SWARM-SLR AI 助手：可扩展系统综述自动化的统一框架

Tim Wittenborg*, Allard Oelen†, Manuel Prinz†
*L3S 研究中心，汉诺威莱布尼茨大学，汉诺威，德国
tim.wittenborg@l3s.uni-hannover.de
†TIB—莱布尼茨科学与技术信息中心，汉诺威，德国{allard.oelen, manuel.prinz}@tib.eu

arXiv:2603.05177v1 [cs.DL] 5 Mar 2026

摘要——尽管目前已有不断发展的工具生态系统来支持系统性文献综述（SLR），但将这些工具整合到用户友好的工作流中仍然充满挑战。用于自动化机器可操作系统性文献综述的精简工作流（SWARM-SLR）统一了工具标注，并提供了连贯且模块化的工作流，但仍面临可扩展性和可用性问题。我们提出 SWARM-SLR AI 助手，这是一种统一框架，将 SWARM-SLR 的结构化方法与基于智能体的助手相结合，通过模块化界面集成研究工具。该框架已集成 SWARM-SLR 的第一阶段，从而实现对话式的大语言模型引导支持和持久化数据存储。为应对工具评估瓶颈，我们提出集中式工具注册库，使开发者能够利用共享元数据模式自主标注和注册工具。初步评估显示可用性有所提升，但在效率、可访问性和透明度之间取得平衡方面仍存在挑战。要实现可扩展的系统综述自动化，还需要进一步发展。

索引词—文献综述、工作流自动化、工具注册表、科研软件、人工智能、大语言模型、众包

I. 引言

尽管针对系统综述（SLR）各个阶段的研究工具日益丰富，但如何将其整合为连贯的工作流仍是一大难题。面向机器执行的系统综述自动化精简工作流（SWARM-SLR）提供了完整、结构化的系统综述流程 [1]，而该工作流的首次完整应用也展现了其能力 [2]。然而，由于可用性方面的挑战，从设置工作流、在 Jupyter Notebook 环境中操作，到安装和切换多种工具，工具标注与使用的实际推广进展缓慢。

我们开发了人工智能辅助研究助手（即 AIssistant），它基于大语言模型，为不同任务和工作流提供指导，并在需要时调用外部可配置工具。SWARM-SLR 和 AIssistant 面临集成瓶颈：AIssistant 依赖对可用工具进行 JSON Schema 标注，而 SWARM-SLR 则依赖填写问卷，以确定工具在整个工作流中的能力定位。这两种方法都难以扩展为自主解决方案。

本研究提出了一个统一的工具注册表，以利用整合 SWARM-SLR 与 AIssistant 工具标注的协同潜力，提供一个统一框架，使开发者能够贡献其解决方案，并帮助用户定制和使用该系统。我们的贡献包括：**1)** 将 SWARM-SLR 集成至 AIssistant 的第一阶段，在统一界面中提供工作流指导，并配备持久化存储层，以支持人类与机器共享访问中间结果和最终结果。**2)** AIssistant 工具注册表的概念设计，集中管理两种方法的工具标注，并允许开发者自主标注工具，供用户今后选择。**3)** 对当前实现进行初步评估，由 18 名问卷受访者针对前 5 个 SWARM-SLR 步骤的实现情况及未来潜在工作提供反馈。

本文结构如下：第 II 节详细介绍我们在第 III 节中所采用方法所需的背景。评估设计与结果详见第 IV 节。这些发现与未来工作在第 V 节中讨论，并于第 VI 节作结。

II. 背景与相关工作

1) SWARM-SLR：最初的 SWARM-SLR 工作流提供了一套结构化方法，用于开展机器可操作的系统综述，识别出映射至 19 个步骤的 65 项需求。这些步骤属于构成系统综述的 8 项主要任务，并被归类为 4 个阶段。这一分类使得能够将工具映射到需求，从而形式化其在系统综述各阶段、任务和步骤上的潜在优势与不足。然而，由于配置开销和去中心化工具，实际采用受到影响。尽管 SWARM-SLR 曾在多个研讨会上展示，添加工具的流程始终未能推广。人们经常询问与他人协作的效率，以及如何避免在众多不同应用程序之间来回切换。AIssistant 有潜力通过提供一个集成平台来应对这些挑战，使用户无需离开该环境即可操作不同工具。

2) ORKG 生态系统：模块化的 SWARM-SLR 旨在利用开放研究知识图谱（ORKG） [3]。自 SWARM-SLR 发布以来，ORKG ASK [4] 已推出，这是一个基于包含 7000 万篇以上论文的预定义语料库进行文献检索的自然语言界面。类似地，John 等人 [5]进一步开发了对本地语料库创建和处理的支持。每当开发出 ORKG ASK 等新工具时 [4]，还需要将其

本工作由德国研究联合会 SE2A 卓越集群资助。

集成到工作流中，同时可能还会引入新的前端，用户需要熟悉这些前端。

3）研究工具生态系统：此前的 SWARM-SLR 工具标注旨在改进 Bosman 和 Kramer 的相关工作 [6], [7]或 Systematic Review Toolbox 等列表[1]。后者在 SWARM-SLR 论文发表前不久曾暂时停止运营，这暴露了此类系统的一个重大缺陷：一旦中心枢纽瘫痪，整个系统便无法使用。对类似的众包工具标注方法进行分析 [8]，，涵盖 Python 包索引（PyPI）等工具或软件包托管服务[2]，以及仅托管相关信息的其他平台，如 awesome 列表[3]，，发现了类似问题，但也找到了若干解决方案：设计一种标准化的标注格式，使任何人都能异步收集相关信息。与其集中存储所有元数据，不如让标注与工具本身共存（例如存放在 GitHub 仓库中），从而使社区能够继续掌控信息基础设施。尤其是工具注册表方法已证明行之有效，例如 Visual Studio Code （ VS Code ）市场[4]，Obsidian 社区插件[5] 或bio.tools [9]。这些基于扩展的环境简化了工具分发与集成。Docker 等容器化解决方案[6] 以及 Kubernetes 等编排框架[7] 则作用于生态系统的另一层，提供一致的运行时环境和可扩展部署能力。

III. 所提出的系统

我们力求将 SWARM-SLR 工作流集成到 AIssistant 中，并通过统一的工具注册表改进两个系统的模块化工具标注，如图 1所示。

1) AIssistant：AIssistant 为人机协作提供工作空间。它通过模块化界面将工具集成到工作流中，并存储中间和最终资产，例如文本或文档。虽然若干智能体在默认设置中预先定义并集成 ORKG ASK、Semantic Scholar 和 arXiv 查询等服务，是核心设计要素之一，即模块化工具定制。工具在模块化文件中定义，用于描述其功能——除文本形式描述功能和参数外，还提供工具执行功能。对函数接口的文本描述确保其与大语言模型的工具调用兼容，使大语言模型能够在认为必要时调用工具。不同工具的函数调用流程各不相同，但通常包括调用特定的 REST 端点来收集信息，且通常依据用户指定的输入。工具执行过程保持简单，以实现系统解耦，确保工具执行以及工具集成都易于进行。

2) SWARM-SLR AIsssitant：SWARM-SLR 的各个步骤在界面中集成为独立的智能体。每个智能体都有自己的系统提示、已启用工具列表以及输入/输出资产列表。将不同步骤拆分为不同的智能体有一个好处：用户能够更好地掌控流程，也更容易修改中间数据。最终，我们的目标并非提供一种完全自动化的方法，而是为用户参与 SWARM-SLR 流程提供 AI 支持的指导。

在当前布局下，要集成此功能，需要先分叉开源代码，然后提交合并请求，并由维护者审核和确认。尽管这种方法仍有改进空间，正如 工具注册表 一节所述，但对于早期的 SWARM-SLR 集成测试而言已经足够。

如图 1所示，我们将 SWARM-SLR 工作流的第一阶段整合进 AIssistant，并在前 5 个步骤中采用手动 JSON 配置。诸如“制定搜索查询”这样的步骤可能会被拆分为多个步骤（参见“综合关键词”）；如果拆分能够减少嵌套任务描述给人类和机器带来的困惑，就应进行拆分。而其他步骤，例如作为阶段 I 结尾的“步骤 2.5：与领域专家重新评估”，本质上无法自动化，因此未予实现。此外，AIssistant 已支持其他几个阶段，例如文献检索和语料库构建。在这些步骤之后，还需要安装其他工具，例如自定义 Python 包[8] 或 Obsidian，这促使我们进一步重新思考该工作流。在

[1] https://systematicreviewtools.com/ [2] https://pypi.org/ [3] https://github.com/sindresorhus/awesome [4] https://marketplace.visualstudio.com/vscode [5] https://obsidian.md/plugins [6] https://www.docker.com/ [7] https://kubernetes.io/

[8] https://pypi.org/project/bnw-tools/

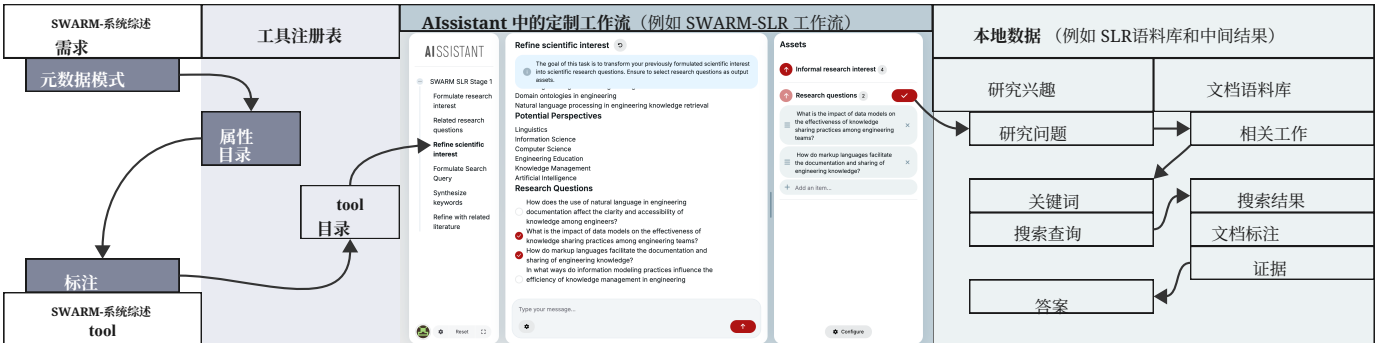

图 1. SWARM-SLR AI 助手概览。SWARM-SLR 的需求映射到工具注册表属性，这些属性用于标注工具。生成的工具目录可在 AIssistant 中使用，为人类和机器访问工具和数据提供统一的界面。

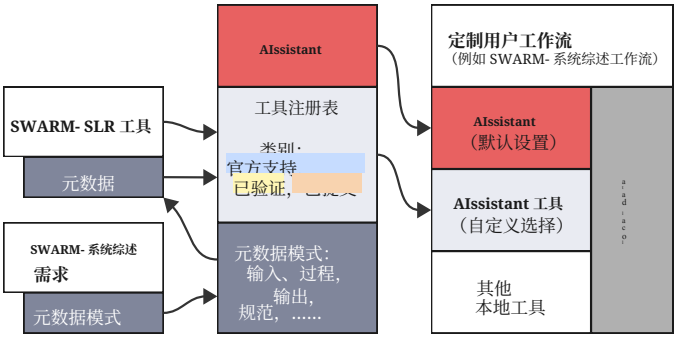

图 2. AIssistant 工具注册表的概念设计提供了一个集中式工具策展点，将需求映射到元数据模式。用户可以浏览特定任务的工具，类似于扩展市场。

当前设计中，SWARM-SLRAIssistant 无法将这些工具整合到同一界面中。为了降低完成整个工作流的门槛，我们重新设计了工具策展。

3) 工具注册表： 我们在 图 2 中展示的概念基于 bio.tools [9], 的示例；在过去十年中，该平台成功推动了超过 30,000 种工具的众包。其概念设计是一个集中式工具注册库，通过用户自主提供的数据异步填充。工具开发者可以通过提供结构统一的元数据来标注其工具，例如使用 SWARM-SLR需求模式。这样，用户便可通过一个中央系统提供和访问工具标注数据，同时即使中央系统发生故障，其工作也不会面临风险。将 SWARM-SLR 工具标注融入 AIssistant 的更大生态系统，可以发挥协同效应，使 SWARM-系统文献综述成为 AIssistant 可支持的多种布局之一。这些变化简化了工具接入，并降低了贡献工具和工作流的技术门槛。

IV. 评估

SWARM-SLR 当前版本在航空航天工程系统综述用例中经过了一年的测试。由于 AIssistant 本身仍处于早期开发阶段，尚未发布，因此目前没有针对它或其 SWARM-SLR 布局开展公开的大规模测试。随着工具能力日益复杂，完成 SWARM-SLR 工作流所需的工具集成也愈发复杂，涉及从简单的大语言模型查询，到在本地硬盘上处理、分析、重组、创建和删除数千份文档的工具。协调工作流并为如此异构的工具设置防护机制，预计会非常耗时。

1）设计：完整的调查设计可在线获取[9]通过点击 SWARM-SLR AIssistant 调查（仅查看）。其页面包括数据保护、一般信息、关于步骤 1.1 至 2.4 的问题、人口统计数据以及未来工作。一般信息说明了调查背景、设计以及仅限于截图的限制。以下 5 页展示了 SWARM-SLR，先是其原始 Jupyter Notebook 设计，然后是在 AIssistant 中的实现。在这些截图下方呈现了两种问题类型：语义差异 UEQ-S [10] 以随机顺序排列，以及一个用于任何其他评论的自由文本字段。所有问题均为可选，且始终提供“无答案”选项。人口统计数据页面询问了已完成的最高教育水平，以及他们在文献综述等主题方面的先前经验。未来工作部分使用图 1 和文字说明了拟议的工具注册表，随后是两个关于未来开发和使用该系统动机的问题。

2）结果： 18 名参与者（10 名博士、8 名硕士）完成了问卷（平均用时 22 分钟，σ：14:20 分钟，中位数：18:34 分钟）。UEQ-S 结果见 图3。显然，AIssistant 优于此前基于 Jupyter Notebook 的实现。七名参与者在文本框中提供了补充反馈，内容从用户界面建议到详尽的意见不等。一名受访者表示，在本地集成开发环境中使用模型上下文协议（MCP）或许能取得相当的效果，“同时保留你所有的工作环境和扩展，而且无需了解助手的术语和‘工作流’。”由于 SWARM-SLR AI 助手的主要目标是为用户提供更多指导，因此，即便通过一个更复杂的设置流程也能达到相同效果，这一点可能并不那么重要。总体反馈非常积极：“*AIssistant* 大大简化了 *Notebook* 用户界面，应能让工具更易使用，从而更易于普及。”11 名参与者表示愿意使用 SWARM-SLR AI 助手，其中 3 人也愿意采用 Jupyter Notebook 方案。两人希望在工作流中加入自己的工具，另有两人希望加入自己的步骤。与此同时，也有人提出了对额外能源消耗及其他成本的担忧，并指出人工智能固有的透明度、幻觉和可复现性问题。评估揭示了一个极具价值的观点：研究工作抽象得越多，人类对质量和可靠性的监督就越少。自动合成信息所节省的时间，到达某个临界点后，就会被用于确保流程及其产出的知识可靠。未来，要实现易用且可靠的系统综述自动化，首先要靠高效传达其处理过程来推动发展。

3) 效度威胁： 有一些值得注意的局限性可能会影响我们工作的普适性。18 名参与者的样本量较小，这本身就构成了限制；尽管结果总体上非常积极，但参与者中可能存在选择偏差，或其他类似因素影响了这一结果。要排除这种可能性，就需要扩大评估规模，但由于 AIssistant 尚未发布，这一目标受到阻碍。我们的工作评估了该系统开发的早期阶段，有助于快速变化的人机协作领域开展类似的开发，但这并不等同于对未来系统进行完整评估。然而，对完整系统进行评估将耗费大量时间，而且可能需要将评估拆分为基于任务的基准测试，以便与现有系统综述系统进行比较。同时，实际

[9]http://github.com/borgnetzwerk/tools/tree/main/scripts/SWARM-SLR/data

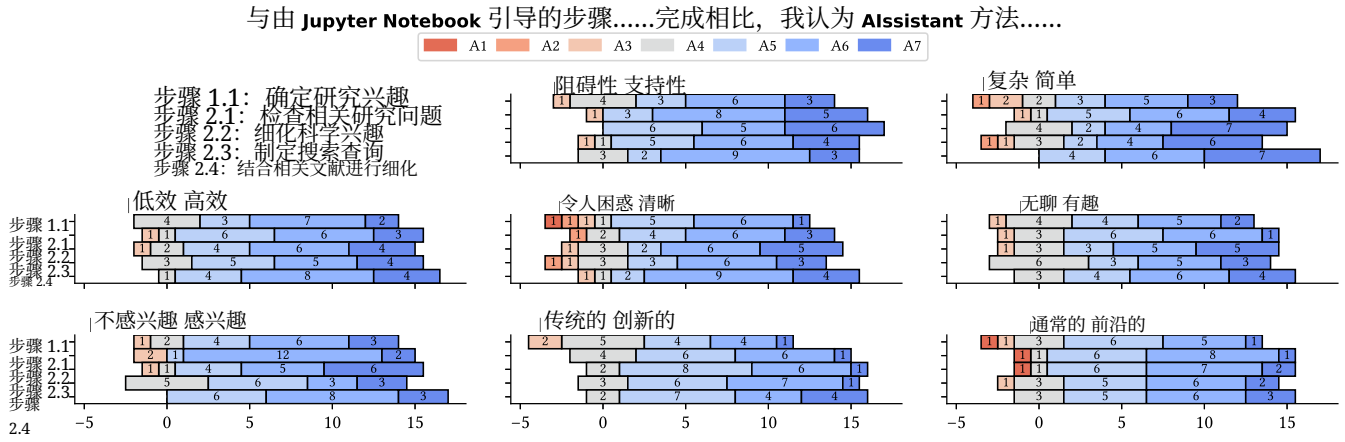

图 3. 用户体验问卷结果（N=18）。总体而言，情感倾向明显为正面。在每个步骤的累计结果中，各项指标有 2–8 项负面倾向和 62–74 项正面倾向。令人困惑 — 清晰 的负面票数最多（8），也是唯一获得强烈负面评价的指标，但另有 3 项强烈常规 评价除外。阻碍 — 支持 的正面票数总计最多（74），其中 21 项为强烈正面评价。复杂 — 容易 多出 6 项强烈正面评价（27），而传统 — 富有创意 最少（8）。

部署的影响以及长期使用反馈目前尚不可得，而这些信息是全面评估我们这项工作的实际价值所必需的。

V. 讨论与未来工作

对 SWARM-SLR 及系统综述自动化进行评估，总体而言仍难以实现规模化。若完整进行评估，耗时过长；若进行分段，则会过度简化并缺乏上下文。SWARM-SLR AI 助手的实现尤其适用于基于智能体或 API 的工具（如 ORKG ASK 或 Zotero），对经过 Python 软件包封装的解决方案（如 SWARM-SLR 的 bnw-tools）支持有限，并且未来还需要进一步整合 Miro 或 Obsidian 等更加异构的软件。该 AIssistant 旨在成为支持任何学术工作流的统一框架。然而，在能耗/时间/财务成本、效率、透明度、互操作性与可复用性、清晰度以及用户体验问卷中提及的诸多其他因素之间进行多目标优化时，仍存在严重限制。

VI. 结论

我们介绍了 SWARM-SLR AI 助手，这是一个模块化的智能体框架，集成了 SWARM-SLR 工作流指南。借助持久化数据层，我们展示了将对话式大语言模型框架与第一阶段的系统文献综述需求相匹配的可行性。此外，我们提出了工具注册表，以简化工具标注、支持去中心化贡献，并降低扩展研究工作流的门槛。18 位受访者参与了初步评估，结果凸显了可用性和可访问性的提升，同时也指出工具透明度和资源效率方面仍面临挑战。尽管 SWARM-SLR AI 助手是迈向易于使用的模块化系统综述自动化的一步，但要实现这一目标，未来仍需付出大量努力，构建一个由众包推动的科研软件生态系统，以支持易于使用且透明的工作流。

致谢

在准备本作品期间，作者使用了**GPT-4.1 (GitHub Copilot)**、**DeepL**、**Grammarly（浏览器插件）**、**LanguageTool（浏览器插件）**，用于：**翻译文本**、**检查语法和拼写**、**改写和重述**，依据CEURGenAIUsageTaxonomy[10]。

参考文献

[1] T. Wittenborg、O. Karras 和 S. Auer，“SWARM-SLR——面向机器可操作系统性文献综述的精简工作流自动化”，载于连接数字图书馆的理论与实践（A. An- tonacopoulos、A. Hinze、B. Piwowarski、M. Co ustaty、G. M. Di Nunzio、F. Gelati 和 N. Vanderschantz 编），（沙姆），第 20–40 页，施普林格·自然瑞士，2024年。[2] T. Wittenborg、I. Baimuratov、L. K. Franz´en、I. Staack、U. R¨omer 和 S. Auer，“知识驱动的航空航天工程——一项系统性文献综述”，2025年5月。arXiv:2505.10142 [cs]。[3] S. Auer、A. Oelen、M. Haris、M. Stocker、J. D’Souza、K. E. Farfar、L. Vogt、M. Prinz、V. Wiens 和 M. Y. Jaradeh，“利用知识图谱改善科学文献的获取”，《图书馆：研究与实践》，第 44 卷，第 3 期，第 516–529 页，2020年。[4] A. Oelen、M. Y. Jaradeh 和 S. Auer，“ORKG ASK：一种神经符号学术搜索与探索系统”，第*20*届语义系统国际会议海报、演示、研讨会和教程联合论文集，2024年。[5] L. John、T. Wittenborg、S. Auer 和 O. Karras，“神经符号学术知识组织的人在回路工作流”，2025年6月。arXiv:2506.03221 [cs]。[6] J. Bosman 和 B. Kramer，“乐于协同工作的工具——每种工具的模板”，2023年。[7] J. Bosman 和 B. Kramer，“400+ 学术交流中的工具与创新——数据收集表”，2023年。[8] T. Wittenborg、R. Gey、O. Karras、D. Mietchen 和 A. Struck，“科研软件发现：我们希望如何搜索科研软件，又希望在哪里找到它？”，技术报告，Zenodo，2025年5月。[9] J. Ison、K. Rapacki、H. M´enager、M. Kala˘š、等，“工具与数据服务注册库：记录生物信息学资源的社区协作”，《核酸研究》，第 44 卷，第 D38–D47 页，2016年1月。[10] M. Schrepp、J. T. Chewski 和 A. Hinderks，“用户体验问卷（UEQ-S）简版的设计与评估”，2017年12月。已接受：2021-09-13T10:56:52Z 出版方：《国际交互式多媒体与人工智能杂志》（IJIMAI）。

[10] https://ceur-ws.org/GenAI/Taxonomy.html