

智枢 OSINT: 基于深度研究的多智能体情报分析框架

丘龙鹏^{1,2}, 陈波^{1,2}, 赵泽亚³, 范意兴^{1,2}, 郭嘉丰^{1,2}, 程学旗^{1,2}

1. 中国科学院网络数据科学与技术重点实验室, 北京 100190;

2. 中国科学院大学, 北京 100190;

3. 北京跟踪与通信技术研究所, 北京 100094

摘要

开源情报分析在辅助复杂决策方面扮演着日益重要的角色。然而, 面对任务复杂多变、多源异构信息处理困难以及深度推理依赖专家经验等挑战, 传统的情报分析流程效率低下。为解决这些问题, 提出了一套面向情报分析的深度研究智能体框架。该框架模拟人类情报分析团队, 构建了一套由任务规划智能体主导, 由信息收集、信息鉴别和洞察生成等专业子智能体协同运作的系统, 旨在实现从任务分解到完整情报报告生成的全流程自动化。为验证框架的有效性, 构建了针对复杂情报任务的评估数据集。实验结果表明, 该框架在生成报告质量维度上显著优于基线模型, 但在网络检索维度的引用能力方面仍有提升空间。

关键词

开源情报; 深度研究; 多智能体系统

中图分类号: TP391

文献标志码: A

doi: 10.11959/j. issn. 2096-0271.2026010

Zhishu OSINT: a multi-agent intelligence analysis framework based on deep research

Qiu Longpeng^{1,2}, Chen Bo^{1,2}, Zhao Zeya³, Fan Yixing^{1,2}, Guo Jiafeng^{1,2}, Cheng Xueqi^{1,2}

1. CAS Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

2. University of Chinese Academy of Sciences, Beijing 100190, China

3. Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China

Abstract

Open-source intelligence analysis plays an increasingly important role in supporting complex decision-making. However, traditional intelligence analysis processes are inefficient due to challenges such as the complexity and variability of tasks, the difficulty of processing multi-source heterogeneous information, and the reliance on expert experience for deep reasoning. To address these issues, this paper proposes a deep research agent framework for intelligence analysis. This framework simulates a human intelligence analysis team, constructing a system led by a "Task Planning Agent" and coordinated by specialized sub-agents, including the "Information Collection Agent," "Information Identification Agent," and "Insight Generation Agent." The framework aims to automate the entire process, from task decomposition to the production of complete intelligence reports. To validate the effectiveness of the framework, we constructed an evaluation dataset for complex intelligence tasks. The results show that the framework

significantly outperforms baseline models in terms of generated report quality (RACT), but still has room for improvement in terms of citation performance in web retrieval (FACT).

Key words
open-source intelligence, deep research, multi-agent systems

0 引言

开源情报（open-source intelligence, OSINT）是基于公开可获取的信息，通过收集、处理和分析等手段形成的情报^[1]，其战略价值日益凸显。对海量、异构且持续增长的公开数据进行深度挖掘与整合，可提炼出有价值的情报线索，辅助解决复杂问题。然而，当前的情报分析流程面临任务复杂多变、多源异构信息处理困难以及深度推理依赖专家经验等一系列严峻挑战，导致整体效率低下，难以满足快速响应、精准决策的现实需求。

近年来，以大语言模型（large language model, LLM，简称大模型）为代表的人工智能技术飞速发展，正在快速革新各个领域的技术体系，特别是以大模型为基础的认知智能体的兴起，为实现从“数据驱动”到“认知驱动”的范式跃迁奠定了关键基础。大模型在自然语言理解、知识推理、信息摘要和内容生成等方面的强大能力，使其成为进行自动化信息处理与初步分析的有力工具^[2]。然而，将单一大模型应用于复杂、多步骤的情报分析任务仍存在明显局限性。这主要体现在：它在处理复杂、多步骤任务时，缺乏宏观的规划与动态调度能力，难以独立完成从任务分解到报告生成的全流程；在处理海量、多源异构信息时，其信息整合与跨源验证能力有限，难以高效地从复杂数据中提炼出有效信息；并且，其对情报线索进行深

度推理与细节推敲的能力不足，难以从零散的信息中构建严谨的因果链或逻辑关系，这导致其难以生成真正有价值的战略洞察。

针对上述挑战，学术界与产业界已展开了一系列探索。现有工作主要聚焦于利用大模型在特定任务上的能力，如信息摘要、情感分析或实体识别^[3]，但大多未能提供一个端到端、具备动态规划和协同工作能力的系统性解决方案。因此，如何构建一个能够整合不同功能模块、实现全流程自动化与智能化协作的框架，成为亟待解决的关键问题。

本文提出了一套面向开源情报的深度研究（deep research）智能体协作框架，该框架并非简单地将大模型作为工具，而是将每个大模型作为具备特定职能的专业化智能体个体，基于“规划-执行-分析-反馈”的闭环模式，构建了一套系统化的职能分工与动态协同机制。通过这种模式，该框架能够将复杂的情报分析任务解耦为若干可由不同智能体独立完成或协同完成的子任务，并实现从任务需求接收、信息自动收集与鉴别、深度推理与洞察，到最终报告生成的全流程自动化，为开源情报分析提供一种全新的、综合性的解决方案。

本文的主要贡献如下。①提出了一套面向开源情报分析的智能体框架。该框架通过设计相互协作的智能体、精细的角色分工和工具集成，有效提升了情报信息获取的全面性、分析的洞察力，并提高了情报报告的生成效率。②构建了一套面向复杂情报任务的评估数据集。笔者人工构建了一批包含复杂推理需求的任务数据集，

为评估多智能体框架在开源情报领域的性能提供了一个标准化的基准。③系统性地验证了框架在解决复杂情报任务上的性能。基于上述数据集的实验有力地验证了本文所提框架在复杂情报任务中的有效性。

1 相关工作

深度研究框架旨在通过整合先进的大模型、信息检索技术与自主推理能力，将人类研究员的完整工作流程实现自动化。它通过模拟人类“研究”这一复杂认知过程，将任务分解、工具调用和多步推理等环节形成一个闭环^[4]，为解决开源情报领域面临的效率、数据融合与虚假信息识别等问题提供了可能。

深度研究技术的发展可以追溯到2023年，当时 n8n^[5]和 Qwen-Agent^[6]等工作流自动化框架奠定了基础。2024 年年末，谷歌 Gemini^[7]首次尝试将多步推理和知识整合应用于深度研究，而 AutoGPT^[8]和 BabyAGI^[9]等早期自主代理框架也为此提供了技术支持。此外，Cline^[10]和 Browserbase^[11]等工具在整合研究流程和浏览器自动化方面也做出了贡献。到了2025年，这一领域取得了重大突破。2025年2月，OpenAI发布了基于 o3 模型的 Deep Research^[12]，它在自主研究规划、跨领域分析和高质量报告生成方面表现出色。同时，Perplexity 也推出了免费的 Deep Research^[13]，主打快速响应和易用性。开源社区也迅速跟进，出现了多个开源替代方案^[10,14-16]，极大地加速了该领域的技术迭代与创新。

尽管深度研究技术在多个领域蓬勃发展，但其在处理开源情报分析这一特定领域的复杂性方面仍面临挑战。开源情报对

信息真实性、偏见识别、多源异构数据（包括非结构化和碎片化数据）深度融合，以及最终形成高度可信、可解释的情报报告有更高的要求。现有框架虽然提供了基础能力，但普遍缺乏针对开源情报特有挑战的定制化机制和高级认知流程。

2 基于深度研究的多智能体情报分析框架

本节详细阐述为应对开源情报分析面临的挑战而设计的智枢 OSINT 智能体协同分析框架。该框架借鉴人类情报分析团队的工作模式，构建了一套由主控智能体与多个专业子智能体协同运作的架构，旨在实现情报分析流程的智能化、自动化与高效化。

2.1 框架概述

深度研究的协同工作流遵循“规划-执行-分析-反馈”的闭环模式。其核心思想是，将复杂且多步骤的情报分析任务解耦为更小的、可管理的子任务，并分配给具备特定能力和工具集成的智能体进行处理。本文定义了4种核心智能体，它们分别是任务规划智能体、信息收集智能体、信息鉴别智能体和洞察生成智能体。这些智能体各司其职，又紧密协作，共同完成情报分析任务。

整体架构如图1所示。

智枢 OSINT 框架流程步骤如下所示。

- 需求理解：当任务规划智能体接收到情报需求后，会迅速将其分解为可执行的子任务。
- 任务规划：对于分解出的每个子任务，制定详细的执行计划，并智能调度专业智能体进行处理。
- 信息收集：随后，对于每个子任务，如果需要获取新的外部信息，调用信息收

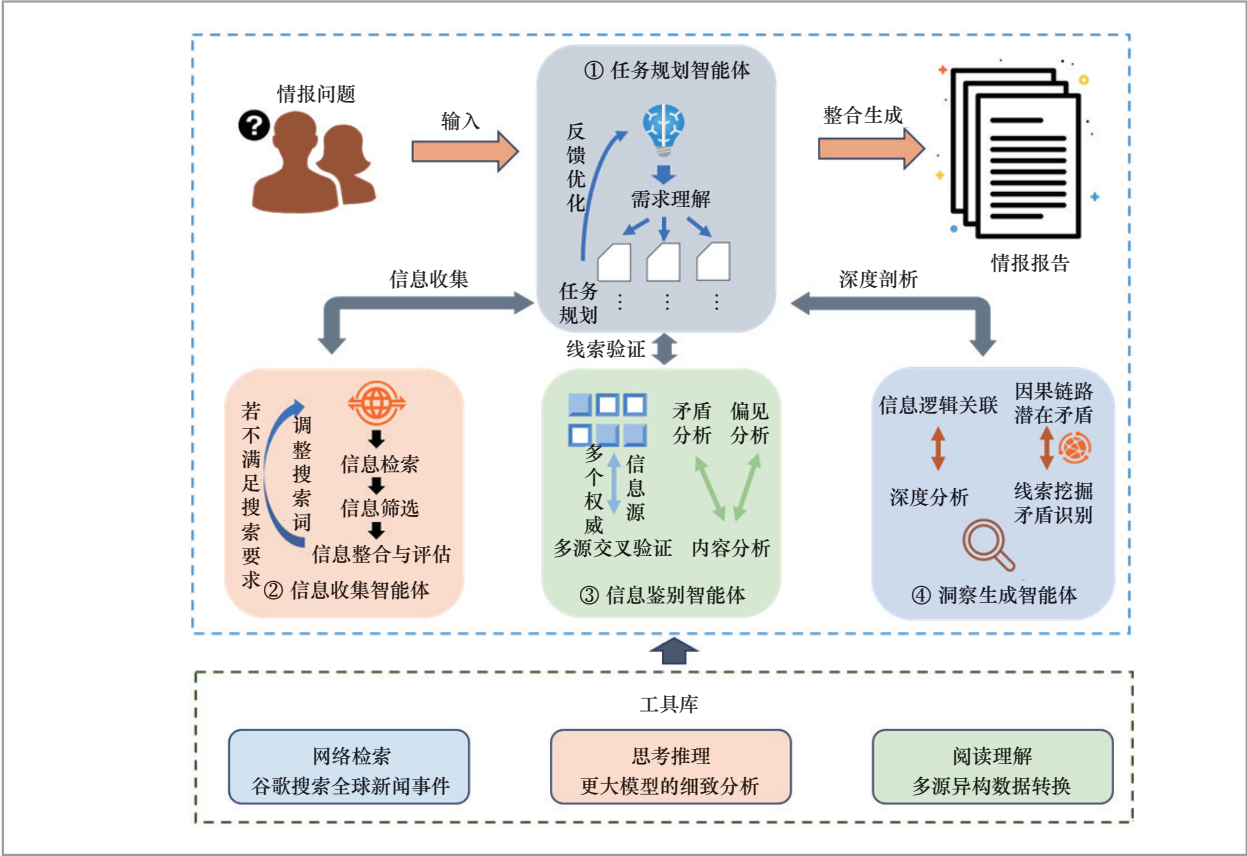


图1 整体架构

集智能体，执行具体的搜索任务。该智能体完成数据收集后，会同时对信息进行初步的可信度分析，并将整理好的数据返回给情报任务规划智能体。

- **线索验证**：任务规划智能体接收到数据后，会调用信息鉴别智能体，对收集到的信息进行多维度交叉验证，以评估其来源可靠性与内容真实性。

- **深度剖析**：任务规划智能体在接收到数据后，会进行评估。如果对信息的可信度或深度存在疑问，它会调用洞察生成智能体进行更深层次的分析。

- **反馈优化**：洞察生成智能体完成分析后，将结果反馈给任务规划智能体。主智能体根据反馈，进行迭代调整，这可能包括重新规划任务、细化搜索条件或要求

进行更深入的分析。

- **整合生成**：这个迭代循环会持续进行，直到所有情报点被充分探索和整合。最终，任务规划智能体将所有信息汇总，生成最终的情报分析报告。

这种迭代与反馈机制确保了系统能够像人类团队一样，在不断获取新信息并进行分析的过程中持续优化其理解和输出。

2.2 智能体角色与职能

本节将深入探讨4种核心智能体。它们通过明确的角色分工和协同机制，共同驱动开源情报分析流程，并承担各自的关键职能。

2.2.1 任务规划智能体

任务规划智能体是整个协同分析框架的大脑和指挥官，其核心目标是将用户提出的复杂、模糊的情报需求转化为一系列可执行、可追踪的子任务，从而解决复杂多变的分析任务，并协调其他专业智能体完成这些任务，最终整合生成完整的情报报告。它模拟了高级情报分析师的思维模式，负责整个分析流程的宏观规划、动态调度与结果整合。其具体处理流程和职能如下。

- 任务分解与规划：该智能体首先接收用户输入的复杂情报需求，将此需求分解为子问题，并对每个子问题规划一系列清晰、可执行的子任务^[17]。
- 动态调度与监控：根据分解出的子任务类型和优先级，该智能体动态选择并调度相应的专业子智能体^[18]。例如，调用信息收集智能体检索相关事件和进展结果。在整个分析过程中，它负责对任务进度进行宏观调控与实时跟踪。
- 迭代反馈：在收到子智能体的结果后，它会进行评估。如果发现信息存在疑问或深度不足，它会根据反馈进行迭代调整，例如要求洞察生成智能体进行更深层次的分析或重新规划搜索策略。
- 整合生成报告：当所有子智能体返回处理结果并完成所有分析后，该智能体将进行最终的信息整合与报告生成，确保报告的完整性、逻辑性和可读性。

2.2.2 信息收集智能体

信息收集智能体是整个系统的外部信息接入点，专门负责与外部信息世界进行交互。它能够针对每项查询，在海量公开信息中高效、准确地获取原始数据，并进行初步筛选与整合，从而有效应对多源异构数据带来的挑战。最终，该智能体提供

的是经过验证和整理的有效知识，而非原始的、未经处理的碎片信息。其核心职能包括信息检索、筛选与整合。

- 信息检索：首先解析指令中的关键词和目标，随后调用预设的搜索、阅读理解等工具，在互联网上进行海量公开信息的检索。
- 信息筛选：在获取到搜索结果后，该智能体并非直接返回原始内容。它会基于信息来源（如新闻机构、官方网站等）和内容质量（如语言流畅性、是否存在明显矛盾）进行初步的可信度评估与筛选。这一步骤旨在过滤掉明显的低质量、虚假或带有偏见的信息。
- 信息整合与评估：在对筛选后的信息进行整理和关键内容提取（如事件、人物、时间、地点等）后，该智能体评估这些信息是否充分满足了原始任务要求。如果发现信息存在不足，将返回并重新执行前两步操作，例如调整搜索词，以确保最终数据的质量和全面性。

2.2.3 信息鉴别智能体

信息鉴别智能体是本框架中负责信息真实性与可靠性深度验证的关键单元，旨在对所有输入信息进行严谨的质量控制，确保后续分析基于可信、无偏见的数据，其核心职能如下。

- 多源交叉验证：接收到一条信息后，自动在多个独立且权威的信息源（如不同国家的主流新闻社、官方机构报告、学术论文）中搜索相同的内容。如果多方来源都证实了同一信息，其可信度就大大提高。如果信息只出现在某个小众或不可靠的网站上，它就会被标记为存疑。
- 内容分析：识别信息中的矛盾点。如果一份报告前后矛盾，或者其中引用的数据与常识不符，该智能体会立即发出警告。它

还可以分析信息的情绪倾向和偏见，判断信息是否带有强烈的个人或组织立场，这有助于识别那些故意误导或带有偏见的假新闻。

2.2.4 洞察生成智能体

洞察生成智能体是该框架的核心分析单元，其职能是对原始信息进行深层次的理解、挖掘与推理。它接收来自主智能体的分析指令，通过更细致的推理分析和调用搜索工具来对比多源信息，从而给出更深入的洞察，从看似杂乱的信息中发现潜在的价值，并将数据转化为具备战略意义的、可操作的情报，有效解决深度推理依赖专家经验的难题。其核心职能如下。

- 深度分析：当任务规划智能体对收集到的信息有疑问或需要更深层次的分析时，会调用此智能体。它会接收经过初步筛选的数据，并利用更强大的推理模型进行深度分析，识别文本中的关键实体、事件、关系及情感倾向，并构建信息间的逻辑关联。

- 线索挖掘与矛盾识别：该智能体擅长进行复杂的逻辑推理和归纳分析。它能够从看似无关的碎片化信息中挖掘隐藏的情报线索，构建复杂的因果链或事件发展路径。同时，对于不确定的信息，它会调用搜索工具，对比不同来源的信息，识别并标记可能存在的矛盾、冲突或不一致之处，并将其作为重要发现回传。

2.3 工具集成

智枢 OSINT 框架的高度自动化和智能化离不开其对多种外部工具的无缝集成。这些工具为智能体提供了执行特定操作的能力，极大扩展了整体框架的功能边界。本文主要选择以下3种工具。

- 网络检索：作为核心的信息检索工

具，它使信息收集智能体能够访问互联网海量的公开数据。通过调用其 API，智能体可以指定搜索地域和语言，获取全球范围内的搜索结果。

- 阅读理解：该工具负责处理多种格式的信息输入，包括但不限于网页内容、PDF 文档、文本文件等，并将其统一转化为标准的 Markdown 格式。这解决了数据异构的难题，确保所有信息都能以文本的方式进行处理。

- 思考推理：该工具是一个调用更大型语言模型的接口，主要供洞察生成智能体在解决高度复杂、需要深层认知能力的问题时使用。

3 实验分析

3.1 数据集

鉴于当前智能体驱动情报分析领域尚缺乏直接可用的公开评估数据集，笔者决定构建一个专用的实验数据集。为确保实验的专业性和相关性，本文从“知识星球”平台的“地缘信息”^①账号下选取了30篇专业报告作为实验数据集。这些报告覆盖了当前地缘政治领域的热点事件和复杂议题，具有高度的专业性和时效性。在实验中，每篇报告的题目被当作需要解决的情报问题，而由人工专家精心编写的报告内容则作为评估输出报告的参考答案，用于衡量生成报告的准确性、深度和全面性。

3.2 实验设置

3.2.1 模型配置

框架中所有智能体的语言模型均基于通义千问（Qwen3）系列模型^[19]构建，其中，

① 已获授权基于相关数据开展研究。

洞察生成智能体需要进行复杂细致的推理分析，笔者为其配置了能力更强的Qwen3-235B-A22B-Thinking-2507模型；其他智能体主要负责任务规划、工具调用和初步信息处理，为其配置了更轻量级的Qwen3-32B模型，以平衡性能与计算成本。

3.2.2 生成参数设置

为确保框架输出的稳定性和可重复性，减少生成内容的随机性，使输出更聚焦于事实和逻辑。温度参数均统一设置为0.3，top-k设置为0.95。

3.2.3 基线方法

为全面评估本文提出的智枢 OSINT 协同分析框架的性能优势，本文选取了3种具有代表性的基线模型进行对比实验。

(1) 单一模型直接生成

将用户的情报问题直接作为输入，不进行任何外部工具调用或多智能体协作，直接生成一份情报报告。此基线模型旨在评估单一通用 LLM 在无辅助情况下的情报生成能力。

(2) 检索增强生成

此基线模型模拟了结合外部信息检索的 LLM 应用。首先，用户的情报问题通过搜索工具获得原始搜索结果，随后作为上下文输入模型中，基于这些信息生成报告。该基线模型旨在评估 LLM 结合外部搜索工具的初步能力，但缺乏深度研究框架中智能体间的高级协作和深度分析。

(3) Hugging Face Deep Research 框架^[20]

本基线模型采用一个在 Hugging Face 社区中广受欢迎的 Deep Research 相关开源框架。笔者根据其官方文档搭建并运行该框架，使其处理相同的情报问题。此基

线模型旨在与现有成熟的深度研究框架进行比较，突出本文提出的智枢 OSINT 在开源情报领域特有挑战上的创新性和针对性。

3.3 评估方法

本文采用多维度评估方法，对智枢 OSINT 生成的情报报告进行量化和定性分析。笔者借鉴了 Deep Research Bench^[21] 这项工作的评估思路，采用同样的评估方式和评估维度。具体而言，分别使用报告质量评估框架 RACE (reference-based adaptive criteria-driven evaluation framework) 和网络检索评估框架 FACT (factual abundance and citation trustworthiness) 对整体框架的不同能力进行针对性考量。

3.3.1 报告质量评估(RACE)

本文使用 Gemini 2.5 Pro 模型进行评估，对框架生成的每份报告进行细致打分。评估将从以下4个关键维度展开。

- 全面性：衡量报告对原始问题涉及的所有相关信息点的覆盖程度，是否遗漏关键事实。
- 洞察力/深度：评估报告是否超越了表面信息的罗列，提供了有价值的分析、关联和深层见解。
- 指令遵循性：考察报告是否严格遵循了用户提出的情报需求和任务指令。
- 可读性：评估报告的语言表达是否流畅、逻辑是否清晰、结构是否合理，以及是否易于理解。

3.3.2 网络检索能力评估(FACT)

本文使用大模型从生成的报告中提取陈述语句及其对应的引用源 URL，随后通

过 Jina Reader API 检索网页的文本内容，并由大模型评估该内容是否提供了足够的证据来支持该陈述。基于这些判断，计算以下两个主要指标。

- 引用准确度：该指标衡量报告引用的精确性，反映其将报告内容与正确来源关联的能力。
- 平均有效引用数：该指标量化了大模型在每个任务中检索和呈现的有价值的、可验证的支持信息的平均数量。

3.4 实验结果与分析

实验结果见表1。

3.4.1 生成报告质量分析(RACT)

RACT 维度主要衡量生成报告的质量，从表1的 RACT 部分可以看出，本文提出的智枢 OSINT 框架在综合得分上表现最佳，达到 0.6063，显著优于所有基线模型。这表明该框架在整体信息检索与初步内容真实性处理方面具有显著优势，特别是在“全面性”和“洞察力”指标上表现突出，印证了框架中“信息收集智能体”高效获取数据和“情报洞察智能体”初步识别潜在线索的有效性。此外，单一模型直接生成在 RACT 的 Overall 和各项子指标上表现优于检索增强生成，这可能是因为简单的搜索结果叠加反而引入了更多噪声。而 Hugging Face Deep Research 框架在

RACT 的各项指标上均表现最差，表明其在开源情报这类信息质量要求高的任务上仍有较大提升空间，分析其生成的报告可发现，报告字数较少，且不少为 JSON 输出格式，整体的报告质量仍有很大提升空间。

3.4.2 报告忠实性与准确性分析(FACT)

FACT 维度旨在评估报告内容对原始信息和分析结果的忠实度，从表1的 FACT 部分可以观察到：在引用准确度方面，检索增强型大模型的得分遥遥领先，其平均有效引用数也高达 10.48，远超其他基线模型。这清晰地表明，直接结合外部搜索结果的基线模型在对原始信息进行引用时展现出更高的数量优势。智枢 OSINT 的引用准确度相对较低，这可能反映了它在生成报告时更倾向于对信息进行深层次的提炼、综合与重述，而非简单地罗列或直接引用原始来源。这种生成方式虽然可能在“洞察力”等维度上表现更优，但同时也对精准溯源和规范引用带来了新的挑战。值得注意的是，Hugging Face Deep Research 框架在 FACT 的两项指标上均为 0，这主要是由于其在生成报告时，较少提供完整的、可追溯的 URL 引用来源，仅提供如“来自新华社”或“根据路透社报道”等概括性的来源描述。此外，在给出了 URL 的情况下，笔者逐一进行了验证，发现这些链接大多返回“Page Not Found”错误。

表1 实验结果

模型/框架	RACE					FACT	
	综合得分	全面性	洞察力	指令遵循性	可读性	引用准确度	平均有效引用数
单一模型直接生成	0.5716	0.5246	0.6339	0.5575	0.5790	3.87%	0.2857
检索增强生成	0.5504	0.4931	0.6184	0.5429	0.5573	57.29%	10.48
Hugging Face Deep Research	0.4593	0.4311	0.4579	0.4739	0.4882	0	0
智枢 OSINT	0.6063	0.6085	0.6559	0.5652	0.5722	17.54%	1.97

4 结束语

本文提出一套基于深度研究的情报协作分析智能体框架，旨在应对开源情报分析领域面临的诸多挑战。针对开源情报实践中普遍存在的任务复杂多变、多源异构信息处理困难以及深度推理依赖专家经验等挑战，笔者设计了一个模拟人类情报分析团队的多智能体协同框架。该框架由任务规划智能体（主控智能体），以及信息收集智能体、信息鉴别智能体和洞察生成智能体等专业子智能体构成，并集成多种工具，实现了从任务分解、信息获取、深度分析到报告生成的端到端智能化 workflows。实验结果显示，本文框架在生成报告质量（RACT）维度上表现出显著优势，尤其在“全面性”和“洞察力”指标上超越了基线模型。这充分验证了本文提出的多智能体协作机制在高效筛选、整合信息以及初步挖掘情报价值方面的有效性。然而，在报告有效引用维度上，尤其是在“引用准确度”和“平均有效引用数”方面，本文所提框架的表现仍有较大的提升空间。

参考文献：

- [1] Ghioni R, Taddeo M, Floridi L. Open source intelligence and AI: a systematic review of the GELSI literature[J]. AI & Society, 2024, 39(4): 1827-1842.
- [2] 曹树金, 曹茹烨. 从 ChatGPT 看生成式 AI 对情报学研究与实践的影响[J]. 复印报刊资料·图书馆情报学, 2023(7): 63-70.
Cao S J, Cao R Y. Influence of generative AI on information science research and practice from ChatGPT[J]. Library Science and Information Science, 2023 (7): 63-70.
- [3] Li W T, Wang C C, Cui X X, et al. COSINT-agent: a knowledge-driven multimodal agent for Chinese open source intelligence[EB]. arXiv preprint, 2025, arXiv: 2503.03215.
- [4] Xu R J, Peng J W. A comprehensive survey of deep research: systems, methodologies, and applications[EB]. arXiv preprint, 2025, arXiv: 2506.12594.
- [5] n8n-io. n8n[Z]. 2025.
- [6] QwenLM. Qwen-Agent[Z]. 2025.
- [7] Citron D. Deep research is now available on Gemini 2.5 pro experimental[Z]. 2025.
- [8] Significant-Gravitas. AutoGPT[Z]. 2025.
- [9] Yoheinakajima. BabyAGI[Z]. 2025.
- [10] Cline. Cline[Z]. 2025.
- [11] Browserbase. Open-operator[Z]. 2025.
- [12] OpenAI. Introducing Deep Research[Z]. 2025.
- [13] Perplexity. Introducing perplexity Deep Research[Z]. 2025.
- [14] Shumer M. OpenDeepResearcher[Z]. 2025.
- [15] Btahir. open_deep_research[Z]. 2025.
- [16] Assafelovic. GPT-Researcher[Z]. 2025.
- [17] Wei J, Wang X Z, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[EB]. arXiv preprint, 2022, arXiv: 2201.11903.
- [18] Park J S, O'Brien J, Cai C J, et al. Generative agents: interactive simulation of human behavior[C]//Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. New York: ACM, 2023: 1-22.
- [19] Yang A, Li A F, Yang B S, et al. Qwen3 technical report[EB]. arXiv preprint, 2025, arXiv: 2505.09388.
- [20] Roucher A, Villanova Del Moral A, Wolf T, et al. "smolagents": a smol library to build great agentic systems[Z]. 2025.
- [21] Du M X, Xu B F, Zhu C W, et al. DeepResearch bench: a comprehensive benchmark for deep research agents[EB]. arXiv preprint, 2025, arXiv:2506.11763.

作者简介



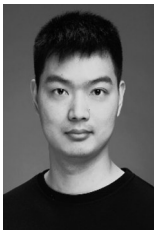
丘龙鹏（2000-），男，中国科学院大学计算技术研究所硕士生，主要研究方向为长文本生成。



陈波（1989-），男，中国科学院大学计算技术研究所工程师，主要研究方向为信息检索与大数据分析系统。



赵泽亚（1990-），女，北京跟踪与通信技术研究所助理研究员，主要研究方向为情报智能分析与知识工程。



范意兴（1990-），男，博士，中国科学院大学计算技术研究所副研究员，主要研究方向为信息检索。



郭嘉丰（1980-），男，博士，中国科学院计算技术研究所研究员，主要研究方向为信息检索、自然语言理解、机器学习与大数据分析系统。



程学旗（1971-），男，博士，中国科学院计算技术研究所研究员、副所长，主要研究方向为大数据分析系统、Web信息检索与数据挖掘。

收稿日期: 2025-10-28
基金项目: 国家自然科学基金项目(No.62372431)
Foundation Item: The National Natural Science Foundation of China (No.62372431)