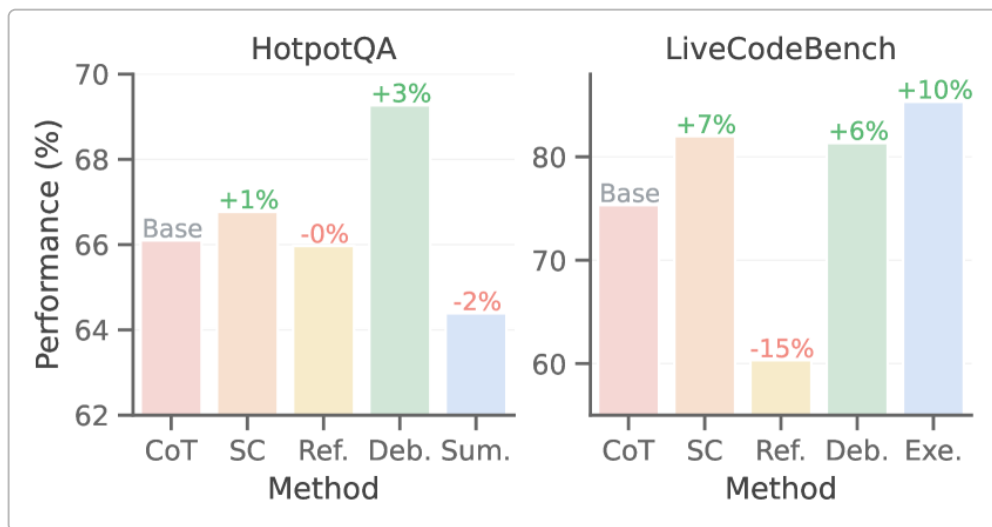


Open Deep Research 多智能体协作架构分析与优化建议

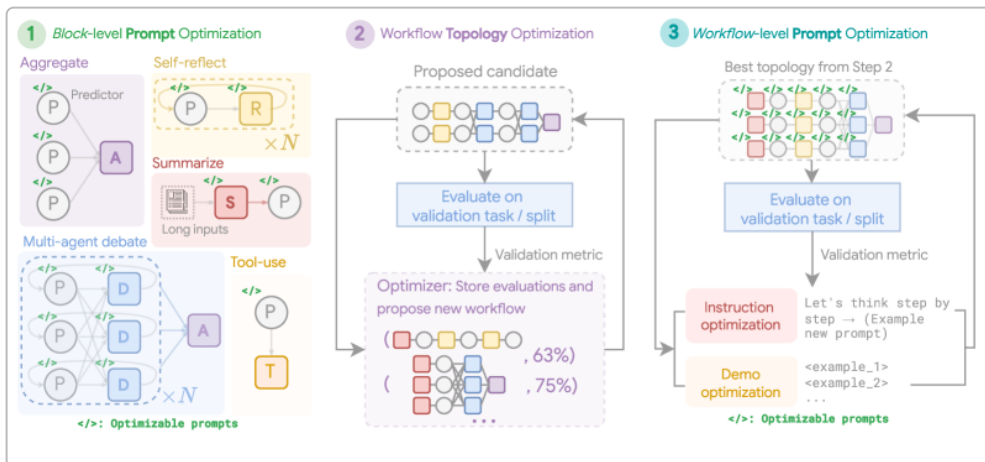
智能体协作机制



Mass框架对不同多智能体拓扑策略的评估结果表明，并非所有协作模式都有效。在HotpotQA任务上，只有Debate拓扑带来约3%的性能提升，其他拓扑影响甚微甚至略有下降；而在LiveCodeBench上，引入Aggregator（Executor）获得了最大的+10%改进¹。这一结果说明Mass认为只有少数特定拓扑对性能有实质贡献，其余拓扑可能无效或带来负面影响¹。

Open Deep Research项目当前采用的是**督导-研究者**架构：一个主管负责规划章节、汇总报告，多个研究者并行撰写各自章节²，是一种简单的并行分工设计。该方案并未利用MASS中提出的更复杂交互拓扑（如多轮反思、交叉辩论或汇总投票等）。鉴于Mass分析的启示，可以尝试引入额外的协作角色。例如，引入反思型(Reflect)智能体作为结果验证者，对研究者初步输出提出批评和改进意见³；采用辩论型(Debate)机制，让多个研究者就同一问题进行多轮讨论，相互提供观点并融合得出更准确的结论⁴；或者让多个研究者并行探索同一问题后再使用聚合型(Aggregate)策略（如多数投票、一致性投票）统一结果⁵。需要注意的是，Mass实验证明并非所有拓扑都适用于所有任务¹，盲目堆叠会增加资源消耗且边际收益递减。因此建议在现有系统中小规模试验上述机制，并结合实际任务特点评估效果。通过引入Reflect、Debate或Aggregate等机制，系统能在原有并行基础上增加协作深度和多样性，有望提升研究结果的准确性和可信度。

提示模板



MASS框架采用分层提示优化策略：左侧为块级提示优化（Block-level Prompt Optimization），中间为 workflow 拓扑优化，右侧为 workflow 级提示优化⁶。当前 open_deep_research 对督导和研究者仅有基础的角色提示，缺乏系统化的优化。根据 MASS 思想，应首先对每个智能体角色进行**块级提示优化**：为督导和研究者分别设计专门的提示模板，并通过实验迭代优化这些提示内容⁷⁶。例如，在督导提示中明确要求生成报告结构和分配研究任务；在研究者提示中给出研究策略指导和输出格式规范。可以加入示例(demonstration)来进一步明确任务要求⁶。

在完成块级优化后，进入**workflow 级提示优化**：将各角色提示结合起来，在全局层面对其进行协同调整⁸⁹。MASS 第3阶段指出，将整个多智能体系统视为一个整体进行提示调优，有助于优化智能体间的相互依赖⁸。在 open_deep_research 中，这意味着要在提示中增加系统级指令（例如要求输出包含引用、步骤说明等），确保不同智能体的输出格式和风格一致。分层式提示设计不仅能增强提示的针对性，还使得各智能体的职责边界更加清晰，有助于提高协作效率和输出质量。

任务规划与 workflow 设计

目前项目对不同任务未灵活调整 workflow 配置：无论主题为何，督导-研究者流程基本一致。MASS 研究显示，不同任务可能需要不同的拓扑结构才能获得最佳效果。例如，Mass 在数学问题(MATH)任务上的优化示例中发现：第一阶段优化后辩论拓扑表现最佳，但随着更多并行聚合器的加入，最终聚合架构超越了单轮辩论¹⁰。这说明针对不同任务，应灵活组合智能体协作模式。

因此，建议在 open_deep_research 中引入任务感知的工作流配置。可以通过整体提示或元策略来选择拓扑：对于复杂推理或有争议的问题，可启用多轮反思和讨论机制；对于信息检索和汇总任务，则使用并行查询加聚合输出。借鉴 MASS 的做法，还可对整个 workflow 进行提示优化，使系统自动适应任务需求⁸¹⁰。例如，在全局提示中加入任务描述和目标要求，让督导智能体在规划时考虑使用何种协作方式。这样的工作流级优化能增强模型的通用性和跨任务鲁棒性，使同一系统适应更多场景。

性能与成本优化

MASS 的成本分析强调**优质提示胜过增加智能体**。实验证明，通过为模型提供更丰富的指导和示例，其性能在单位 token 开销上远优于简单地堆叠更多 Agent¹¹。在优化提示之后，加入自洽性(self-consistency)策略进一步提高了性能/成本比，而纯粹扩展并行 Agent 则很快出现效益递减¹¹。因此，在资源有限的情况下，建议

open_deep_research优先通过改进提示质量和利用自洽性来提升性能，而非盲目增加智能体数量。同时，可利用缓存、并行查询等技术降低调用开销；在可能时使用较小或开源模型进行预处理，仅在关键生成环节调用高成本模型。合理控制每个Agent的查询数量和回答长度，也能有效降低推理成本。总之，应参考MASS策略，聚焦“优化单体”而非无限扩展群体，以获得较佳的性能/成本权衡 ¹¹ ¹²。

可解释性

引入MASS的分阶段设计有助于提升系统可解释性。MASS强调先优化个体，再组合系统 ¹⁰。对应到open_deep_research，即首先确保每个角色（督导、研究者）的提示和输出格式明确，再进行整体协作调优。在此过程中，可为每个Agent定义清晰的输入输出结构和行为规范，便于跟踪和分析各自的贡献。分阶段优化还使调试更容易：如首先验证单个研究者在块级提示下能独立完成任务，然后再观察他们在多智能体环境中的交互情况。

此外，MASS建议在提示中加入明确指令和示例来消除歧义 ⁷ ⁶。开放式Deep Research可借鉴这一点，为每个角色的提示添加示例输出或要求逐步解释其推理过程，以增强输出的可理解性。最后，MASS的整体提示优化阶段确保提示能反映智能体间的相互依赖，使整个系统的工作流程和决策过程更加透明 ⁸ ¹⁰。按照这一思路设计的系统，其各部分协作关系更清晰，有助于用户理解和信任生成结果。

小结

综上所述，我们建议open_deep_research在设计多智能体系统时，借鉴MASS框架的优化思路：

- **协作拓扑优化**：在现有督导-研究者并行结构基础上，尝试引入Reflect（验证）、Debate（辩论）或Aggregate（投票）等交互机制 ⁵ ⁴；同时剔除无效拓扑以降低复杂度 ¹。
- **提示分层优化**：为不同角色设计个性化提示并进行块级优化 ⁷ ⁶，再通过全局提示优化协调各部分，使智能体协作更加连贯 ⁸ ⁹。
- **工作流动态配置**：根据任务特点灵活调整智能体流程，例如复杂推理任务时增设讨论回合，事实类任务时强化并行检索和聚合 ¹⁰ ⁸。通过工作流级提示优化增强多场景下的适应性。
- **成本效益考虑**：以优化提示和自洽性为主，避免盲目增加代理数。合理使用缓存、模型大小和并行度等手段，降低token开销 ¹¹。
- **增强可解释性**：采用分阶段提示设计让各Agent角色职责明确；在提示中加入示例或要求中间步骤输出；整体提示设计突出智能体间依赖 ⁸ ¹⁰，提高系统透明度。

这些改进措施参考了MASS论文的分析 and 实验结果，有助于提高open_deep_research的协作效率、性能与可解释性。所有建议需结合实际任务需求进行验证，以获得最佳的系统表现。

参考文献：分析依据了MASS论文中的实验与结论 ¹¹ ¹⁰ ⁵ ⁸ 及Open Deep Research项目文档 ²。

¹ ³ ⁴ ⁵ ⁶ ⁷ ⁸ ⁹ ¹⁰ ¹¹ ¹² [2502.02533] Multi-Agent Design: Optimizing Agents with Better Prompts and Topologies

<https://arxiv.org/html/2502.02533>

² GitHub - langchain-ai/open_deep_research

https://github.com/langchain-ai/open_deep_research