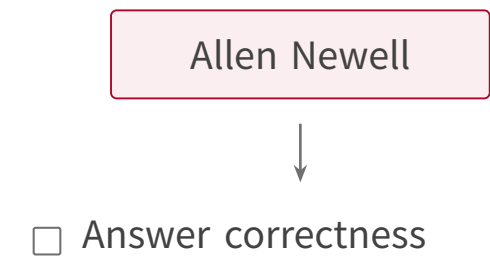
03 / AGENT MODELING

Open-ended reports need richer training signals

What should the reward measure?

VERIFIABLE SHORT ANSWER

OPEN-ENDED SYNTHESIS



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

40

03 / 智能体建模

开放式报告需要更丰富的训练信号。

奖励应衡量什么？

可验证的简短答案：Allen Newell → 答案正确性。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

开放式报告需要更丰富的训练信号

第 10 讲 · 深度研究智能体：任务、基准与方法

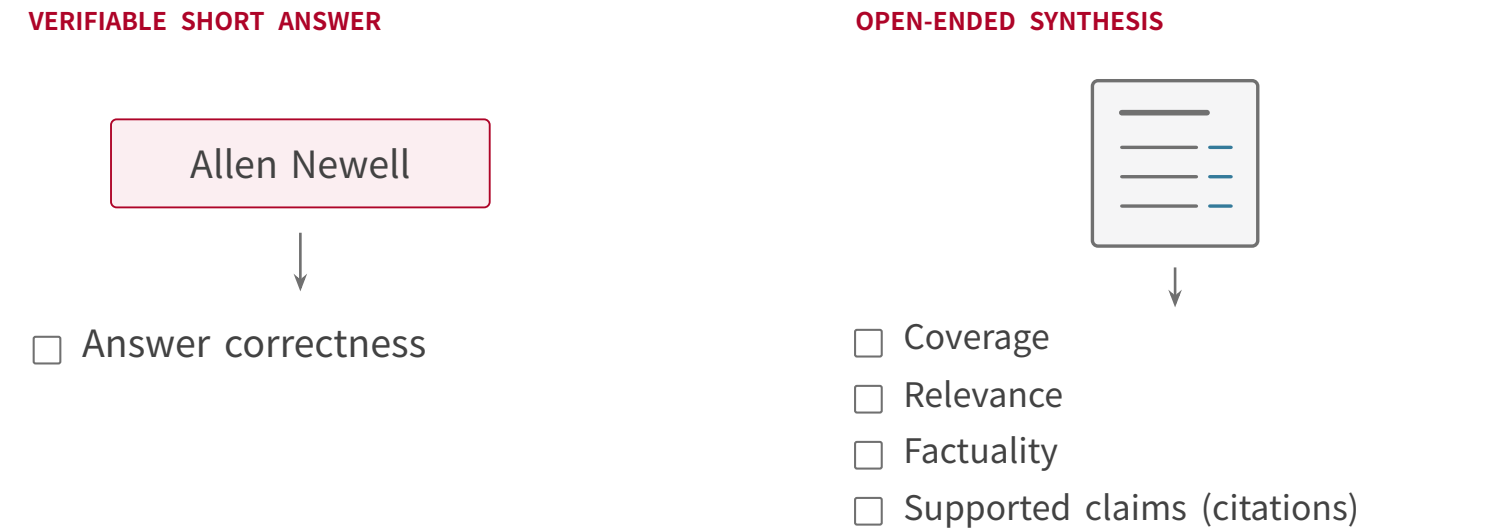
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

Open-ended reports need richer training signals

What should the reward measure?



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

40

03 / 智能体建模

开放式报告需要更丰富的训练信号。

奖励应衡量什么？

可验证的简短答案：答案正确性。

开放式综合：覆盖度、相关性、事实准确性、得到支持的主张（含引用）。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

开放式报告需要更丰富的训练信号

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

DR Tulu: evolving rubrics for RL



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.. Rubric reward shown; auxiliary rewards omitted.

41

03 / 智能体建模

DR Tulu：用于强化学习的演进式评分标准。

问题 x：文献综合 → 策略 π_θ （搜索并推理）→ 采样搜索轨迹 y_1 、 y_2 、 y_3 → 报告 A、报告 B、报告 C。

图中显示评分标准奖励；未显示辅助奖励。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

DR Tulu：用于强化学习的演进式评分标准

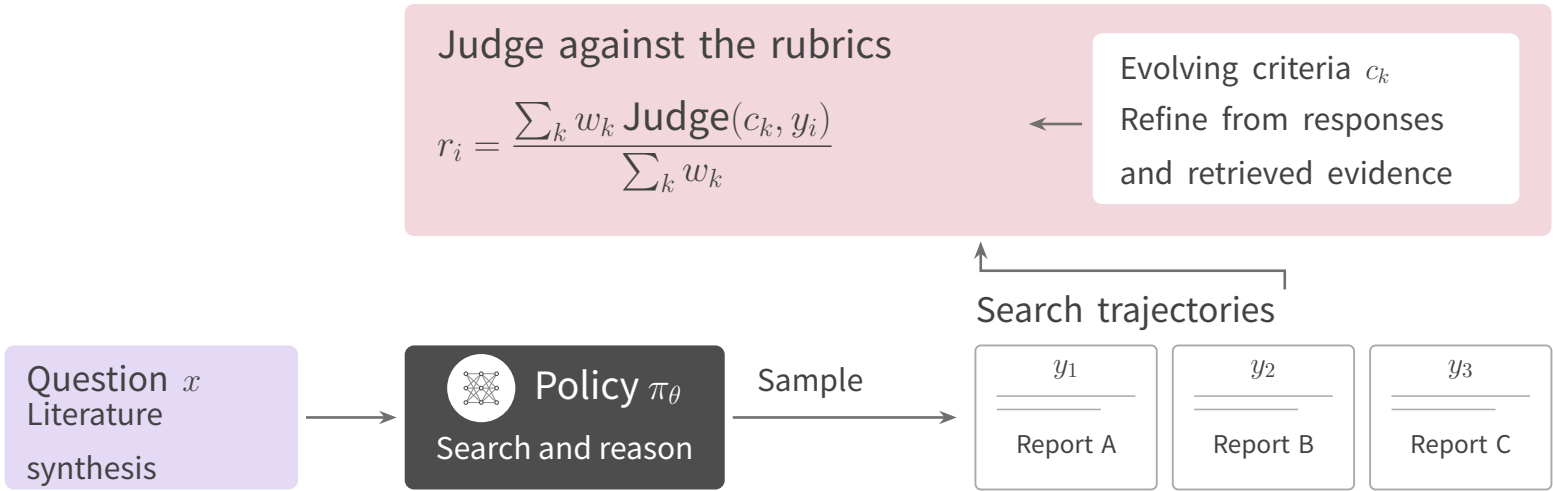
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

DR Tulu: evolving rubrics for RL



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.. Rubric reward shown; auxiliary rewards omitted.

41

03 / 智能体建模

DR Tulu：用于强化学习的演进式评分标准。

使用评分标准进行评判： $r_i = [\sum_k w_k \text{Judge}(c_k, y_i)] / [\sum_k w_k]$ 。

不断演进的标准 c_k ：根据回答和检索到的证据进行优化。

问题 x ：文献综合 $\rightarrow$ 策略 π_θ （搜索并推理） $\rightarrow$ 采样轨迹 y_1 、 y_2 、 $y_3 \rightarrow$ 报告 A、B、C。

图中显示评分标准奖励；未显示辅助奖励。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

公式与变量保留原样。

DR Tulu：用于强化学习的演进式评分标准

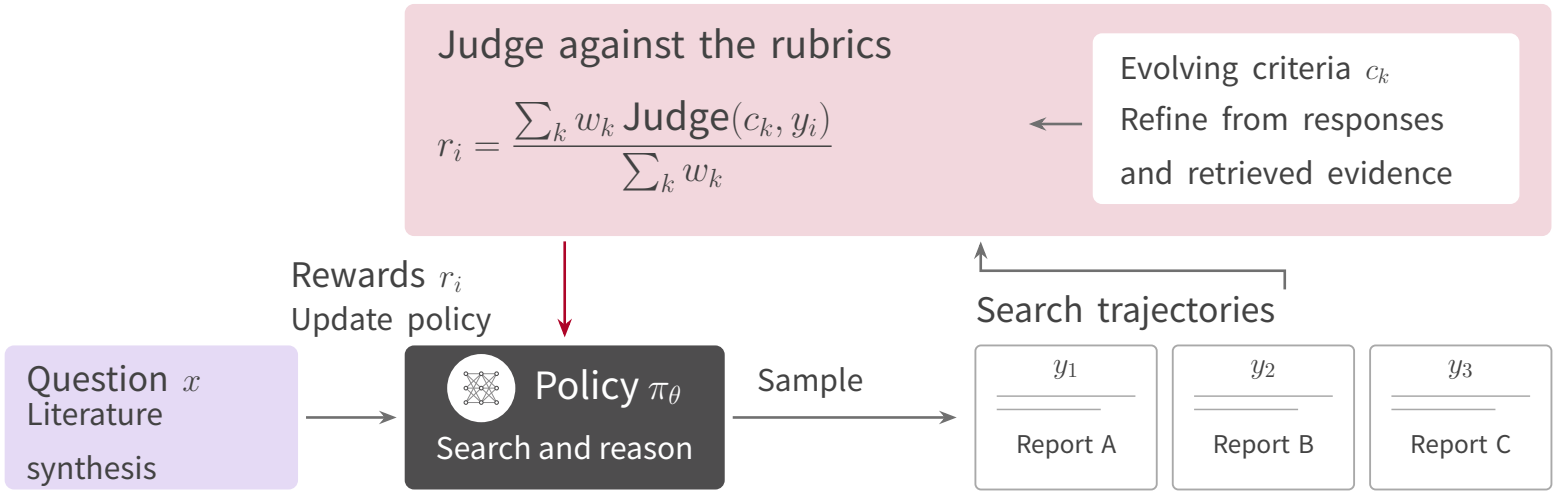
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

DR Tulu: evolving rubrics for RL



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.. Rubric reward shown; auxiliary rewards omitted.

41

03 / 智能体建模

DR Tulu：用于强化学习的演进式评分标准。

使用评分标准进行评判： $r_i = [\sum_k w_k \text{Judge}(c_k, y_i)] / [\sum_k w_k]$ 。

不断演进的标准 c_k ：根据回答和检索到的证据进行优化。

奖励 $r_i \rightarrow$ 更新策略。

问题 x ：文献综合 $\rightarrow$ 策略 π_θ （搜索并推理） $\rightarrow$ 采样轨迹 y_1 、 y_2 、 $y_3 \rightarrow$ 报告 A、B、C。

图中显示评分标准奖励；未显示辅助奖励。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

公式与变量保留原样。

DR Tulu：用于强化学习的演进式评分标准

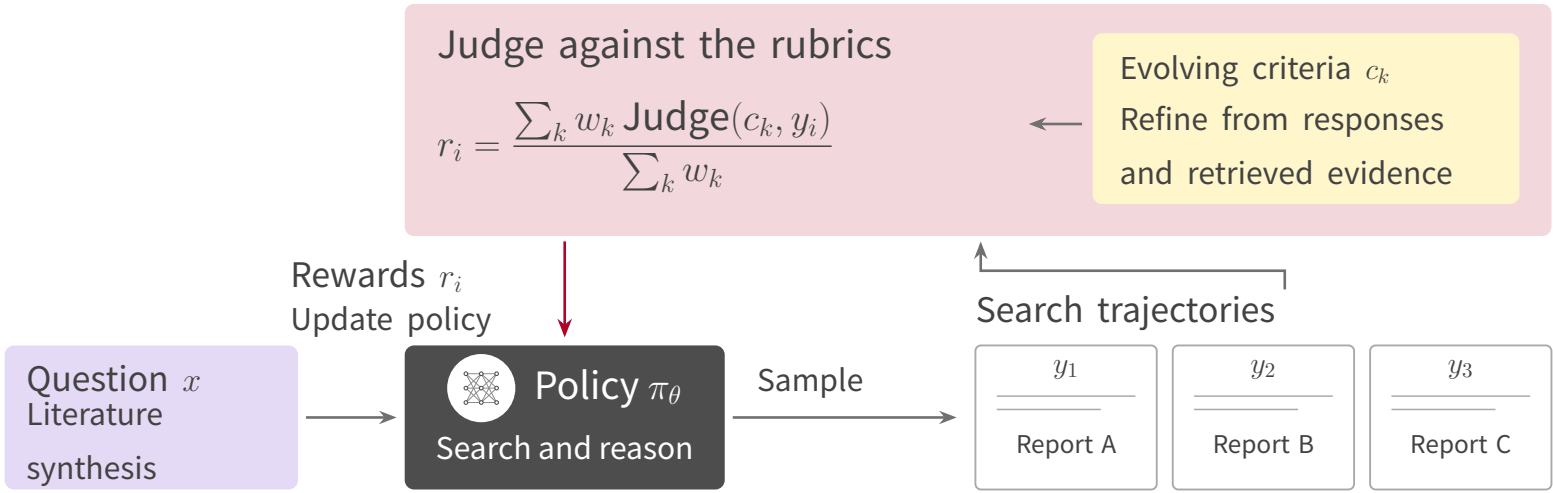
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

DR Tulu: evolving rubrics for RL



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.. Rubric reward shown; auxiliary rewards omitted.

41

03 / 智能体建模

DR Tulu：用于强化学习的演进式评分标准。

使用评分标准进行评判： $r_i = [\sum_k w_k \text{Judge}(c_k, y_i)] / [\sum_k w_k]$ 。

不断演进的标准 c_k ：根据回答和检索到的证据进行优化。

奖励 $r_i \rightarrow$ 更新策略。

问题 x ：文献综合 $\rightarrow$ 策略 π_θ （搜索并推理） $\rightarrow$ 采样轨迹 y_1 、 y_2 、 $y_3 \rightarrow$ 报告 A、B、C。

图中显示评分标准奖励；未显示辅助奖励。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

公式与变量保留原样。

DR Tulu：用于强化学习的演进式评分标准

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

How the rubrics evolve during training

User query

Describe how genetically engineering T cells could be used as an anti-inflammatory therapy for IBD by changing cytokine signaling between T cells and macrophages.

Generate rubrics with search

Persistent rubrics

Cites "IL-10-engineered T cells **reduced colitis severity**"

States "engineering T cells to overexpress IL-10 **suppresses macrophage TNF- α** "

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

42

03 / 智能体建模

评分标准在训练中如何演进。

用户问题：描述如何通过改变 T 细胞与巨噬细胞之间的细胞因子信号，将基因工程 T 细胞用于 IBD 抗炎治疗。

生成带搜索的评分标准。持续有效的评分标准：a. 引用 “IL-10 工程化 T 细胞降低结肠炎严重程度” ； b. 指出 “工程化 T 细胞过表达 IL-10 会抑制巨噬细胞 TNF- α ” 。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

评分标准在训练中如何演进

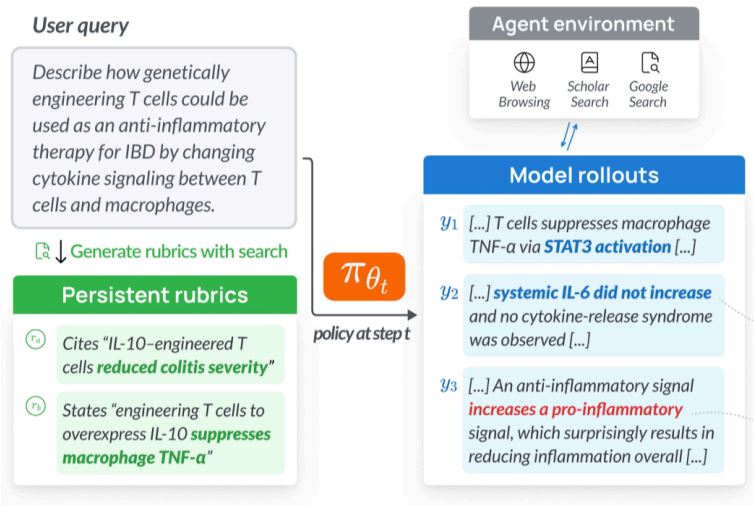
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

How the rubrics evolve during training



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

03 / 智能体建模

评分标准在训练中如何演进。

用户问题：描述如何通过改变 T 细胞与巨噬细胞之间的细胞因子信号，将基因工程 T 细胞用于 IBD 抗炎治疗。

持续有效的评分标准：引用“IL-10 工程化 T 细胞降低结肠炎严重程度”；指出“工程化 T 细胞过表达 IL-10 会抑制巨噬细胞 TNF-α”。

智能体环境：Web 浏览、学术搜索、Google 搜索。

策略 π_{θ} 在每个训练步骤生成采样轨迹： y_1 认为 T 细胞通过 STAT3 activation 抑制巨噬细胞 TNF- α ； y_2 认为 systemic IL-6 未增加且未观察到 cytokine-release syndrome； y_3 声称抗炎信号增强促炎信号，却意外降低整体炎症。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

生物医学机制用语及缩写保留原样。

评分标准在训练中如何演进

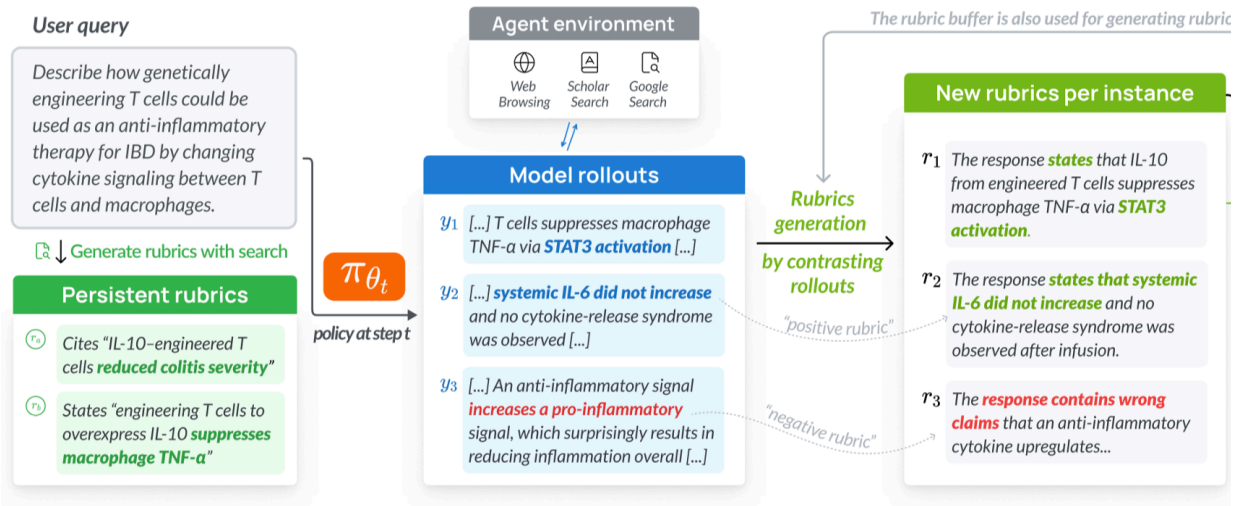
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

How the rubrics evolve during training



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

03 / 智能体建模

评分标准在训练中如何演进。

系统根据不同采样轨迹生成新评分标准： r_1 指出 IL-10 从工程化 T 细胞释放后，经 STAT3 activation 抑制巨噬细胞 TNF- α ； r_2 指出输注后 systemic IL-6 未增加，且未观察到 cytokine-release syndrome； r_3 指出回答包含错误主张，声称抗炎细胞因子会使促炎细胞因子上调。

通过对比采样轨迹生成评分标准；从正向轨迹生成“正向评分标准”，从错误轨迹生成“负向评分标准”。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

小字生物医学示例按可辨内容转写。

评分标准在训练中如何演进

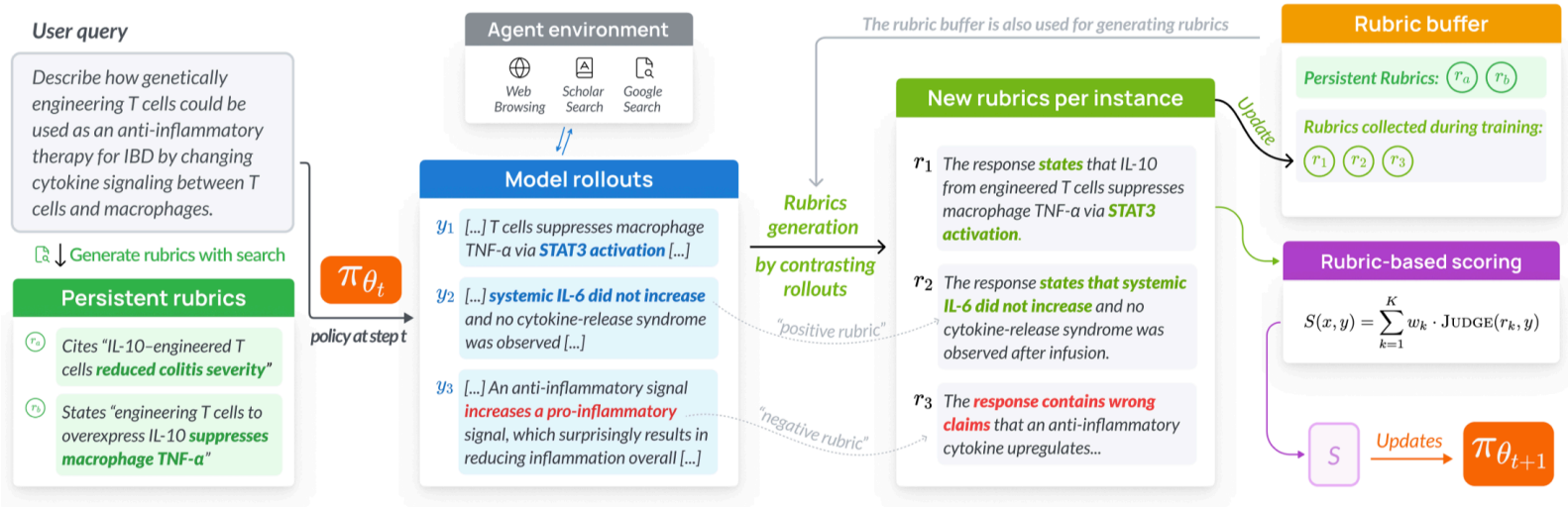
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

How the rubrics evolve during training



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

03 / 智能体建模

评分标准在训练中如何演进。

用户问题、持续评分标准、智能体环境与采样轨迹用于产生训练期间收集的新评分标准。

评分标准缓冲区：持久评分标准 r_a 、 r_b ；训练期间收集的评分标准 r_1 、 r_2 、 r_3 。

基于评分标准的打分： $S(x,y)=\sum_{k=1}^K w_k \cdot \text{Judge}(r_k,y)$ 。

分数 S 更新策略 $\pi_{\theta_{t+1}}$ 。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

公式与变量保留原样。

评分标准在训练中如何演进

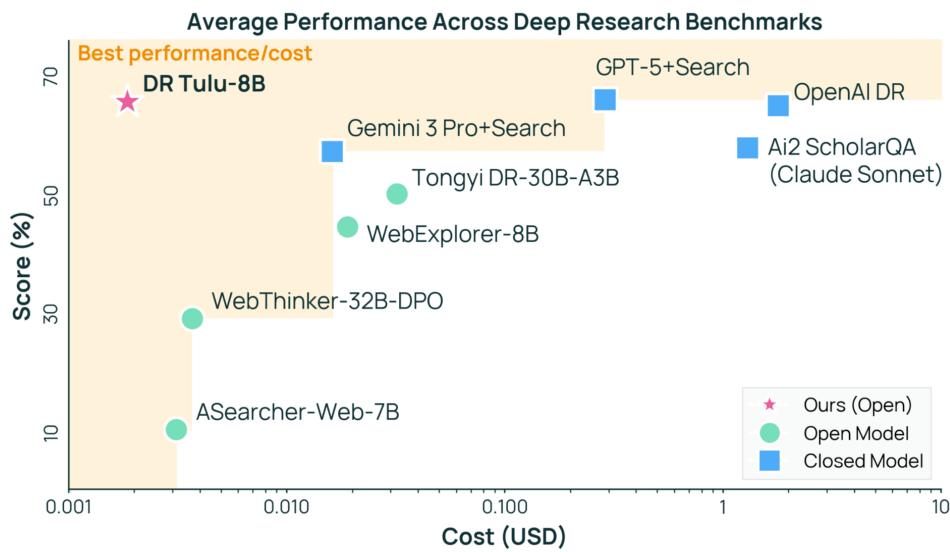
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

DR Tulu: cost and research performance



DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

43

03 / 智能体建模

DR Tulu：成本与研究表现。

深度研究基准的平均表现；纵轴为得分（%），横轴为成本（美元，0.001、0.010、0.100、1、10，对数刻度）。图例：Ours (Open)、Open Model、Closed Model。

图中模型：DR Tulu-8B（表现 / 成本最佳）、WebThinker-32B-DPO、ASearcher-Web-7B、WebExplorer-8B、Tongyi DR-30B-A3B、Gemini 3 Pro+Search、GPT-5+Search、OpenAI DR、Ai2 ScholarQA (Claude Sonnet)。

Shao*、Asai* 等, ICML 2026 (Oral) 。DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

散点未标注具体坐标数值；未根据图中位置估算各模型分数或成本。

DR Tulu：成本与研究表现

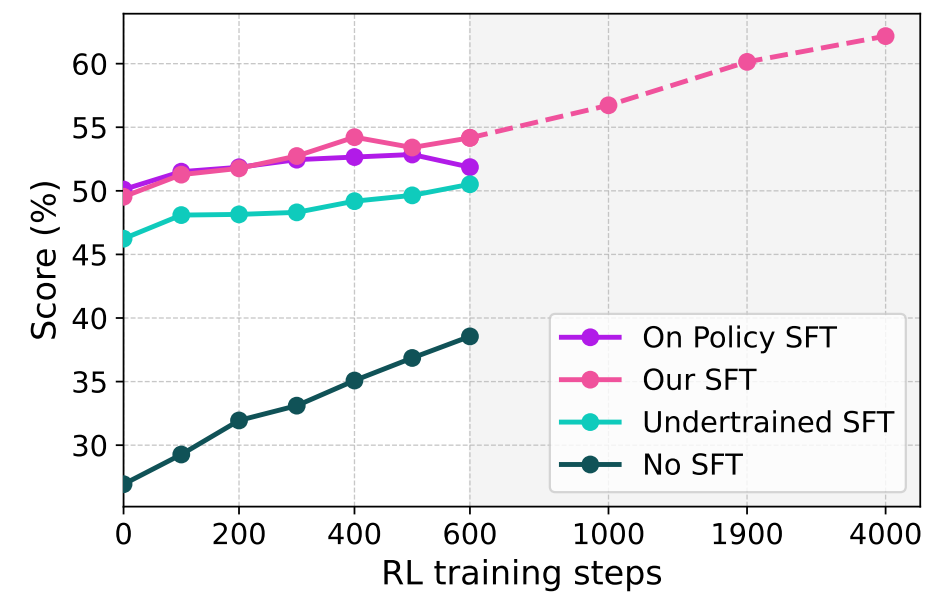
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

03 / AGENT MODELING

DR Tulu: SFT as an RL warm start



Even 5% SFT helps.
The full SFT mixture gives the strongest RL results.

Gray region: unequal step spacing.

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

44

03 / 智能体建模

DR Tulu: 监督微调作为强化学习热启动。

得分 (%) 随 RL training steps 增加而变化；图例：On Policy SFT、Our SFT、Undertrained SFT、No SFT。刻度展示训练步数0–4,000和得分约30–60%以上。

即使只使用5%的 SFT 数据也有帮助。完整 SFT 混合数据带来最强的 RL 结果。

灰色区域表示训练步数间隔不等。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

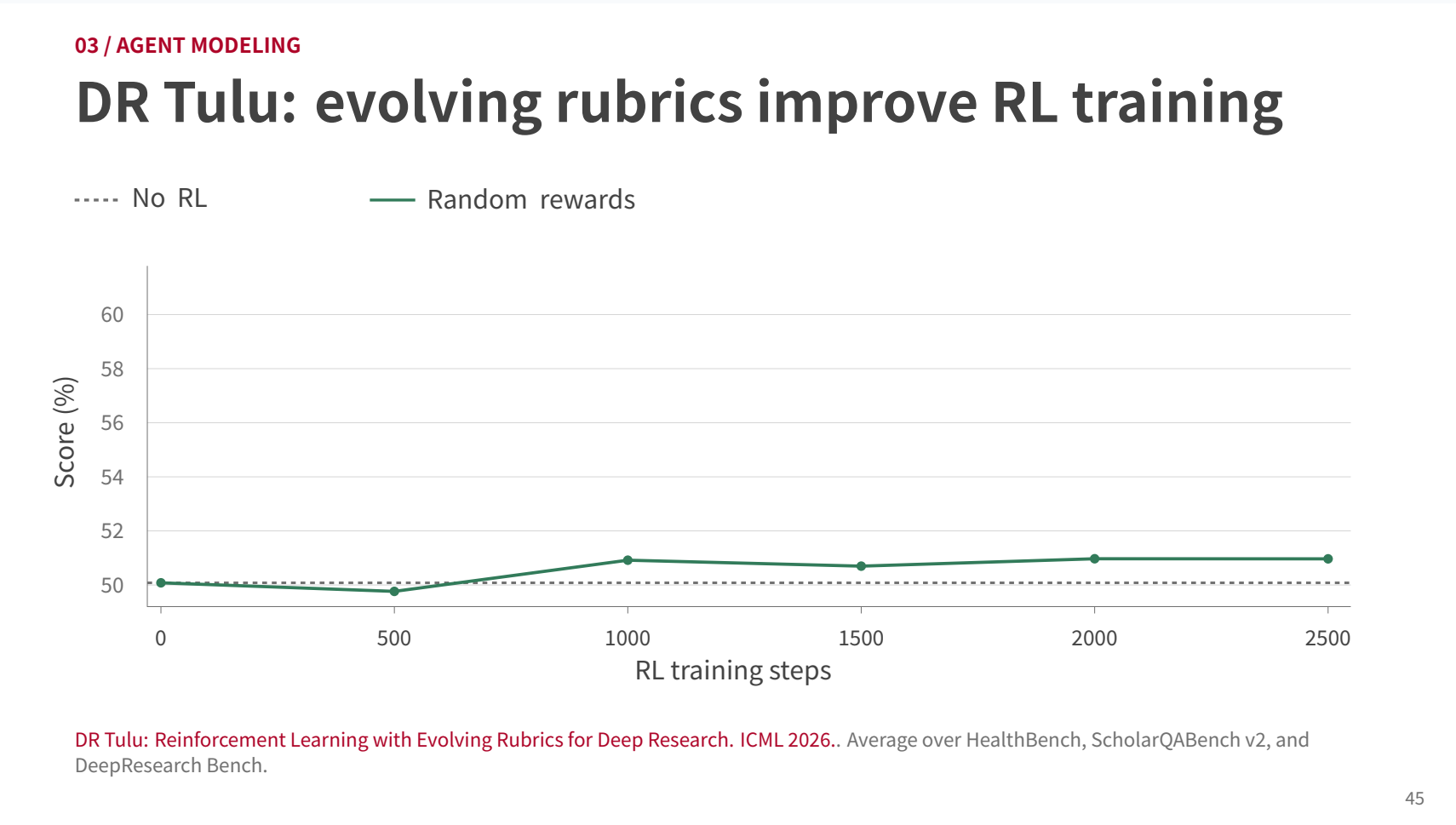
曲线各点未标数值，未估读未标注的数据点。

DR Tulu: 监督微调作为强化学习热启动

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



DR Tulu：演进式评分标准改善强化学习训练

03 / 智能体建模

DR Tulu：演进式评分标准改善强化学习训练。

图例：No RL、Random rewards。纵轴为得分（%），横轴为 RL training steps（0–2,500）；基准线约为50%，随机奖励曲线在约50%附近。

得分取 HealthBench、ScholarQABench v2 与 DeepResearch Bench 的平均值。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

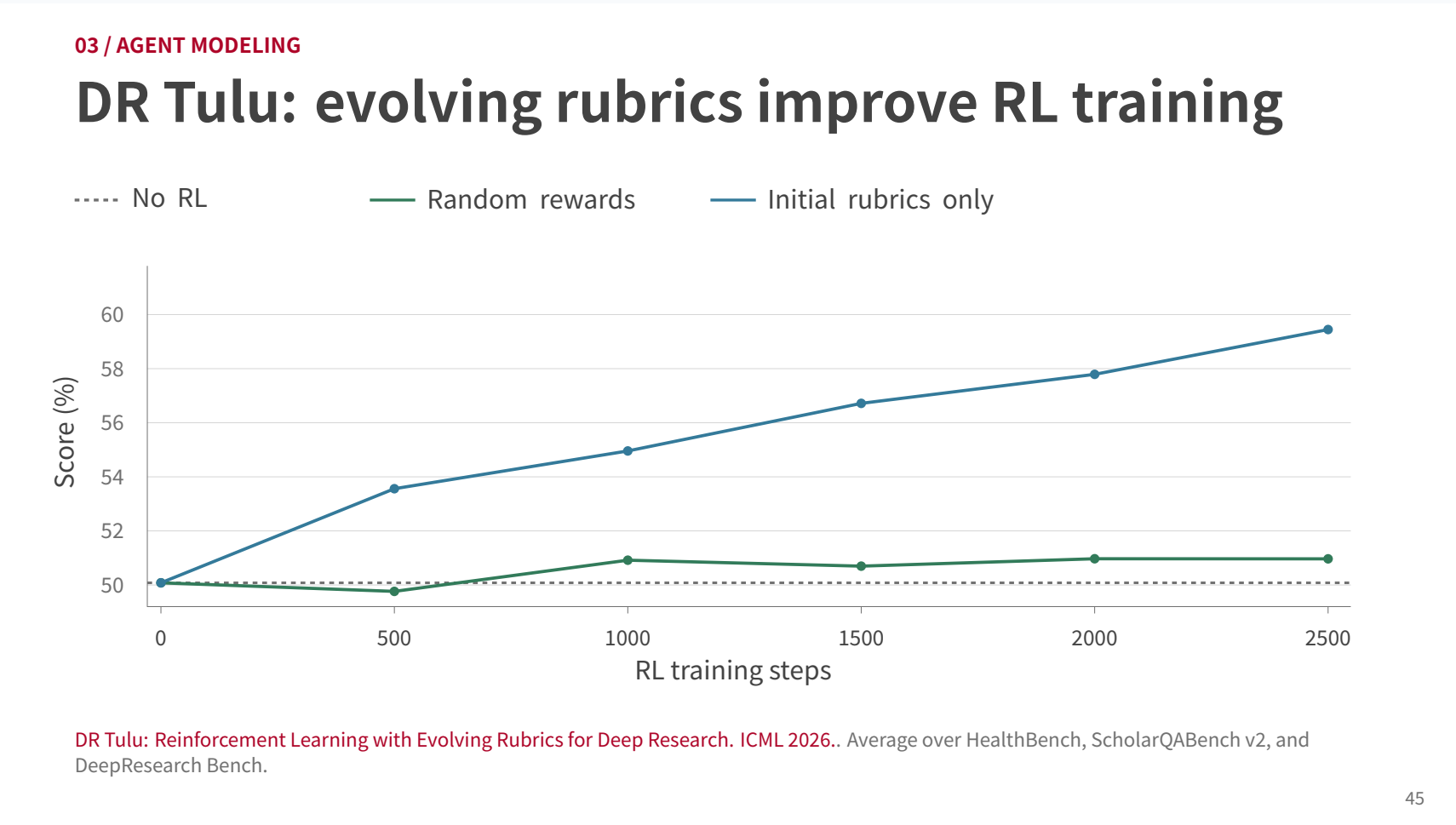
译注 / 与原文翻译分开

曲线数值未逐点标注；不对未标数值作精确估计。

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



03 / 智能体建模

DR Tulu：演进式评分标准改善强化学习训练。

图例：No RL、Random rewards、Initial rubrics only。纵轴得分（%）；横轴 RL training steps（0–2,500）。仅使用初始评分标准的曲线随训练提升至约59%。

得分取 HealthBench、ScholarQABench v2 与 DeepResearch Bench 的平均值。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

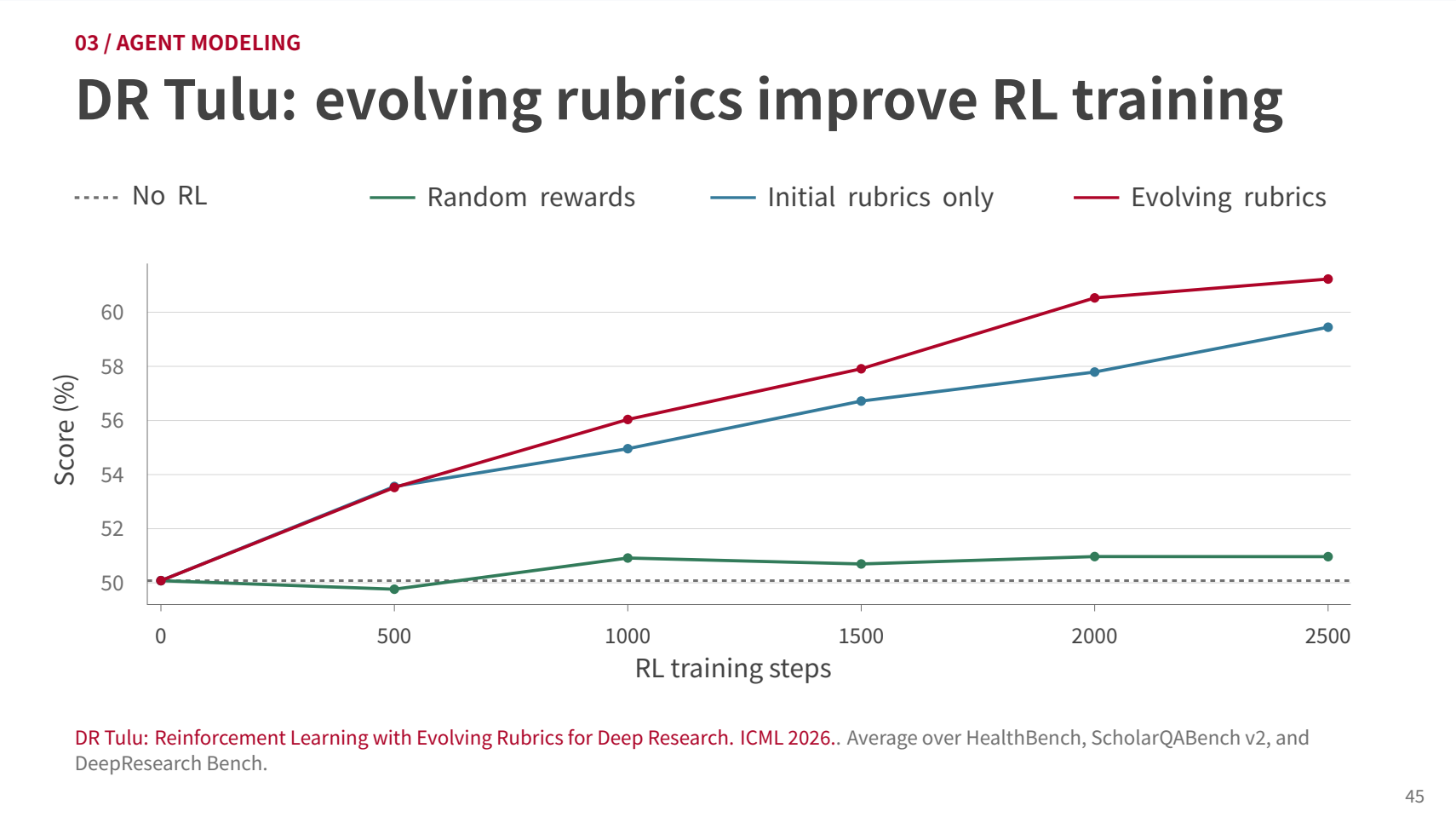
曲线数值未逐点标注；不对未标数值作精确估计。

DR Tulu：演进式评分标准改善强化学习训练

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



03 / 智能体建模

DR Tulu：演进式评分标准改善强化学习训练。

图例：No RL、Random rewards、Initial rubrics only、Evolving rubrics。纵轴得分（%），刻度50–60；横轴 RL training steps（0–2,500）。演进式评分标准曲线高于仅用初始评分标准的曲线。

得分取 HealthBench、ScholarQABench v2 与 DeepResearch Bench 的平均值。

DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

曲线数据点未标注具体数值，未作推断。

DR Tulu：演进式评分标准改善强化学习训练

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04

Retrieval for Deep Research

04

深度研究中的检索

深度研究中的检索

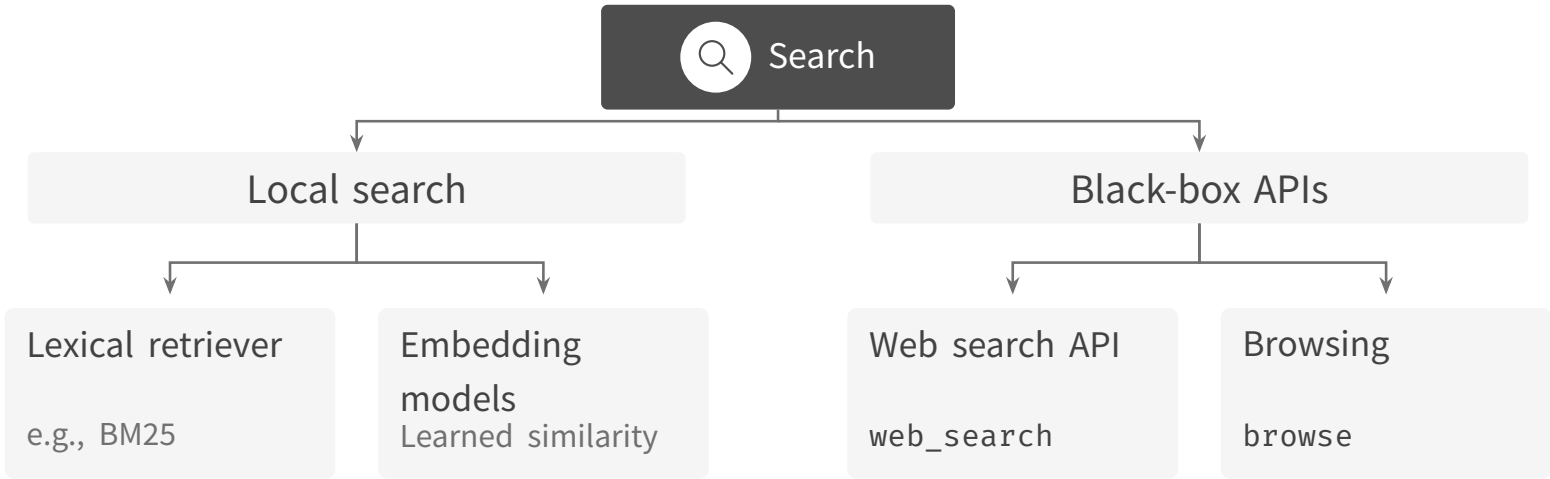
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

Search tools for deep research



Other tools: Python execution (Tongyi); scholarly search (Tongyi, DR Tulu).

Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020.; Tongyi DeepResearch Technical Report. arXiv, 2025.; DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

47

04 / 检索

深度研究的搜索工具。

本地搜索：词法检索器（例如 BM25）； 嵌入模型（学习得到的相似度）。

黑盒 API：网络搜索 API（web_search）； 浏览（browse）。

其他工具：Python 执行（Tongyi）； 学术搜索（Tongyi、DR Tulu）。

Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020. · Tongyi DeepResearch Technical Report. arXiv, 2025. · DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research. ICML 2026.

译注 / 与原文翻译分开

工具与 API 标识保留原样。

深度研究的搜索工具

第 10 讲 · 深度研究智能体：任务、基准与方法

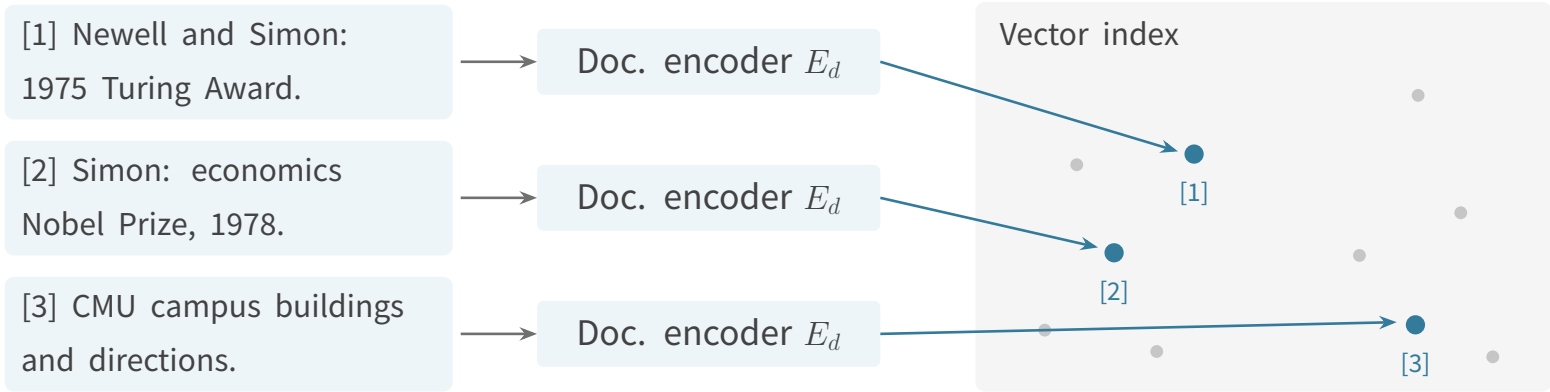
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

Embedding retrieval finds nearby documents

First, encode the document collection and build a vector index.



Score each document: $s(q, d) = E_q(q)^\top E_d(d)$; retrieve the top k .

Dual-encoder retrieval: Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020..

48

04 / 检索

嵌入检索会找到相近的文档。

首先，为文档集合编码并建立向量索引。

文档编码器 E_d ；向量索引。

文档 [1]：Newell 和 Simon：1975 年图灵奖。

文档 [2]：Simon：1978 年诺贝尔经济学奖。

文档 [3]：CMU 校园建筑与路线指引。

为每份文档计算分数： $s(q, d) = E_q(q)^\top E_d(d)$ ；检索分数最高的前 k 份文档。

双编码器检索：《Dense Passage Retrieval for Open-Domain Question Answering》，EMNLP 2020。

译注 / 与原文翻译分开

公式中的编码器与变量记号按原样保留。

嵌入检索会找到相近的文档

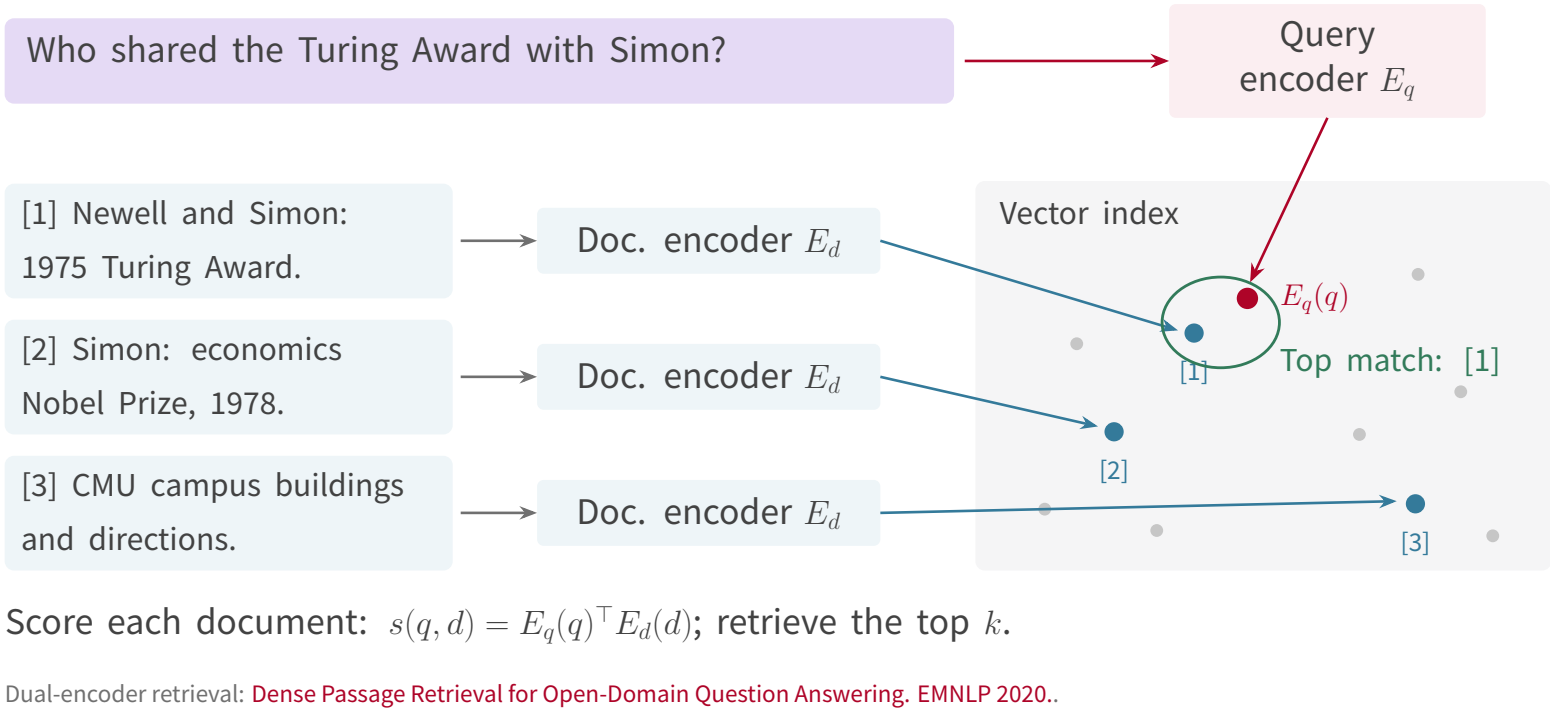
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

Embedding retrieval finds nearby documents



04 / 检索

嵌入检索会找到相近的文档。

问题：谁与 Simon 共同获得了图灵奖？

查询编码器 E_q ；查询表示 $E_q(q)$ 。

文档编码器 E_d ；向量索引。

文档 [1]：Newell 和 Simon：1975 年图灵奖。

文档 [2]：Simon：1978 年诺贝尔经济学奖。

文档 [3]：CMU 校园建筑与路线指引。

最匹配项：[1]。

为每份文档计算分数： $s(q, d) = E_q(q)^\top E_d(d)$ ；检索分数最高的前 k 份文档。

双编码器检索：《Dense Passage Retrieval for Open-Domain Question Answering》，EMNLP 2020。

译注 / 与原文翻译分开

编码器与文档编号保留原样。

嵌入检索会找到相近的文档

第 10 讲 · 深度研究智能体：任务、基准与方法

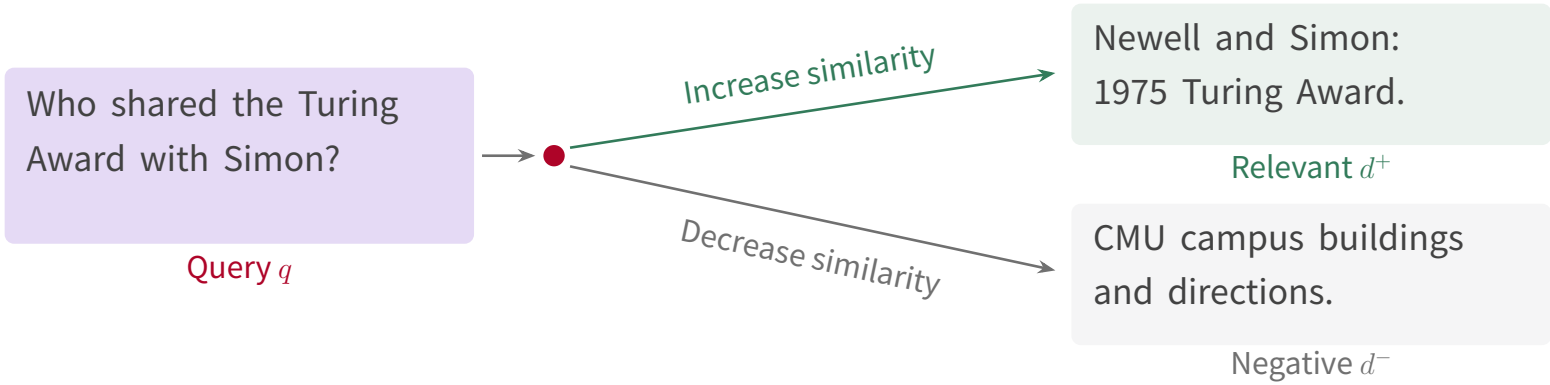
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

Train embeddings to rank useful evidence higher

Contrastive learning compares a relevant passage with negative passages.



For encoder design, negative sampling, and indexing, see Advanced NLP and IR courses.

Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020..

49

04 / 检索

训练嵌入，使有用证据排名更高。

对比学习将相关段落与负例段落进行比较。

问题：谁与 Simon 共同获得了图灵奖？查询 q 。

提高与相关文档 d^+ 的相似度：Newell 和 Simon：1975 年图灵奖。

降低与负例文档 d^- 的相似度：CMU 校园建筑与路线指引。

有关编码器设计、负例采样和索引，参见高级 NLP 与信息检索课程。

《Dense Passage Retrieval for Open-Domain Question Answering》，EMNLP 2020。

译注 / 与原文翻译分开

q 、 d^+ 、 d^- 等数学记号按原样保留。

训练嵌入，使有用证据排名更高

第 10 讲 · 深度研究智能体：任务、基准与方法

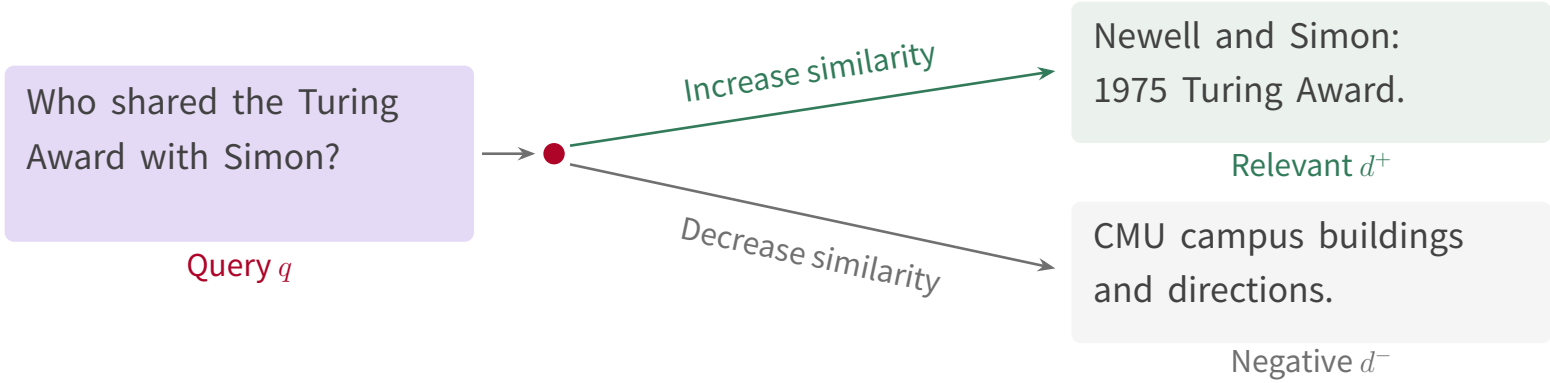
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

Train embeddings to rank useful evidence higher

Contrastive learning compares a relevant passage with negative passages.



$$\mathcal{L} = -\log \frac{\exp s(q, d^+)}{\exp s(q, d^+) + \sum_{d^-} \exp s(q, d^-)}$$

For encoder design, negative sampling, and indexing, see Advanced NLP and IR courses.

Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020..

49

04 / 检索

训练嵌入，使有用证据排名更高。

对比学习将相关段落与负例段落进行比较。

问题：谁与 Simon 共同获得了图灵奖？查询 q。

提高与相关文档 d⁺ 的相似度：Newell 和 Simon：1975 年图灵奖。

降低与负例文档 d⁻ 的相似度：CMU 校园建筑与路线指引。

损失函数：L = -log[exp s(q, d⁺) / (exp s(q, d⁺) + ∑{d⁻} exp s(q, d⁻))].

有关编码器设计、负例采样和索引，参见高级 NLP 与信息检索课程。

《Dense Passage Retrieval for Open-Domain Question Answering》, EMNLP 2020。

译注 / 与原文翻译分开

损失函数沿用幻灯片中的记号；分母对负例文档 d⁻ 求和。

训练嵌入，使有用证据排名更高

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

The retriever often sees only the query

Reasoning is not passed to the retriever.

Only the latest query is passed.

Ranked docs

AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

50

04 / 检索

检索器通常只接收查询。

智能体：原始任务、先前推理、先前查询、先前证据、当前推理 τ_t ，以及最新查询 q_t 。

只把最新查询传给检索器。

推理不会传给检索器。

检索器返回排序后的文档。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

τ_t 和 q_t 为原文变量符号。

检索器通常只接收查询

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

AgentIR: retrieval with reasoning and a query

Agent

Original task

Previous reasoning

Previous query

Previous evidence

⋮

Current reasoning τ_t

Latest query q_t

1 / RETRIEVER INPUT

Reasoning τ_t + Query q_t

2 / RETRIEVER TRAINING

Learn which documents help this reasoning step.

AgentIR

Ranked docs

AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

51

04 / 检索

AgentIR：将推理与查询一同用于检索。

智能体包含原始任务、先前推理、先前查询、先前证据、当前推理 τ_t 和最新查询 q_t 。

1 / 检索器输入：推理 τ_t + 查询 q_t 。

2 / 检索器训练：学习哪些文档有助于当前这一步推理。

AgentIR 返回排序后的文档。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》, COLM 2026。

译注 / 与原文翻译分开

τ_t 和 q_t 为原文变量符号。

AgentIR：将推理与查询一同用于检索

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 108 | 原页 106

第 10 讲 · 深度研究智能体：任务、基准与方法

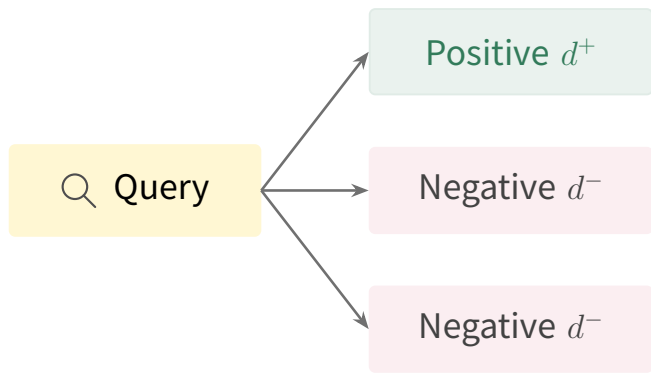
EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

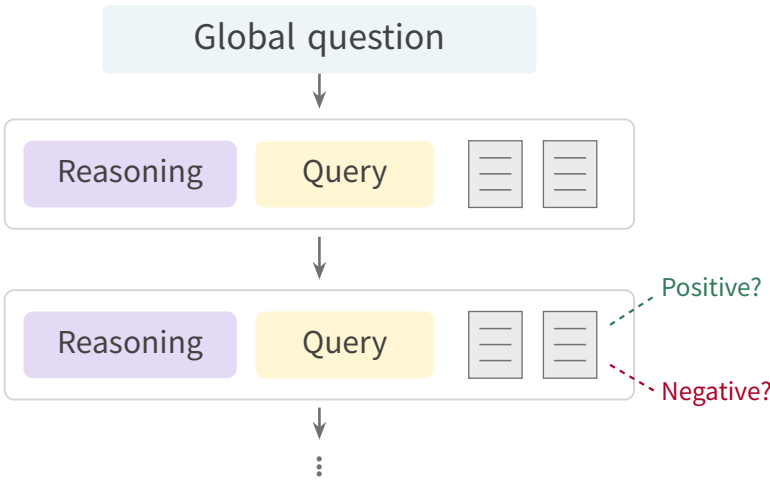
Intermediate searches need local labels

Standard retriever training



Known relevance labels

Deep research trajectory



Which documents help this turn?

AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

52

04 / 检索

中间检索需要局部标签。

标准检索器训练：查询对应一个正例 d^+ 和负例 d^- ；已知相关性标签。

深度研究轨迹：全局问题逐步展开为“推理—查询”回合，并对文档作正例或负例判断。

关键问题：哪些文档有助于这一回合？

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

d^+ 、 d^- 等记号保留原样。

中间检索需要局部标签

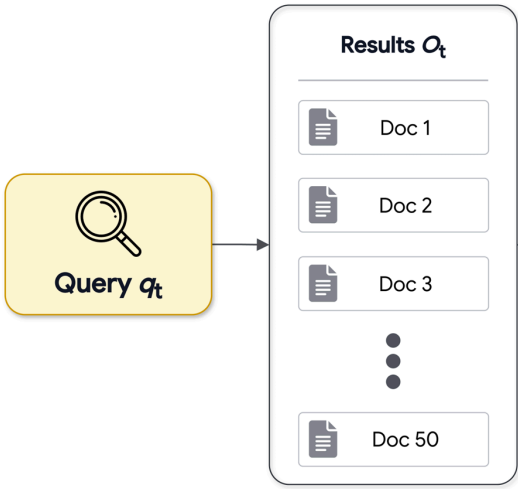
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

DR-Synth: labels for intermediate searches



AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

53

04 / 检索

DR-Synth 为检索器提供由智能体生成的信息需求。

DR-Synth：为中间搜索生成标签。

结果 O_t ：文档 1、文档 2、文档 3、.....、文档 50。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》 , COLM 2026。

译注 / 与原文翻译分开

本页为逐步展开的图示，按当前页可见内容翻译。

DR-Synth 为检索器提供由智能体生成的信息需求

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

DR-Synth: labels for intermediate searches

Query q_t

Results O_t

Doc 1

Doc 2

Doc 3

⋮

Doc 50

Add Global Golds P

Global Gold 1

⋮

Global Gold 5

Doc 1

⋮

Doc 45

AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

53

04 / 检索

DR-Synth 为检索器提供由智能体生成的信息需求。

DR-Synth：为中间搜索生成标签。

结果 O_t ：文档 1、文档 2、文档 3、.....、文档 50。

添加全局金标准 P ：全局金标准 1、.....、全局金标准 5；并列出文档 1 至文档 45。

译注 / 与原文翻译分开

本页在前页图示上增加“全局金标准”文档集合；省略号按图示保留。

DR-Synth 为检索器提供由智能体生成的信息需求

个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 111 | 原页 109

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

DR-Synth: labels for intermediate searches

Query q_t

Results Q_t

Doc 1

Doc 2

Doc 3

...

Doc 50

Add Global Golds P

Global Gold 1

...

Global Gold 5

Doc 1

...

Doc 45

Global Query Q

Listwise Rerank

Global Answer A

AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

53

04 / 检索

DR-Synth 为检索器提供由智能体生成的信息需求。

DR-Synth：为中间搜索生成标签。

结果 O_t ：文档 1、文档 2、文档 3、.....、文档 50；加入全局金标准 P 。

全局查询 Q 输入列表式重排器（Listwise Rerank），并结合全局答案 A 。

译注 / 与原文翻译分开

本页在前页图示上增加全局查询、列表式重排与全局答案模块。

DR-Synth 为检索器提供由智能体生成的信息需求

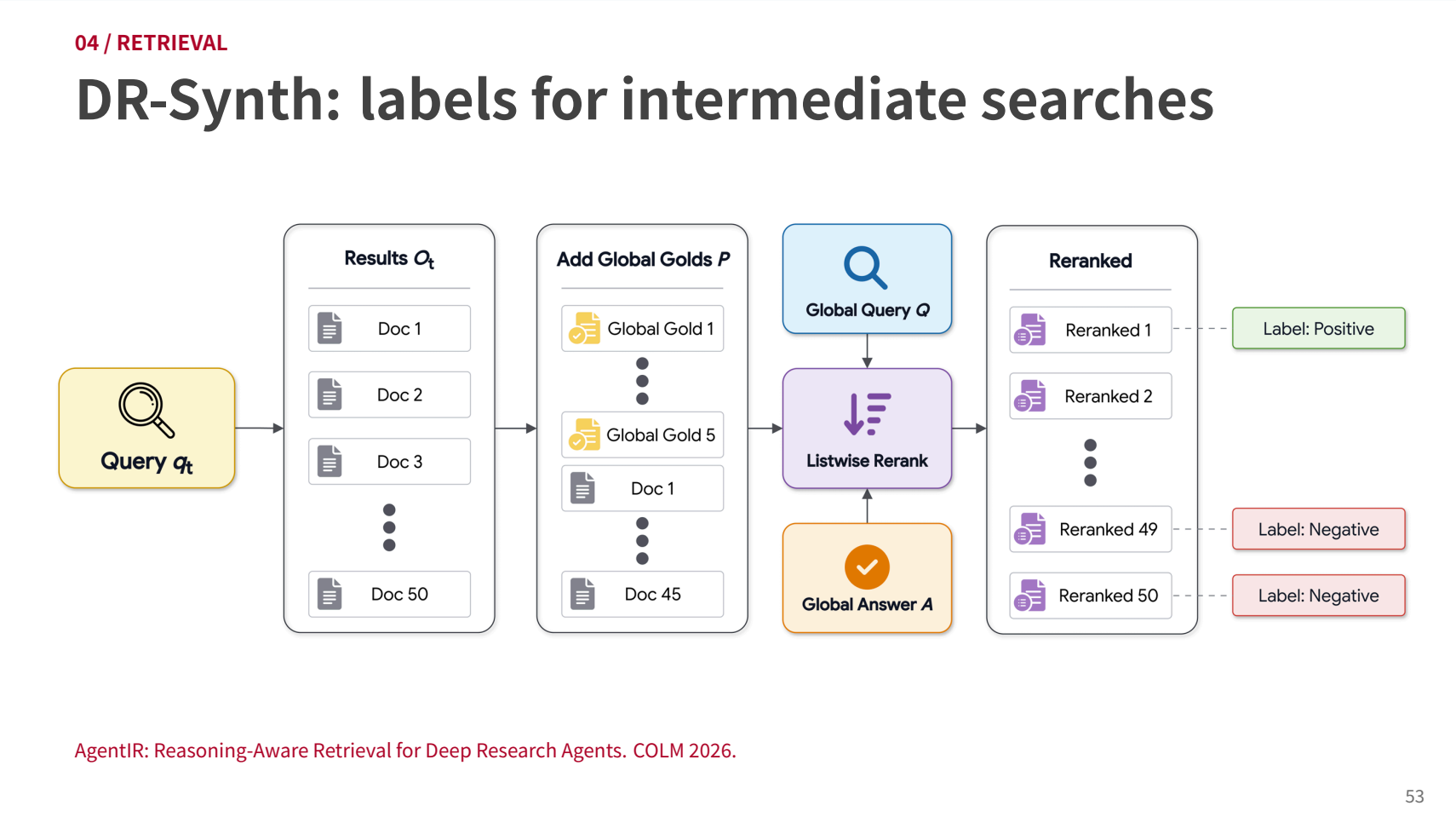
个人学习译文 · 非 CMU 官方译本 · 译注独立标记 · 原始课件来自 cmu-agents.com

双语 PDF 112 | 原页 110

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

DR-Synth 为检索器提供由智能体生成的信息需求。

DR-Synth：为中间搜索生成标签。

结果 O_t ：文档 1、文档 2、文档 3、.....、文档 50；加入全局金标准 P 。

全局查询 Q 与全局答案 A 用于列表式重排；重排后前列示例为重排文档 1、重排文档 2，末列示例为重排文档 49、重排文档 50。

标签：正例（Positive）；标签：负例（Negative）。

译注 / 与原文翻译分开

图示将重排列表靠前文档标为正例，并将靠后的文档 49、50 标为负例。

DR-Synth 为检索器提供由智能体生成的信息需求

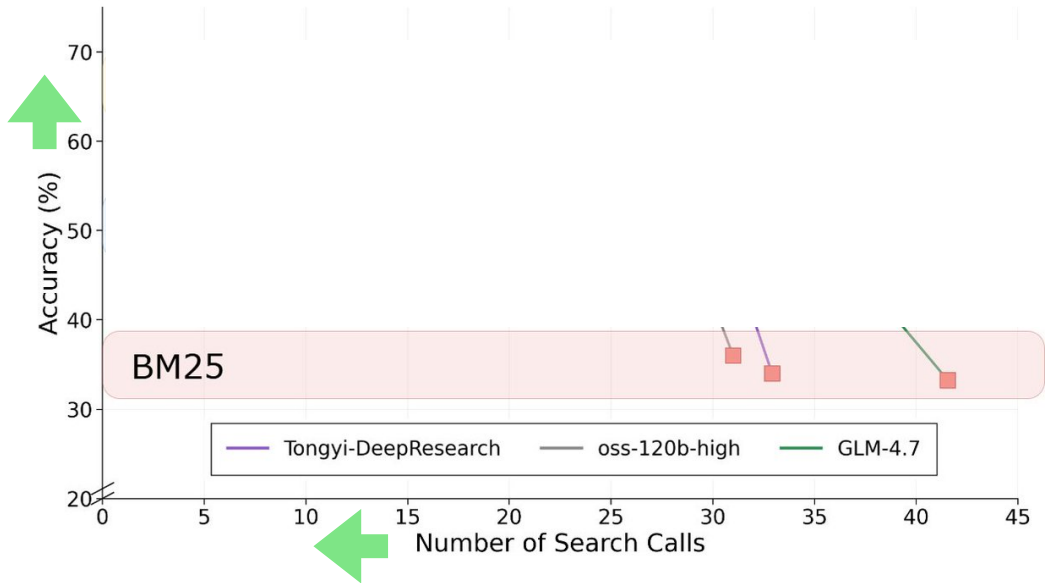
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

AgentIR: higher accuracy with fewer searches



AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

54

04 / 检索

AgentIR：用更少搜索获得更高准确率。

AgentIR：用更少搜索找到更多证据。

纵轴：准确率（%）；横轴：搜索调用次数。

基线检索器：BM25。

系列：Tongyi-DeepResearch、oss-120b-high、GLM-4.7。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

图中散点没有逐点标注数值；未从坐标位置估读具体数值。

AgentIR：用更少搜索获得更高准确率

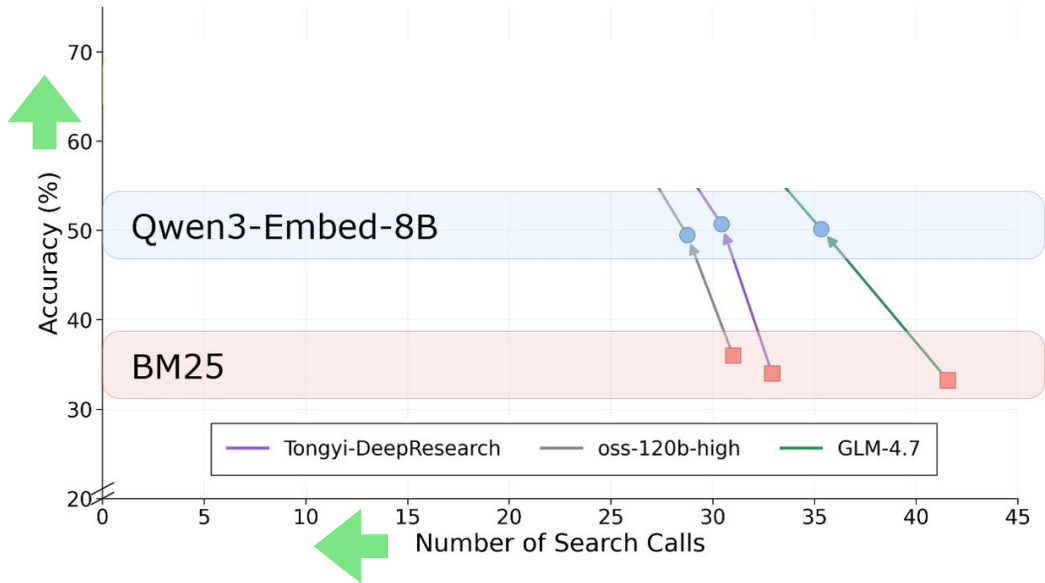
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

AgentIR: higher accuracy with fewer searches



AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

54

04 / 检索

AgentIR：用更少搜索获得更高准确率。

AgentIR：用更少搜索找到更多证据。

纵轴：准确率（%）；横轴：搜索调用次数。

检索器：BM25、Qwen3-Embed-8B。

系列：Tongyi-DeepResearch、oss-120b-high、GLM-4.7。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

图中散点没有逐点标注数值；未从坐标位置估读具体数值。

AgentIR：用更少搜索获得更高准确率

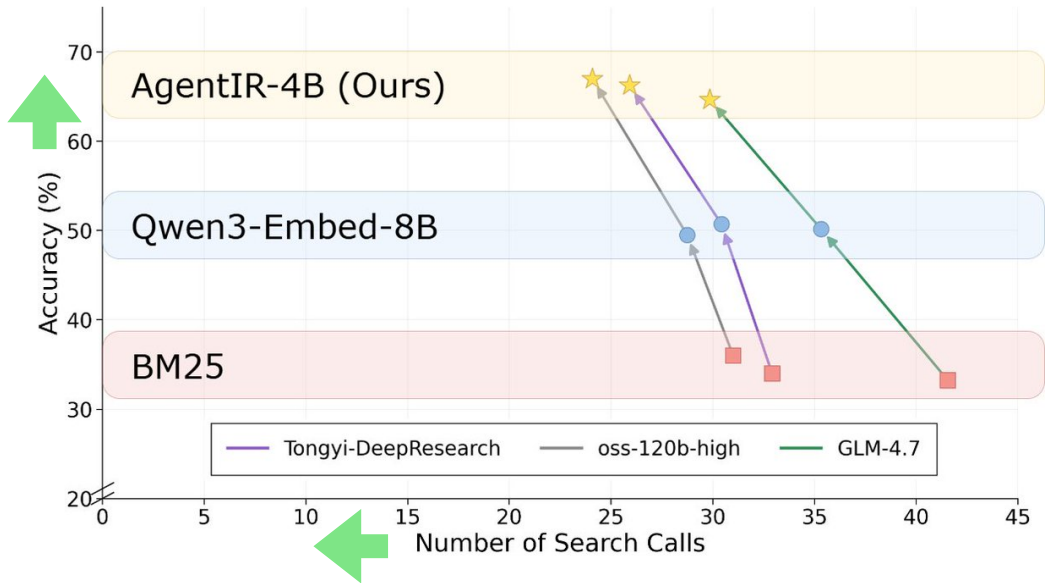
第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

AgentIR: higher accuracy with fewer searches



AgentIR: Reasoning-Aware Retrieval for Deep Research Agents. COLM 2026.

54

04 / 检索

AgentIR：用更少搜索获得更高准确率。

AgentIR：用更少搜索找到更多证据。

纵轴：准确率（%）；横轴：搜索调用次数。

检索器：BM25、Qwen3-Embed-8B、AgentIR-4B（作者提出）。

系列：Tongyi-DeepResearch、oss-120b-high、GLM-4.7。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

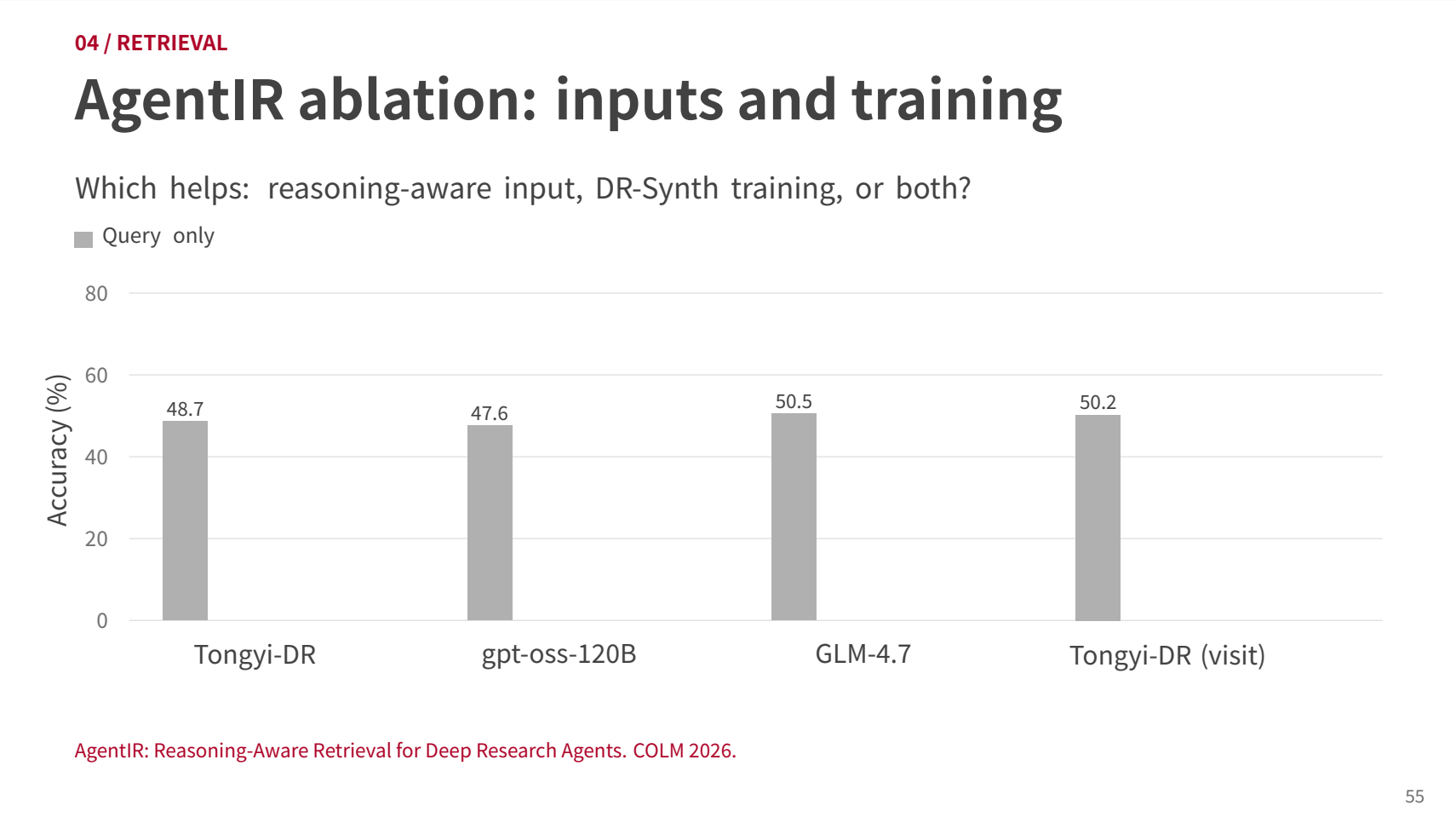
图中散点没有逐点标注数值；未从坐标位置估读具体数值。

AgentIR：用更少搜索获得更高准确率

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

AgentIR 消融实验：输入与训练。

哪一种有帮助：具备推理感知能力的输入、DR-Synth 训练，还是两者兼有？

图例：仅查询（Query only）。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：48.7；gpt-oss-120B：47.6；GLM-4.7：50.5；Tongyi-DR（visit）：50.2。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

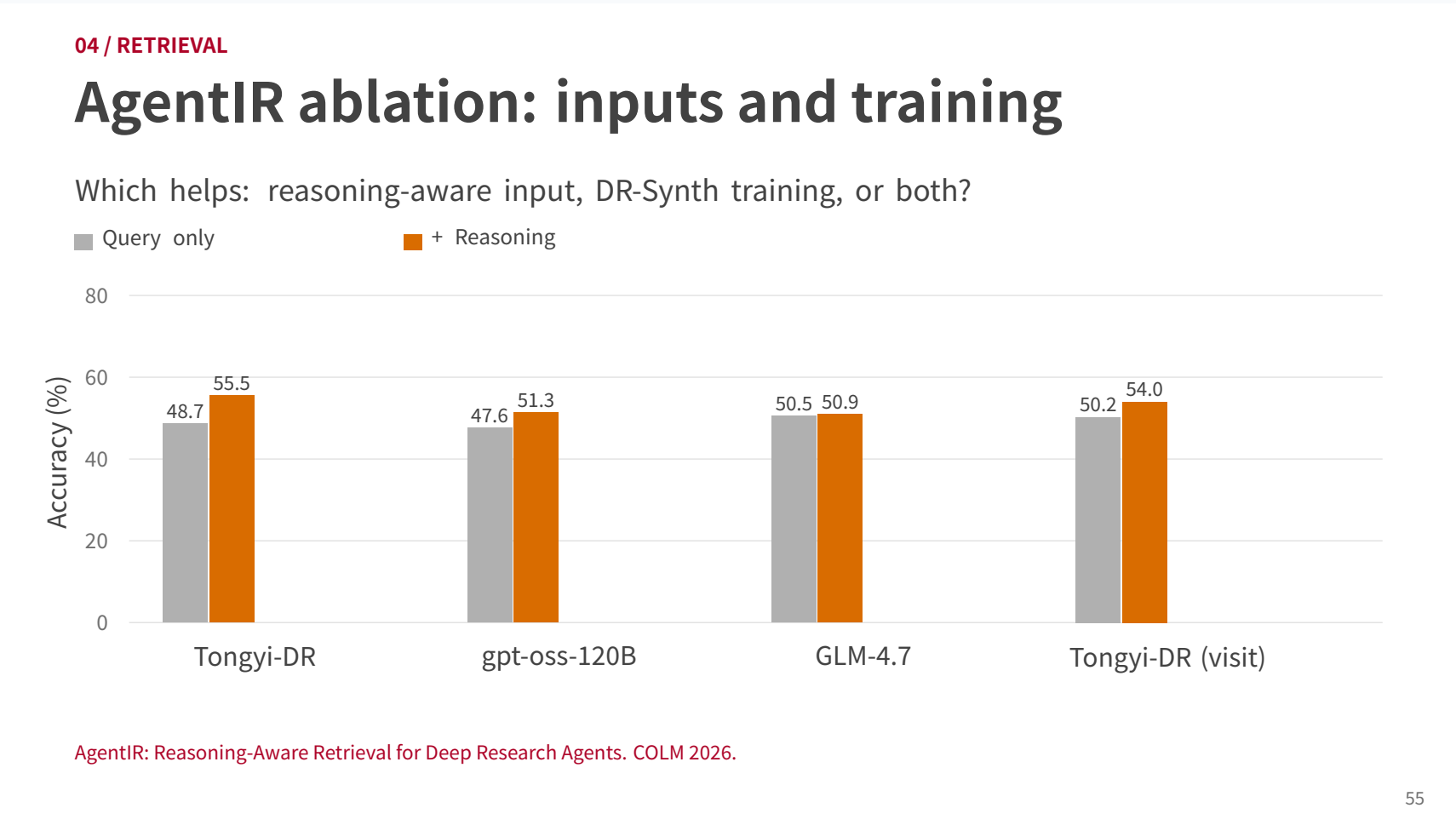
模型名及图例标签保留原样。

AgentIR 消融实验：输入与训练

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

AgentIR 消融实验：输入与训练。

哪一种有帮助：具备推理感知能力的输入、DR-Synth 训练，还是两者兼有？

图例：仅查询；+ 推理。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：仅查询 48.7，+ 推理 55.5。

gpt-oss-120B：仅查询 47.6，+ 推理 51.3。

GLM-4.7：仅查询 50.5，+ 推理 50.9。

Tongyi-DR (visit)：仅查询 50.2，+ 推理 54.0。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

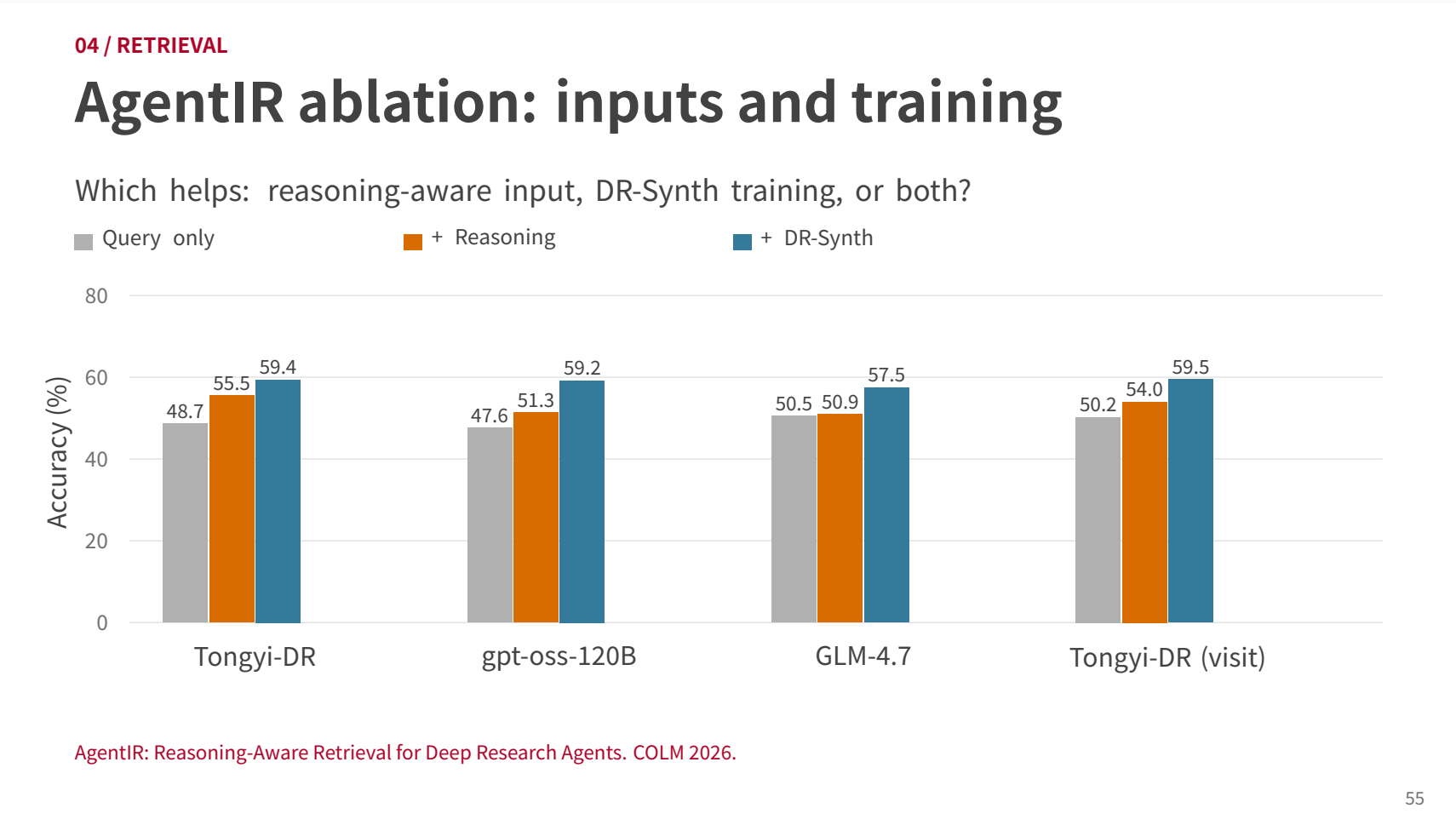
准确率按柱状图标注值抄录。

AgentIR 消融实验：输入与训练

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

AgentIR 消融实验：输入与训练。

哪一种有帮助：具备推理感知能力的输入、DR-Synth 训练，还是两者兼有？

图例：仅查询；+ 推理；+ DR-Synth。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：仅查询 48.7，+ 推理 55.5，+ DR-Synth 59.4。

gpt-oss-120B：仅查询 47.6，+ 推理 51.3，+ DR-Synth 59.2。

GLM-4.7：仅查询 50.5，+ 推理 50.9，+ DR-Synth 57.5。

Tongyi-DR（visit）：仅查询 50.2，+ 推理 54.0，+ DR-Synth 59.5。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

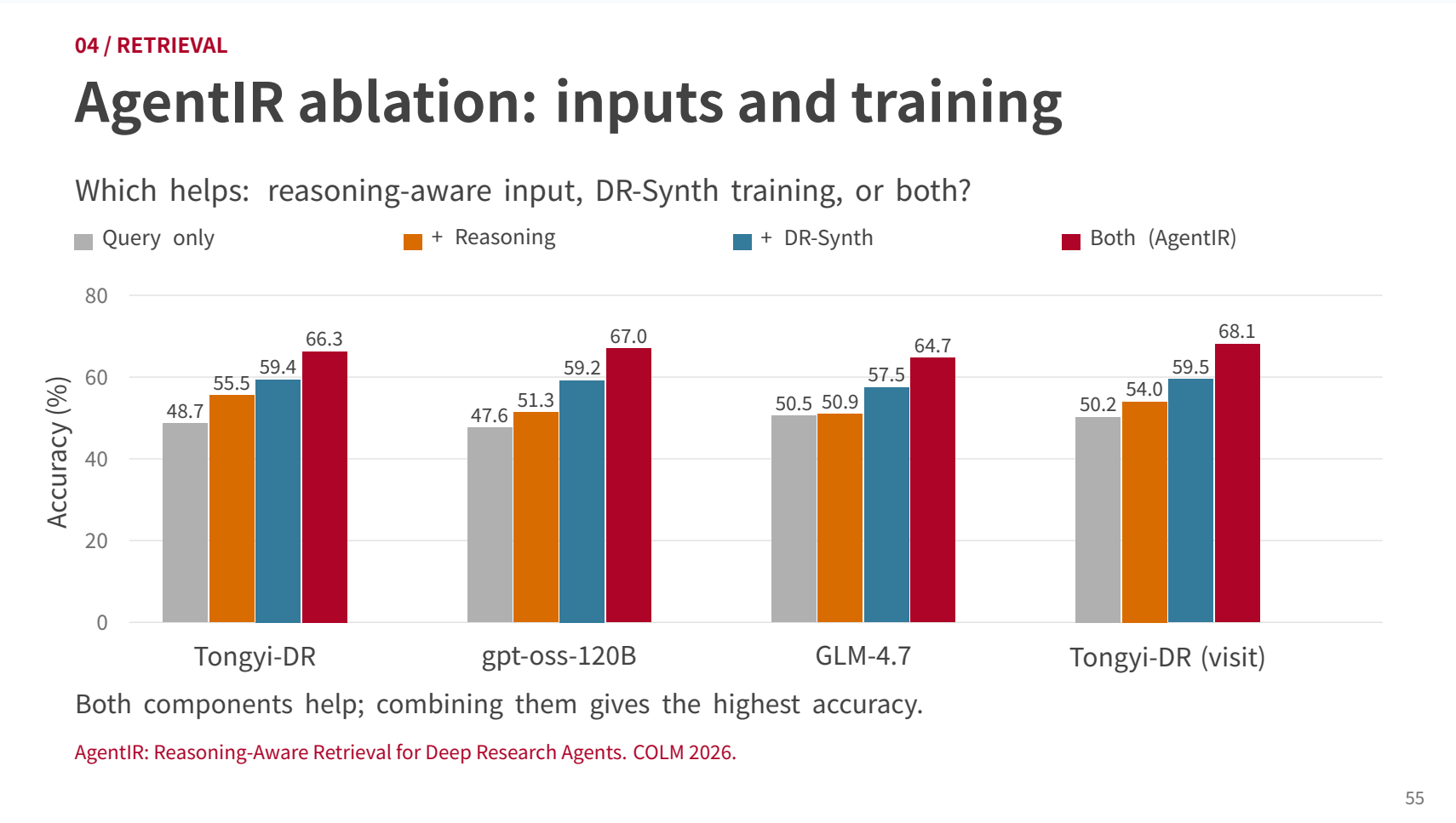
准确率按柱状图标注值抄录。

AgentIR 消融实验：输入与训练

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



AgentIR 消融实验：输入与训练

04 / 检索

AgentIR 消融实验：输入与训练。

哪一种有帮助：具备推理感知能力的输入、DR-Synth 训练，还是两者兼有？

图例：仅查询；+ 推理；+ DR-Synth；两者兼有（AgentIR）。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：仅查询 48.7，+ 推理 55.5，+ DR-Synth 59.4，两者兼有 66.3。

gpt-oss-120B：仅查询 47.6，+ 推理 51.3，+ DR-Synth 59.2，两者兼有 67.0。

GLM-4.7：仅查询 50.5，+ 推理 50.9，+ DR-Synth 57.5，两者兼有 64.7。

Tongyi-DR（visit）：仅查询 50.2，+ 推理 54.0，+ DR-Synth 59.5，两者兼有 68.1。

两个组成部分都有效；将它们结合起来可获得最高准确率。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

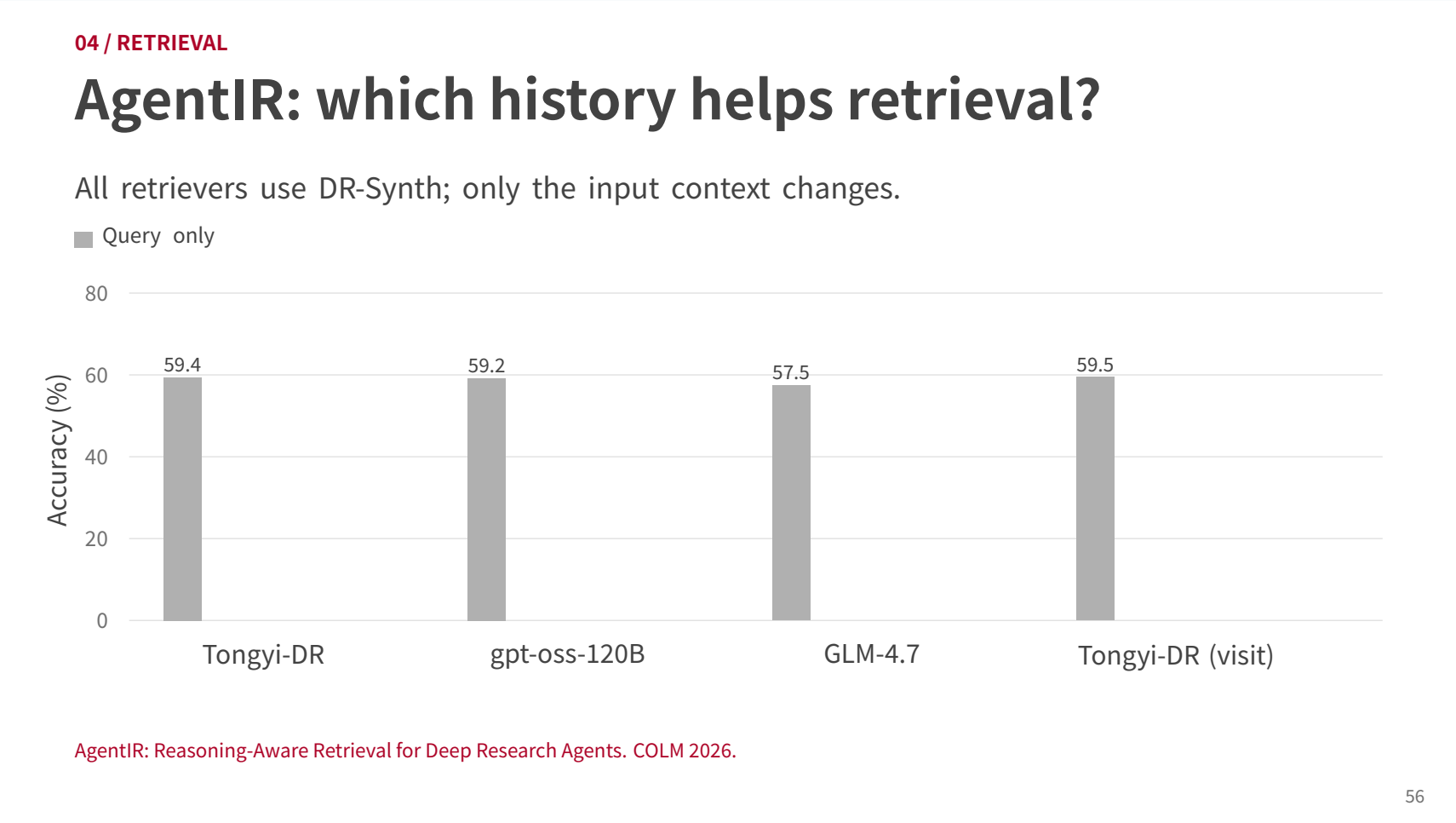
译注 / 与原文翻译分开

准确率按柱状图标注值抄录。

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

AgentIR：哪些历史信息有助于检索？

所有检索器都使用 DR-Synth；只有输入上下文不同。

图例：仅查询（Query only）。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：59.4；gpt-oss-120B：59.2；GLM-4.7：57.5；Tongyi-DR（visit）：59.5。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

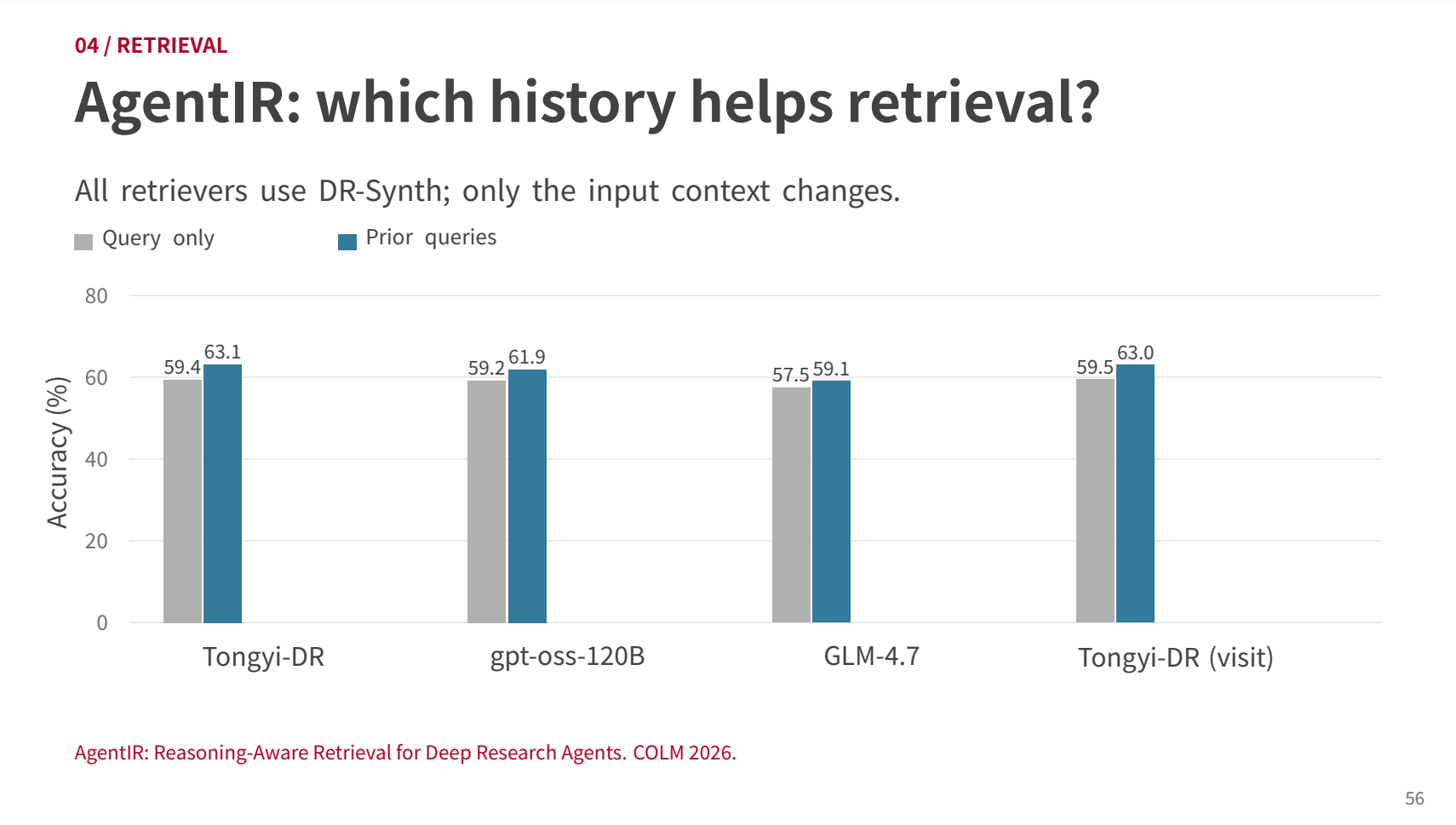
准确率按柱状图标注值抄录。

AgentIR：哪些历史信息有助于检索？

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

AgentIR：哪些历史信息有助于检索？

所有检索器都使用 DR-Synth；只有输入上下文不同。

图例：仅查询；先前查询。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：仅查询 59.4，先前查询 63.1。

gpt-oss-120B：仅查询 59.2，先前查询 61.9。

GLM-4.7：仅查询 57.5，先前查询 59.1。

Tongyi-DR（visit）：仅查询 59.5，先前查询 63.0。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

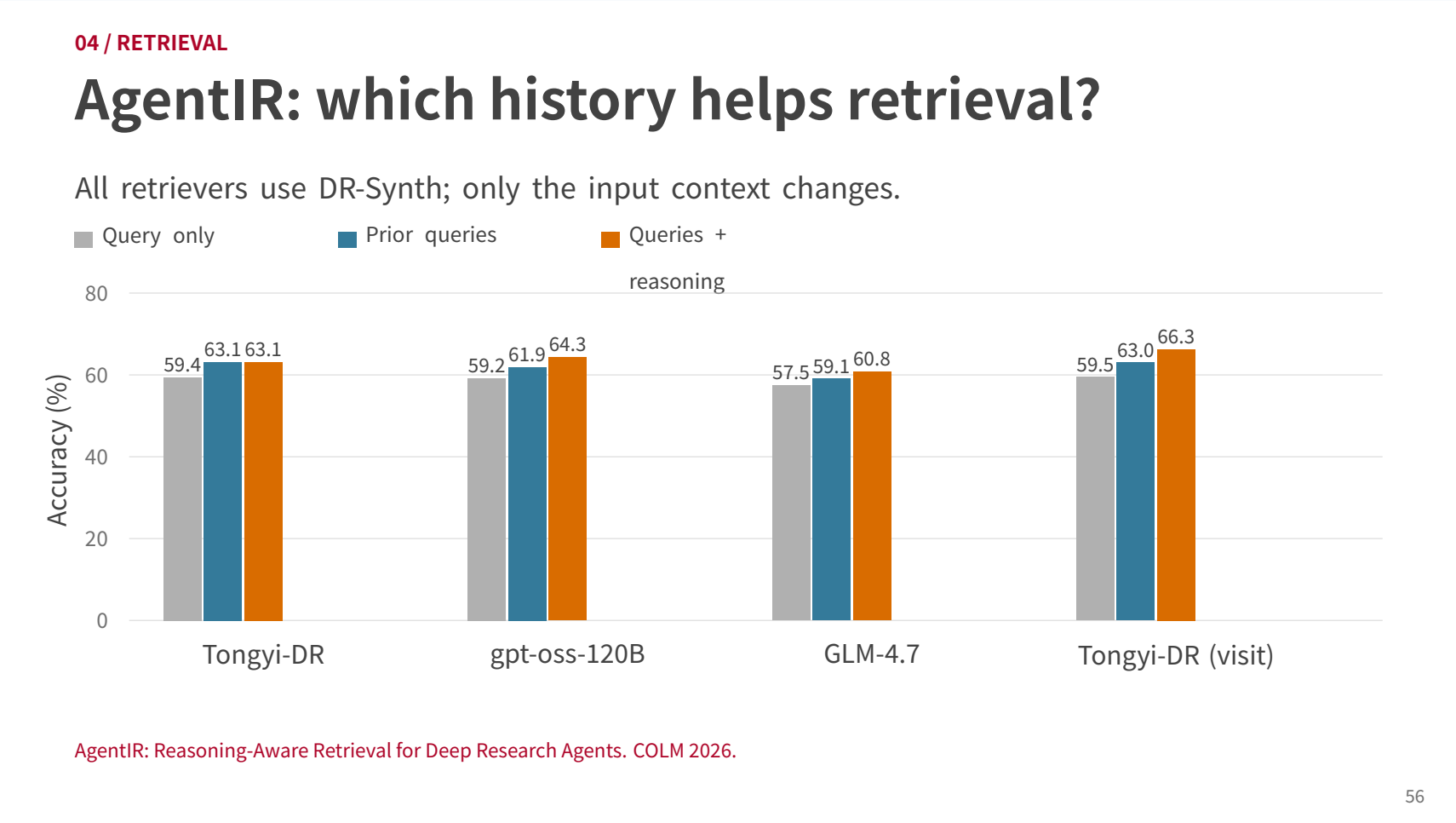
准确率按柱状图标注值抄录。

AgentIR：哪些历史信息有助于检索？

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

AgentIR：哪些历史信息有助于检索？

所有检索器都使用 DR-Synth；只有输入上下文不同。

图例：仅查询；先前查询；查询 + 推理。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：仅查询 59.4，先前查询 63.1，查询 + 推理 63.1。

gpt-oss-120B：仅查询 59.2，先前查询 61.9，查询 + 推理 64.3。

GLM-4.7：仅查询 57.5，先前查询 59.1，查询 + 推理 60.8。

Tongyi-DR（visit）：仅查询 59.5，先前查询 63.0，查询 + 推理 66.3。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

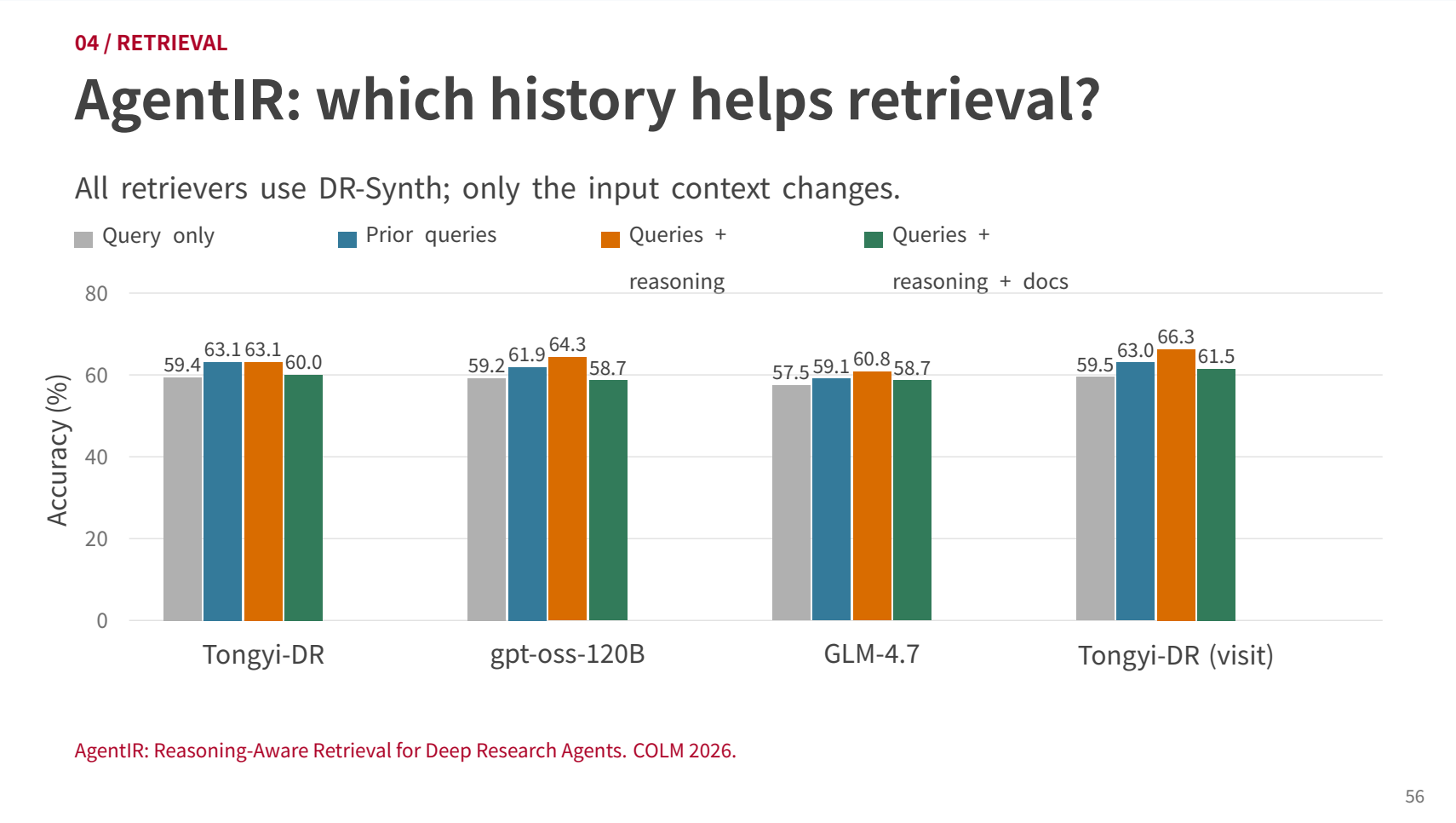
准确率按柱状图标注值抄录。

AgentIR：哪些历史信息有助于检索？

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

AgentIR：哪些历史信息有助于检索？

所有检索器都使用 DR-Synth；只有输入上下文不同。

图例：仅查询；先前查询；查询 + 推理；查询 + 推理 + 文档。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：仅查询 59.4，先前查询 63.1，查询 + 推理 63.1，查询 + 推理 + 文档 60.0。

gpt-oss-120B：仅查询 59.2，先前查询 61.9，查询 + 推理 64.3，查询 + 推理 + 文档 58.7。

GLM-4.7：仅查询 57.5，先前查询 59.1，查询 + 推理 60.8，查询 + 推理 + 文档 58.7。

Tongyi-DR (visit) ：仅查询 59.5，先前查询 63.0，查询 + 推理 66.3，查询 + 推理 + 文档 61.5。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

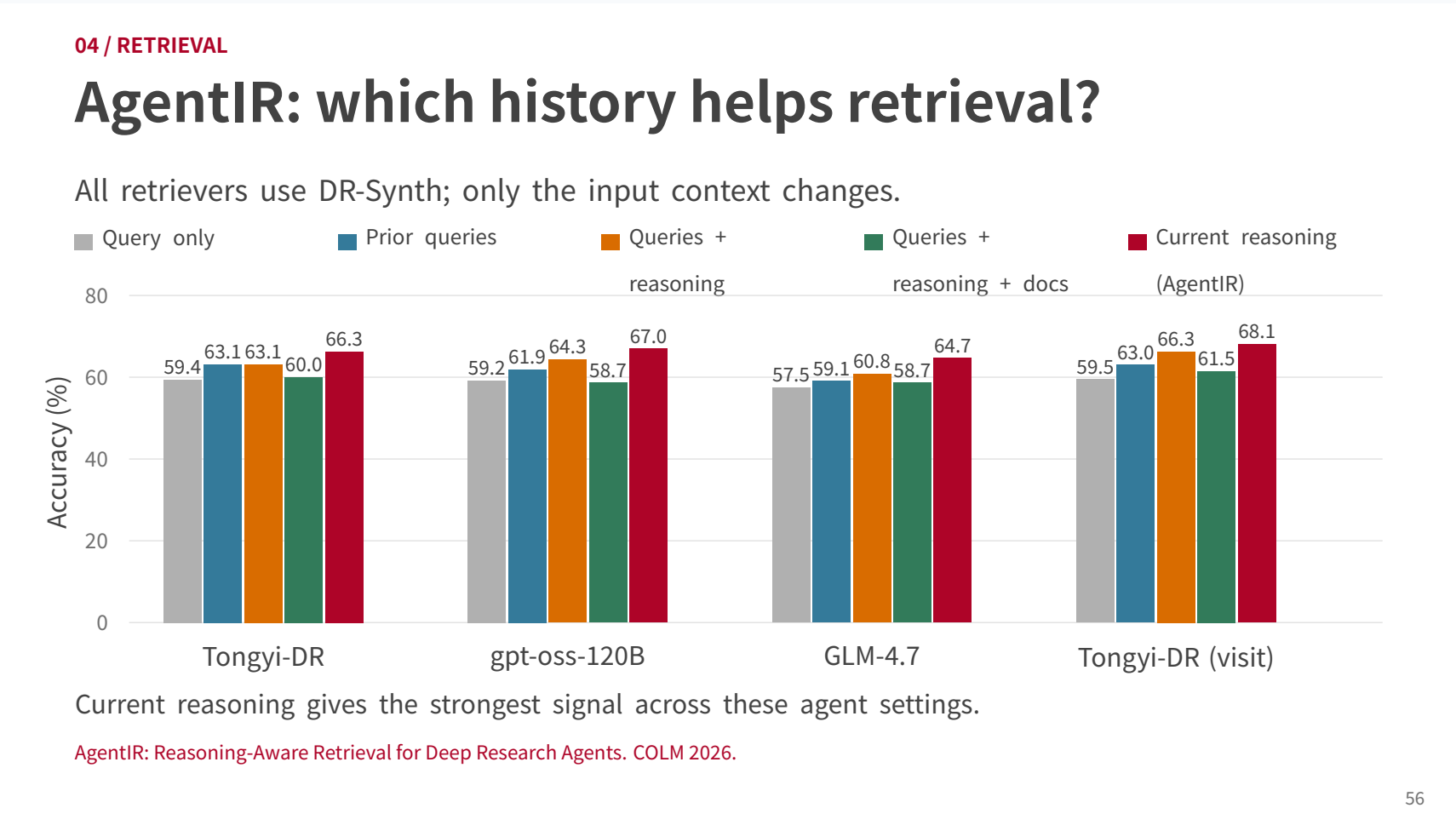
准确率按柱状图标注值抄录。

AgentIR：哪些历史信息有助于检索？

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分



04 / 检索

AgentIR：哪些历史信息有助于检索？

所有检索器都使用 DR-Synth；只有输入上下文不同。

图例：仅查询；先前查询；查询 + 推理；查询 + 推理 + 文档；当前推理（AgentIR）。纵轴：准确率（%），刻度为 0、20、40、60、80。

Tongyi-DR：仅查询 59.4，先前查询 63.1，查询 + 推理 63.1，查询 + 推理 + 文档 60.0，当前推理 66.3。

gpt-oss-120B：仅查询 59.2，先前查询 61.9，查询 + 推理 64.3，查询 + 推理 + 文档 58.7，当前推理 67.0。

GLM-4.7：仅查询 57.5，先前查询 59.1，查询 + 推理 60.8，查询 + 推理 + 文档 58.7，当前推理 64.7。

Tongyi-DR (visit)：仅查询 59.5，先前查询 63.0，查询 + 推理 66.3，查询 + 推理 + 文档 61.5，当前推理 68.1。

当前推理在这些智能体设置中提供了最强信号。

《AgentIR: Reasoning-Aware Retrieval for Deep Research Agents》，COLM 2026。

译注 / 与原文翻译分开

准确率按柱状图标注值抄录。

第 10 讲 · 深度研究智能体：任务、基准与方法

EN / 英文官方原页 · 保持原样

ZH / 中文译文 · 第 1/1 部分

04 / RETRIEVAL

Summary

01 Tasks

Deep research combines many searches with evidence synthesis.

02 Evaluation

Assess answer quality and citation support; audit the rubrics too.

03 Training

Learn from demonstrations, then improve through task feedback.

04 Retrieval

Give retrievers the agent’s information need and train for it.

57

04 / 检索

总结。

01 任务：深度研究将多次搜索与证据综合结合起来。

02 评估：评估答案质量和引文支持情况；也要审查评分标准。

03 训练：先从示范中学习，再通过任务反馈改进。

04 检索：向检索器提供智能体的信息需求，并针对这一需求训练检索器。

总结