

AI4Drawing-学术研究

产品进攻点

- 高质量科研结构图草稿生成 + 可编辑交付 + 复刻元信息包 + 结构正确性验证 (LLM 候选 + Checker + Repair) + 多图一致性。

A. 快速产出型 (2/3/4 区, 应用向会议/期刊可行)

A1. 科研结构图草稿 “可用性” 评测体系：从主观审美到可量化指标

研究问题

如何定义并验证 “科研结构图草稿” 的可用性指标，使其与真实科研工作流强相关（草稿不是终稿，但必须能显著降低定稿成本）？

核心创新点

提出一套针对 “草稿” 的指标体系，不再以美观为主，而以**结构正确性 + 编辑成本 + 一致性**为核心。

可行方法

- 定义指标四件套：
 - a. 可渲染/可导入通过率 (Mermaid/PlantUML/draw.io/PPTX)
 - b. 结构完整性 (必备节点/边、关键链路覆盖、孤立节点数)
 - c. 编辑距离 (草稿 IR → 专家终稿 IR 的最小修改代价)
 - d. Time-to-final (用户定稿耗时, AB 实验)
- 形成一个小规模 benchmark (例如 100-300 个图套件样本, 覆盖方法/架构/流程三类)

评测/实验

- 人类专家制作终稿 IR 作为 gold; 评估不同系统输出草稿的编辑距离与时间节省。
- 对比: 传统模板法 vs LLM-only vs LLM+Checker vs LLM+Checker+Repair。

适合投稿

应用型 AI/可视化/软件工具与测评类期刊与会议 (容易落地、结果可复现)。

A2. “复刻元信息包” 标准：科研图的符号表/风格表/概念字典如何最小完备

研究问题

为了让科研人员在传统工具中快速定稿，草稿系统需要输出哪些最小元信息集合 (图标、配色、布局规则、术语映射) 才能最大化复刻效率？

核心创新点

把“草稿→终稿”的信息断层形式化：提出**Replication Pack**的最小完备字段集，并用用户实验验证其充分性。

可行方法

- 设计多种“元信息包版本”（精简版/标准版/富版）
- 用户实验：给同一草稿图，分别配不同元信息包，让用户在 draw.io/PPT 中定稿；统计 Time-to-final 与错误率。

评测/实验

- 指标：定稿耗时、修改次数、术语一致性错误、颜色/符号不一致率。
- 结果将直接指导产品规格，也可形成规范化输出标准。

适合投稿

HCI/软件工程实践/可视化应用类期刊会议。

A3. “LLM 生成 + 约束检查” 在科研结构图上的收益边界：哪些错误能被系统性压下去

研究问题

在科研结构图任务中，LLM-only 的主要失败模式是什么？引入约束检查（Checker）后，哪些失败模式能被显著消除，哪些仍需人工确认？

核心创新点

给出一个系统性的 failure taxonomy（错误类型谱系），并用实验量化 Checker 的收益边界。

可行方法

- 错误类型：孤立节点、缺失关键链路、未定义引用、循环/层级冲突、跨图术语漂移、颜色 token 越界等。
- 实现一个轻量 Checker，统计拦截率、误报率、修复成功率。

评测/实验

- 输出：一份“错误类型 × 频率 × 可自动修复性”矩阵。
- 这类结果往往很受应用社区欢迎，且能形成后续顶会工作的基座。

A4. 多图一致性管理的最小机制：全局符号表/风格表在科研图套件中的作用

研究问题

论文/基金通常是一套图。如何用最小机制保证跨图一致性（术语、符号、颜色、模块命名）并量化其对理解与定稿效率的影响？

核心创新点

提出跨图一致性指标与最小实现（StyleSheet + SymbolTable + Glossary），验证它对用户效率与错误率的提升。

评测/实验

- 组别：无全局表 vs 有全局表（锁定）
 - 指标：一致性错误率、定稿时间、用户主观理解评分。
-

B. 高影响型（顶会/顶刊潜力：需要更强的理论与系统闭环）

B1. 科研结构图的通用中间表示 IR：可验证、可编译、可编辑的“图形编译器”范式

研究问题

能否为科研结构图建立类似编译器的形式化框架：
自然语言/文档 → 图 IR（AST）→ 多目标后端（draw.io/PPTX/SVG/Mermaid）
并在 IR 层面完成可验证性与一致性约束？

核心创新点

- 把“画图”变成“编译”：IR 是核心贡献。
- IR 同时表达结构语义（节点/边/分组/层级/泳道/异常）与风格 token（可复刻）。
- 支持多后端确定性编译，确保可编辑交付物。

技术难点（顶会点）

- IR 的表达力与最小完备性证明（覆盖科研高频图类的同时保持简洁）。
- 约束系统与静态分析：对图结构做形式化校验，给出可解释错误报告。

评测/实验

- 端到端：成功生成并导入多后端的通过率；
- IR 级：编辑距离、结构正确率、跨图一致性；
- 用户研究：Time-to-final。

适合投稿

软件工程/程序语言（PL）与 AI 工程交叉、顶级可视化/HCI、以及 ML Systems 方向。

B2. 约束满足式生成：把 LLM 从“生成器”降级为“提案器”，通过 Checker/Repair 达到可证明可靠

研究问题

如何将科研结构图生成建模为“约束满足 + 搜索/修复”的问题：LLM 只提出候选，系统用约束与修复策略收敛到满足规范的草稿？

核心创新点

- 把“幻觉”从模型缺陷转化为系统可控变量。
- 引入 Repair 策略（最小编辑距离修复、冲突最小化、必要时输出人工确认点）。
- 在强逻辑任务上实现比 LLM-only 更高的可靠性下界。

技术路线

- Plan → IR candidate → Check → Repair（迭代）
- Repair 可用：规则修复、局部重生成、搜索（beam/MCTS）、或混合策略。
- 关键是建立“可证明”与“可解释”的校验链条。

评测/实验

- 可靠性：结构错误率显著降低；
- 可解释性：每次修复输出变更日志；
- 代价：推理开销与收益曲线（可靠性-成本 trade-off）。

适合投稿

顶会 ML Systems / Agents / SE+AI 交叉；顶刊软件工程与智能系统方向。

B3. 多图一致性作为一等任务：跨图概念对齐、符号一致与“全局状态”的学习/推理

研究问题

如何在生成“图套件”时维护全局一致性状态（符号表/风格表/术语表），并在跨图修改时保持一致性不崩溃？

核心创新点

- 把一致性从“提示词技巧”升级为系统任务：全局状态是模型外部事实。
- 概念对齐（concept alignment）：同一概念在不同图中如何映射到同一符号/颜色/命名。
- 支持跨图联动修改（例如模块改名，所有图同步变更且保持布局约束）。

可行方法

- 构建概念字典与别名系统；
- 约束：同概念同 token；
- 需要时做检索增强（RAG）以复用已有定义。

评测/实验

- 一致性指标体系 + 用户理解实验；
 - 与单图生成方法对比，证明“套件化一致性”价值。
-

B4. “草稿到终稿”的人机协作最优分工：何时 AI 生成、何时人类编辑、何时系统自动修复

研究问题

在高风险科研图交付中，如何形式化地优化人机协作流程，使总成本最小、错误最少？

核心创新点

- 将流程建模为决策问题：AI 生成/自动修复/请求人工确认的策略最优化。
- 目标函数：Time-to-final、错误风险、用户认知负担、返工成本。
- 可以形成理论化贡献（决策策略、交互策略），也可以形成系统贡献（交互闭环实现）。

评测/实验

- 用户研究不可或缺：不同策略对效率与信任的影响。
- 输出：协作策略建议与可复现实验。

适合投稿

顶级 HCI / 可视化 / 人机协作智能系统方向。

B5. 数据与基准：以 IR/复刻包为标注对象的科研结构图数据集与评测基准

研究问题

能否建立公开数据集与基准，标注不是 PNG，而是“Diagram IR + 复刻元信息包 + 约束集合”，用于系统性推进领域发展？

核心创新点

- 数据标注目标从“图像”转向“可编辑结构资产”，对科研场景更实用。
- 评测不仅看渲染成功，还看结构正确、跨图一致、编辑距离与 Time-to-final。

可行路径

- 从开源 draw.io/PlantUML/Mermaid 仓库抽取结构（注意许可合规）。
- 建立专家小规模标注集作为高质量 test set。
- 发布 baseline：LLM-only、模板法、LLM+Checker。

顶会潜力原因

基准与数据集往往是高影响工作的核心组成部分，尤其当它对齐真实工作流并可复现。

C. 论文选题建议：按团队能力从小到大组合成“可连续发表”的研究线

为了兼顾学生团队可落地与长期顶会潜力，建议形成一条阶梯式路线：

1. 第一篇（快）：A1 + A3

- 草稿评测体系 + LLM-only failure taxonomy + Checker 增益实证
- 目标：构建话语权与可复现 benchmark

2. 第二篇（中）：A4 + A2

- 跨图一致性机制 + 复刻包最小完备性（用户实验）
- 目标：形成你产品最关键的差异化“标准化交付物”

3. 第三篇（冲顶）：B1 或 B2（最好合并为“图形编译器 + 约束满足生成”）

- IR 设计 + 多后端编译 + Check/Repair 可靠闭环
- 目标：顶会/顶刊级系统贡献

D. Agent 范畴学术问题研究

D1. 面向科研结构图草稿生成的 Agent 动作空间与观测空间对齐方法

研究问题

在科研结构图草稿生成与编辑任务中，如何设计对 LLM 友好且可验证的 **Observation/Action Space**，使 Agent 的决策更稳定、结构错误更少、修复成本更低？

核心创新点

- 将图生成/编辑抽象为一组可组合的“动作原语”（AddNode/AddEdge/Group/Align/ApplyStyleToken/Export/RunChecker/RequestClarification）。
- 将当前图状态以结构化 IR（AST）+ 校验摘要 + 全局符号表的形式作为观测，降低自由生成带来的漂移。
- 研究动作粒度（粗粒度 vs 细粒度）与稳定性/效率的权衡，给出可复用的动作空间设计准则。

评测/实验

- 对比多种 action granularity：一步生成整图 vs 分步构建（节点→连边→分组→样式）。
- 指标：结构错误率、修复轮次、编辑距离、推理步数、可渲染/可导入通过率。
- 失败案例分析：哪些观测信息缺失会导致幻觉/漂移。

适合投稿

Agent Systems / ML Systems / 软件工程+AI / 应用型可视化与工具方向（中高水平会议/期刊）；若提出通用设计准则与系统化实验，可冲顶会相关赛道。

D2.Checker 驱动的可验证 Diagram Drafting Agent：从候选生成到约束满足的闭环控制

研究问题

如何构建一个以 Checker 为裁决核心的 Agent 框架，使 LLM 仅作为“候选提案器”，通过 **Check→Repair** 迭代达到可证明的结构正确性与一致性？

核心创新点

- 提出“候选—校验—修复”闭环：LLM 提案 IR，Checker 判定约束满足，Repair 以最小编辑距离修复。
- 约束分层（语法/结构/逻辑/一致性/风格 token）与“风险标注”机制：不确定处不硬猜，显式输出需人工确认点。
- 形成可解释的“修复日志”，提升科研场景的可信度与可审计性。

评测/实验

- 基线对比：LLM-only vs LLM+Checker vs LLM+Checker+Repair。
- 指标：通过率、结构错误率、修复成功率、误报率、生成开销、人工确认次数。
- 消融：移除某一类约束（如跨图一致性约束）对结果的影响。

适合投稿

ML Systems / Agents / 软件工程智能化方向；若引入形式化约束语言与可证明性质，可向顶会/顶刊冲刺。

D3.跨多幅科研图的一致性记忆：符号表/风格表/术语表作为 Agent 的结构化长期记忆

研究问题

在生成“图套件”（方法图+架构图+流程图）时，如何让 Agent 维护并更新全局一致性状态，使同一概念在多图中符号、颜色、命名保持一致，并支持跨图联动修改？

核心创新点

- 将跨图一致性显式建模为结构化记忆：SymbolTable + StyleSheet + Glossary。
- 研究记忆策略：何时写入、何时锁定、何时回滚，避免记忆污染导致错误扩散。
- 提出跨图一致性度量（符号一致率、命名一致率、token 一致率）与“冲突仲裁”机制。

评测/实验

- 组别：无记忆 vs 仅风格记忆 vs 风格+符号+术语全记忆；锁定策略不同的对比。

- 指标：跨图一致性、错误传播率、用户定稿时间、返工次数。
- 任务：对同一研究内容生成 3 幅图并在修改需求后保持全局一致。

适合投稿

HCI/可视化/协作系统、Agent 系统方向；强调“套件级生成”与一致性可冲更高水平 venues。

D4.多代理协作的 Diagram Drafting Pipeline：专门化 micro-agents 与仲裁机制

研究问题

如何将科研图草稿生成拆分为多个专门化代理（规划/生成/风格/一致性/校验修复），并设计有效的仲裁与通信协议，使系统整体在可靠性与效率上优于单体 Agent？

核心创新点

- 设计 micro-agent 分工：Plan Agent（抽取结构与约束）、Draft Agent（IR 候选）、Style Agent（token 分配+复刻包）、Consistency Agent（概念对齐）、Check/Repair Agent（闭环收敛）。
- 提出“仲裁器”选择策略：当代理输出冲突时如何决策（基于约束优先级、置信度、历史表现）。
- 形成一套可复用的 message schema（agent 间协议）与任务编排策略。

评测/实验

- 对比：单体 Agent vs 多代理流水线；不同分工组合的消融。
- 指标：通过率、结构错误率、修复轮次、总 token 成本/时间成本。
- 分析：哪些子任务适合分离成专门代理，哪些合并更划算。

适合投稿

Agents / ML Systems / 工程化智能系统方向；若提供通用编排与仲裁理论，可冲顶会系统类赛道。

D5.最小澄清问题集：高风险科研图任务中的主动提问与多候选交付策略

研究问题

当输入需求存在歧义时，Agent 应该何时提问、提什么、问多少，才能在不增加用户负担的前提下最大限度降低结构错误与返工成本？

核心创新点

- 提出“最小澄清问题集”策略：把歧义定位为可枚举变量（连接关系/层级归属/对比项/异常分支等）。
- 设计“双交付模式”：
 - 模式A：提 1-3 个关键澄清问题后生成单草稿；

- 模式B：不提问，输出 2-3 个候选草稿并标注风险与推荐。
- 将提问/多候选选择建模为决策问题（信息增益 vs 用户成本）。

评测/实验

- 用户研究：比较“先问再画”与“多候选直接交付”的效率与满意度。
- 指标：Time-to-final、错误率、用户认知负担（主观量表）、返工次数。
- 结果输出：可执行的澄清策略指南与可复现实验设置。

适合投稿

顶级 HCI / 人机协作智能系统 / 可视化交互方向；该课题天然适合做高质量用户研究，顶会潜力大。

D6.面向科研结构图草稿的 Agent 评测环境：DiagramDraftArena 的任务定义与可复现基准

研究问题

如何构建一个可复现的 Agent 评测环境，使科研结构图草稿生成从“主观评判”变为“可测、可比、可迭代”的标准任务？

核心创新点

- 定义环境：输入（文本/结构要点）—工具（编译器/渲染器/Checker/导出器）—成功判据（约束通过+交付物完备）。
- 将“草稿交付”定义为一等任务：必须输出可编辑源文件+复刻包+校验报告。
- 提供评测协议：结构正确性、编辑距离、跨图一致性、Time-to-final 的统一计算方式。

评测/实验

- 基线：模板法、LLM-only、LLM+Checker、LLM+Checker+Repair、单体 Agent vs 多代理。
- 指标：成功率、成本曲线（成功率 vs token/时间）、稳定性（重复生成差异）。
- 输出：公开 benchmark + baseline 实现 + leaderboard 设计（可选）。

适合投稿

Agents / ML Systems / Evaluation & Benchmarking 顶级与高水平 venues；若基准被社区采用，影响力上限高。

D7.草稿到终稿的人机协作最优分工：基于策略的生成/修复/人工确认调度

研究问题

在高风险科研图交付中，如何形式化地优化人机协作流程，使总成本最小、错误最少？

核心创新点

- 将流程建模为决策问题：AI 生成/自动修复/请求人工确认的策略最优化。
- 目标函数：Time-to-final、错误风险、用户认知负担、返工成本。
- 既可形成理论化贡献（决策策略、交互策略），也可形成系统贡献（交互闭环实现）。

评测/实验

- 用户研究不可或缺：不同策略对效率与信任的影响。
- 输出：协作策略建议与可复现实验。

适合投稿

顶级 HCI / 可视化 / 人机协作智能系统方向。

建议的“论文路线组合”（Agent 角度）

- **第一篇（快、易落地）**：Checker 驱动闭环（条目2）+ 失败模式分析（可作为实验部分）
- **第二篇（差异化增强）**：跨图一致性记忆（条目3）+ 套件生成评测
- **第三篇（冲顶）**：DiagramDraftArena 基准（条目6）或 动作空间对齐（条目1）+ 系统化理论与实验
- **HCI 冲顶线**：最小澄清问题集（条目5）+ 用户研究（强论文形态）