

# [ AI for Science ] 도메인 특화 Agent - 공개 사례

('25.8.18. 이제현. [jehyun.lee@gmail.com](mailto:jehyun.lee@gmail.com), <https://jehyunlee.github.com>)

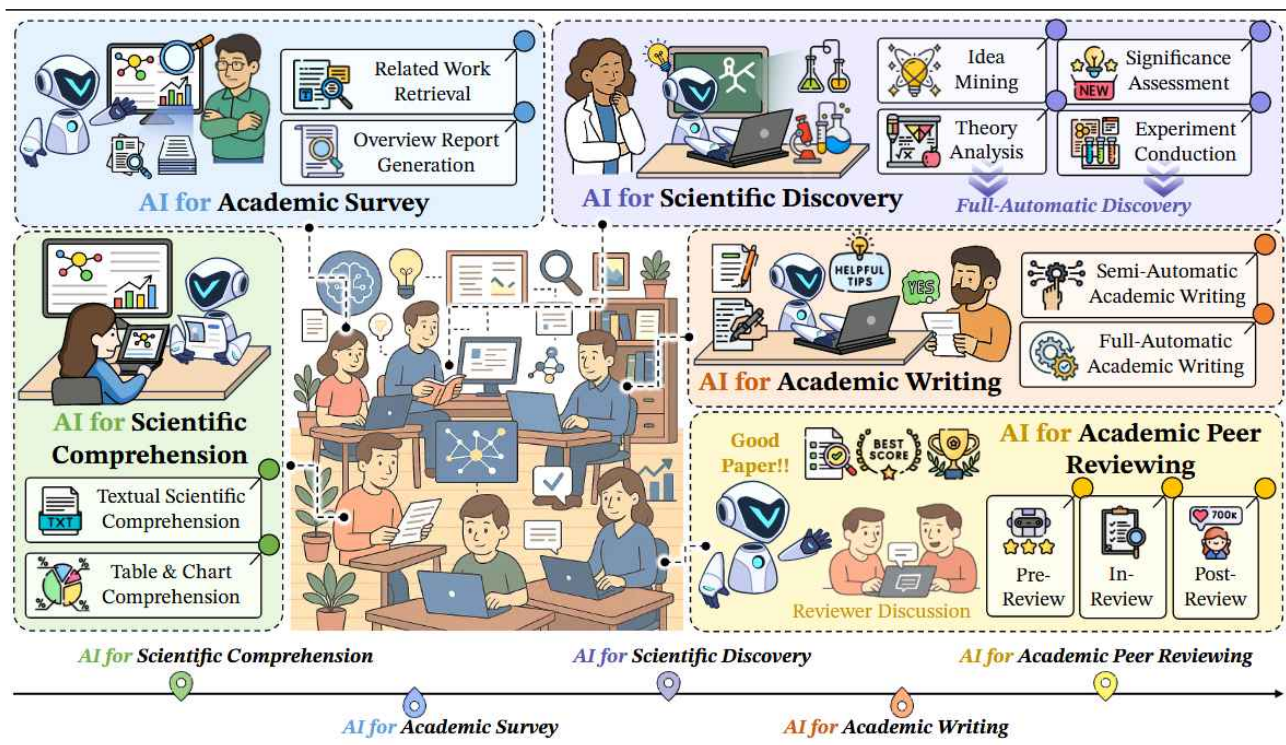
## 1 도입 및 요약

### □ R&D AI 에이전트

- (파운데이션 모델 성능 향상) 논리 추론(reasoning), 다중 감각(multimodal), 입출력 용량(context size) 증대, 응답 지연(latency) 감소 등 거대언어모델 수준을 뛰어넘는 **파운데이션 모델 성능 향상**으로 **AI를 연구 파트너로 간주**할 수 있는 단계로 발전
- (AI 모델 활용 기능 확장) 검색 증강 생성(RAG: Retrieval Augmented Generation) 과 API(Application-Program Interface), MCP(Model Context Protocol) 등 파운데이션 모델의 **도구 활용 가능성 증대에 기반한 기능 확장**
- (R&D AI Agent) AI의 의사결정능력과 도구 활용 역량을 활용하여 연구개발 전 과정을 자동화·최적화하여 **복잡한 과제를 자율적으로 해결하는 지능형 시스템**.  
※ Bayesian, Heuristic 등 경량 모델 및 소재/기계 등 도메인 특화 파운데이션 모델 활용 가능
- (에이전트 기반 연구 동향) 문헌·지식 처리형(검색/요약/리뷰), 가설/연구설계형(아이디어·계획 수립), 도구 사용(API, MCP 기반 작업), 자율실험실 등으로 구분되며 **에이전트간 의사소통을 통한 협업** 가능  
※ 문헌 탐색 → 가설 생성 → 시뮬레이션/실험/분석 → 논문 작성/리뷰 전체 또는 부분 파이프라인 연결  
※ 에이전트 도구 활용 및 에이전트간 공유 저장소를 통해 누적 성능 향상

## □ 연구 활용 AI 에이전트

- (에이전트 분류) 과학 이해, 학술 조사, 과학적 발견, 학술 작문 및 심사로 구분
  - ※ 과학 이해(Scientific Comprehension): 단일 논문에서 정보 추출, 해석, 종합
  - ※ 학술 조사(Academic Survey): 여러 기존 논문을 종합, 구조화하여 연구 분야에 대한 포괄적 개요 제공
  - ※ 과학적 발견(Scientific Discovery): 새로운 과학 가설을 수립하고 실험이나 시뮬레이션을 통해 검증
  - ※ 학술 작문(Academic Writing): 과학 원고 작성, 수정 및 형식화하여 출판 보조
  - ※ 동료 심사(Academic Peer Reviewing): 과학 원고에 대한 구조적, 객관적 검토를 제공하여 품질 및 효율성 향상
- (에이전트 기술) 연구 주제 및 에이전트의 범위에 따라 연구 주기 전체 또는 일부를 담당하며 에이전트의 동작을 원활하게 하기 위한 요소 기술 연구 활발
  - 거대언어모델은 과학적 발견 외 단계들에서 필수적으로 사용되며, 거대언어모델의 언어능력, 다중 감각 및 추론능력이 에이전트의 성능을 결정
  - 연구 에이전트의 성능은 문헌 검색, 실험 수행, 데이터 분석 등 도구 활용능력에 따라 크게 좌우되기 때문에 도구 및 호출 신뢰성 확보 필수
  - 도메인별 차이는 과학적 발견 단계에서 두드러지며, 동일 도메인에서도 연구 방법 (컴퓨터과학, 시뮬레이션, 실험연구)에 따라 다르게 데이터 편향으로 인해 일반화 난이도 높음



## □ 학술 조사 및 동료 심사

- (PaperQA2) 과학 정보 처리 에이전트 (2024)
  - 질문-증거-근거 추적 중심으로 **정보 탐색** 및 **모순 탐지 에이전트**(ContraCrow) 포함
  - **박사급 연구원보다 높은 성능** 확보
  - **단점**: 검색 및 요약이 많아 **실행 비용 높음**
- (MARG) 심사 자동화 멀티 에이전트 (2024)
  - 논문 전장을 분할병렬 처리하여 거대언어모델의 입출력 용량 제한 극복
  - **분야별 전문 평가 에이전트** 활용 및 **정서**(Refinement) 단계를 거쳐 심사 품질 향상
  - **단점**: 에이전트간 **일부 메시지 누락**, 그림 등 **시각 정보 미활용**

## □ 가설·연구 설계

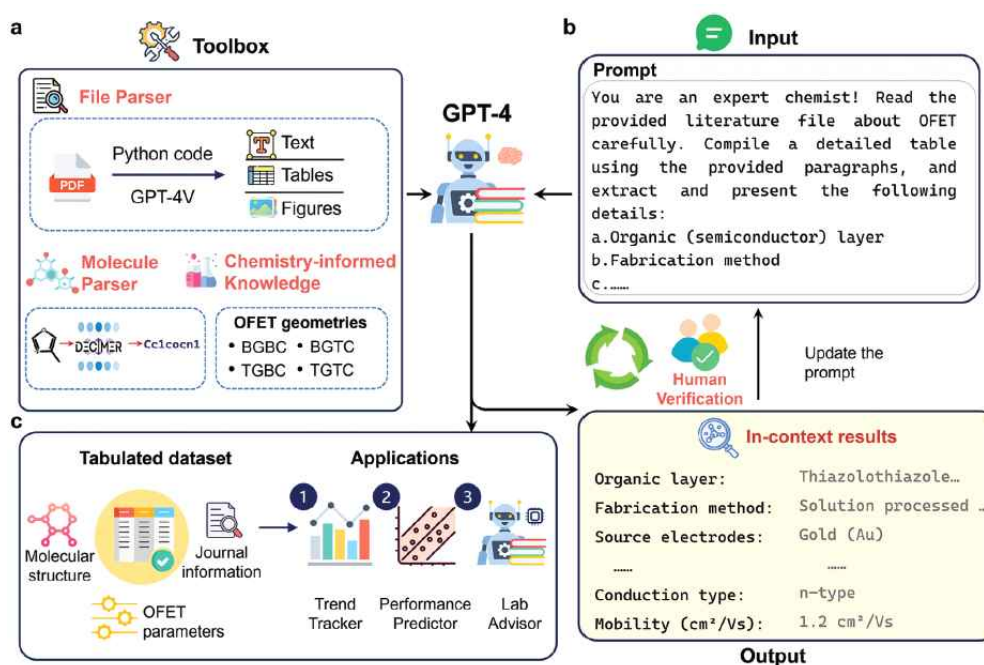
- (AI co-scientist) 분업형 다중 에이전트 (2025)
  - **연구 가설 생성·토론·진화**를 통해 일부 생물의학 가설을 **실험 검증**
  - 특정 분야의 **연구 가속 효과 확인**
  - **단점**: 생성 가설의 **참신성** 및 **타 분야 일반화 가능성 불명확**
- (Virtual Lab) 계층적 다중 에이전트 (2025)
  - PI(수석연구원), 과학 비평가 에이전트 정의, “**팀 회의**”와 “**개별 회의**”로 의사결정
  - SARS-CoV-2 나노바디 신규 설계 제시
  - **단점**: 타 분야 **일반화 가능성 불명확**, 오래된 문헌/코드 인용
- (SciAgents) 온톨로지·지식그래프 활용 (2024)
  - **two-track 전략** ① 사전 프로그래밍 상호 작용, ② 자동 에이전트 자체 조직화
  - **분업형 다중 에이전트 활용 아이디어 탐색** 및 유사 아이디어 제거
  - **단점**: LLM의 내재적 한계로 **환각 발생**, 타 분야 **일반화 가능성 불명확**

## □ 도구 사용 계열

- (AnyTool) 실세계 REST API 49분류 16,464종 제어 (2024)
  - API 문서 수집 → **지시 생성** → **API 호출**로 도구 사용 데이터셋 구성
  - **자기 반성 메커니즘**: 실패 원인 분석 후 반복적 재시도를 통한 성능 향상
  - **단점**: LLM(GPT-4)의 **성능 한계**가 **도구 호출 성능**에 직접적 영향

## □ 도메인 특화 에이전트

- (OFET Agent) 유기 반도체 문헌 데이터 추출, 특성 예측, 최적화 제시 (2025)
  - GPT-4와 ML을 결합하여 문헌 277편의 표와 그림에서 **공정/구조 파라미터 추출**
  - 실험 검증: 제안된 최적화 방안으로 전하 이동도 162% 향상
  - 단점: LLM(GPT-4)의 환각 가능성 및 타 분야 일반화 가능성 불명확
- (MOF-LLM) 추출한 문헌 데이터를 기반으로 신소재 설계 및 물성 예측 (2025)
  - 멀티모달 데이터 처리 1만건 이상 문헌의 텍스트, 표, 이미지에서 **데이터 추출**
  - 자동+수동 혼합 검증: LLM 자동화 장점과 전문가 검증 결합
  - 단점: LLM의 전문용어 처리 능력 제한 및 원문 형식 불일치로 인한 오류 가능성
- (VASPilot) MCP 활용 VASP 기반 소재물성계산 자동화 (2025)
  - MCP 활용: 다양한 도구와의 호환성 확보
  - 오류 자동 수정: 계산 중 발생하는 오류를 자동으로 파악, 조건 수정 및 재시도
  - 단점: DeepSeek 모델 대상으로 타 LLM사용시 호환성 검증 필요
- (Tippy) 제약 산업의 DMTA(Design-Make-Test-Analyze) 사이클 자동화 (2025)
  - Safety Guardrail Agent 도입: 합성 중 위험 반응, 불법 물질, 보안 침해 탐지
  - 실제 화합물 적용: Ensitrelvir 기반 COVID 약물 후보 설계 및 분석
  - 단점: 보안상 이유로 합성 파라미터 미공개, 타 분석기기 호환성 불명확

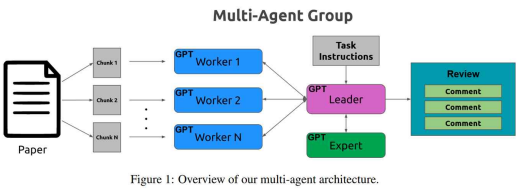
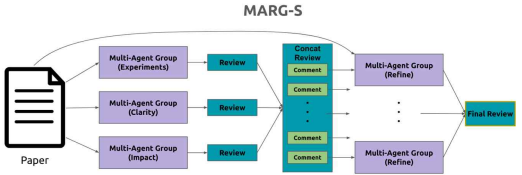
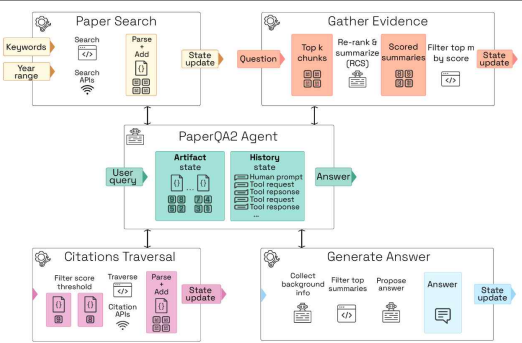
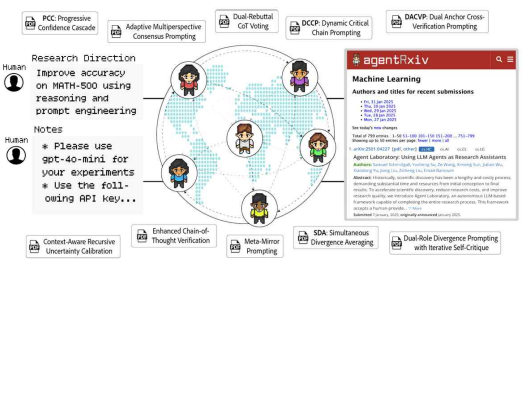


1. Collins, K. M. et al. Building machines that learn and think with people. *Nat Hum Behav* 8, 1851–1863 (2024).
2. Chen, Q. et al. AI4Research: A Survey of Artificial Intelligence for Scientific Research. Preprint at <https://doi.org/10.48550/arXiv.2507.01903> (2025).
3. Skarlinski, M. D. et al. Language agents achieve superhuman synthesis of scientific knowledge. Preprint at <https://doi.org/10.48550/arXiv.2409.13740> (2024).
4. D'Arcy, M., Hope, T., Birnbaum, L. & Downey, D. MARG: Multi-Agent Review Generation for Scientific Papers. Preprint at <https://doi.org/10.48550/arXiv.2401.04259> (2024).
5. Gottweis, J. et al. Towards an AI co-scientist. Google Blog (2025) doi: <https://doi.org/10.48550/arXiv.2502.18864>.
6. Ghafarollahi, A. & Buehler, M. J. SciAgents: Automating Scientific Discovery Through Bioinspired Multi-Agent Intelligent Graph Reasoning. *Advanced Materials* n/a, 2413523.
7. Qin, Y. et al. ToolLLM: Facilitating Large Language Models to Master 16000+ Real-world APIs. Preprint at <https://doi.org/10.48550/arXiv.2307.16789> (2023).
8. Guo, Z. et al. StableToolBench: Towards Stable Large-Scale Benchmarking on Tool Learning of Large Language Models. Preprint at <https://doi.org/10.48550/arXiv.2403.07714> (2025).
9. Du, Y., Wei, F. & Zhang, H. AnyTool: Self-Reflective, Hierarchical Agents for Large-Scale API Calls. Preprint at <https://doi.org/10.48550/arXiv.2402.04253> (2024).
10. Zhang, Q. et al. Large-Language-Model-Based AI Agent for Organic Semiconductor Device Research. *Advanced Materials* 36, 2405163 (2024).
11. Kang, Y. et al. Harnessing Large Language Models to Collect and Analyze Metal–Organic Framework Property Data Set. *J. Am. Chem. Soc.* 147, 3943–3958 (2025).
12. Liu, J. et al. VASPilot: MCP-Facilitated Multi-Agent Intelligence for Autonomous VASP Simulations. Preprint at <https://doi.org/10.48550/arXiv.2508.07035> (2025).
13. Fehlis, Y. et al. Accelerating Drug Discovery Through Agentic AI: A Multi-Agent Approach to Laboratory Automation in the DMTA Cycle. Preprint at <https://doi.org/10.48550/arXiv.2507.09023> (2025).



연구	대표 그림	특징
<b>AI co-scientist</b> <a href="#">Gottweis et al., 2025</a>	<p>The AI co-scientist system design</p>	<p><b>설명</b></p> <ul style="list-style-type: none"> <li>자연어 기반 연구 가설/제안 생성 <b>다중 에이전트 시스템</b></li> <li><b>전문 에이전트:</b> Generation, Reflection, Ranking, Evolution, Proximity, Meta-review</li> <li><b>자연어 상호작용:</b> 목표, 제약, 피드백</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li><b>자가 개선:</b> 가설 생성/토론/진화 토너먼트 프레임워크</li> <li><b>실험 검증:</b> 신규 약물 재배치 후보 및 신규 후성유전학적 표적의 실험적 검증을 통해 유효성 확인</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li>AI co-scientist의 <b>독창성 불분명:</b> 평가가 기존 논문 재현에 국한됨</li> <li><b>시스템 일반화 가능성 불분명</b></li> </ul>
<b>The Virtual Lab</b> <a href="#">Swanson et al., 2025</a>	<p>The Virtual Lab system design</p>	<p><b>설명</b></p> <ul style="list-style-type: none"> <li>과학 연구용 LLM 에이전트 시스템</li> <li>인간 연구자가 <b>PI(수석 연구원)</b> 및 <b>과학 비평가 에이전트</b> 정의</li> <li>PI가 주제 기반 전문 과학자 에이전트 자동 생성, "팀 회의" 및 "개별 회의"로 연구 진행</li> <li>PI가 토론 안내, 요점 종합, 결정, 질문, 요약</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>새로운 SARS-CoV-2 나노바디 설계</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li><b>오래된 문헌/코드 문제 가능성:</b> RAG 또는 미세 조정을 통해 정보 업데이트 가능</li> <li>샌드박스 환경에서 도구 설치 탐색 필요</li> <li><b>시스템 일반화 가능성 불분명</b></li> </ul>
<b>General LLMs for Clinical Review</b> <a href="#">Luo et al., 2025</a>	<p>General LLMs for Clinical Review system design</p>	<p><b>설명</b></p> <ul style="list-style-type: none"> <li>의료 텍스트 이해/생성, 방대한 문헌 동화, 구조화된 임상 리뷰 생성</li> <li>일부 LLM 플랫폼은 정확한 참조 제공</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>AI 생성 리뷰를 체계적으로 비교</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li><b>품질 비교 기준:</b> 생성 시리뷰와 인간 리뷰간 품질 차이를 다루지 않음</li> <li><b>전문가 다양성:</b> 전문가 패널을 중국인과 영국인으로만 구성</li> </ul>

<p><b>ResearchAgent</b> Baek et al., 2024</p>	<p><b>(A) Scientific Knowledge Sources</b></p> <p><b>Paper:</b> Language Models are Few-Shot Learners (...) Here we show that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching (...). Specifically, we train GPT-3, (...)</p> <p><b>Entity-Centric Knowledge Store</b></p> <table border="1"> <thead> <tr> <th>Entity A</th> <th>Entity B</th> <th>Occurrence</th> </tr> </thead> <tbody> <tr> <td>GPT-3</td> <td>Physics</td> <td>78</td> </tr> <tr> <td>CoT</td> <td>CoT</td> <td>17,326</td> </tr> </tbody> </table> <p><b>Academic Graph</b></p> <p><b>(B) Systematic Approach for Research Idea Generation</b></p> <p><b>Paper:</b> GPT-3</p> <p><b>Academic Graph:</b> RLHF, Physics</p> <p><b>Knowledge Store:</b> CoT</p> <p><b>Research Ideas:</b> Problem Identification → Method Development → Experiment Design</p> <p><b>Reviews &amp; Feedback:</b> Reviewing Agents → Human-induced Criteria → Human Judgements</p>	Entity A	Entity B	Occurrence	GPT-3	Physics	78	CoT	CoT	17,326	<p><b>설명</b></p> <ul style="list-style-type: none"> <li>LLM 기반 연구 아이디어 생성 및 개선: 문제 정의, 방법론, 실험 설계</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li><b>entity-centric knowledge augmentation:</b> LLM이 연구 아이디어 생성시 외부 지식 활용 방법. <ul style="list-style-type: none"> <li>학제간 통찰 등 기존 문헌으로는 어려울 수 있는 참신한 아이디어 획득 가능.</li> </ul> </li> <li><b>Academic Graph 활용:</b> 연관 과학 문헌에 대한 이해 확장</li> <li><b>Knowledge Store 활용:</b> 개념 및 학제간 관계에 대한 백과사전식 조망</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li><b>제한적 Knowledge Store:</b> 다른 많은 엔티티 및 상호 연결성을 포함하지 못함</li> <li><b>환각 가능성:</b> LLM 중심으로 환각 가능성 존재. 실험적 검증 필수</li> <li><b>실증 미흡:</b> ResearchAgent가 제시한 가설의 품질을 인간 전문가와 모델이 평가했으나, 실험 등 실증으로 이어지지 못함.</li> </ul>
Entity A	Entity B	Occurrence									
GPT-3	Physics	78									
CoT	CoT	17,326									
<p><b>Sakana AI</b> Lu et al., 2024</p>	<p><b>Idea Generation</b></p> <p>LLM Idea/Plan Innovation → Novelty Check Sem. Scholar → Idea scoring / archiving</p> <p><b>Experiment Iteration</b></p> <p>Experiment Template → Code Δ via LLM &amp; aider → Experiment Exec Script → Numerical Data/Plots → Update Plan → Experiments</p> <p><b>Paper Write-Up</b></p> <p>Manuscript Template → Text Δ via LLM &amp; aider → Manuscript → LLM Paper Reviewing</p>	<p><b>설명</b></p> <ul style="list-style-type: none"> <li><b>완전 자동 개방형 과학 발견 시스템:</b> 아이디어 생성, 실험 반복, 논문 작성 <ul style="list-style-type: none"> <li>아이디어 생성: LLM으로 연구 방향 브레인스토밍, 자기 성찰, 유사 문헌 필터링</li> <li>실험 반복: Aider(오픈소스 코딩 도구)로 실험 실행/결과 시각화/오류 수정, 메모 작성</li> <li>논문 작성: 템플릿 활용 LaTeX 작성</li> <li>LLM 생성 리뷰: 논문 품질, 참신함, 흥미도 등</li> </ul> </li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>ML연구 프로세스 전체 자동화: AI 스레드 통합</li> <li>LLM 리뷰어가 논문 판단</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li>소규모 실험 집중: 계산 효율성/제약</li> <li>미래 토큰 정보 유출(Future Token Information Leakage): 모델 성능 부풀림 문제</li> <li>텍스트 전용, ML 연구로 제한</li> </ul>									
<p><b>SciAgents</b> Ghafarollahi et al., 2024</p>	<p><b>a</b></p> <p><b>b</b></p> <p><b>Global knowledge graph</b></p> <p><b>Ontologist:</b> Filters the list of nodes and relationships, defines each term, and discusses the connections between them.</p> <p><b>Scientist 1:</b> Based on the analysis of proposals and relationships identified by the Ontologist, crafts a detailed research proposal encompassing seven key components: hypothesis, outcome, mechanisms, design principles, anticipated properties, comparison, and novelty.</p> <p><b>Critic:</b> Provides a summary of the research idea, conducts a thorough critical scientific review highlighting strengths and weaknesses, and suggests improvements.</p> <p><b>c</b></p>	<p><b>설명</b></p> <ul style="list-style-type: none"> <li><b>자동 과학 발견:</b> LLM/온톨로지 지식 그래프 활용</li> <li><b>two-track 전략:</b> ① 사전 프로그래밍 AI-AI 상호 작용, ② 완전 자동 에이전트 자체 조직화</li> <li><b>전문 역할 부여:</b> Ontologist, Scientists, Critic, Assistant, 그룹 채팅 관리자 <ul style="list-style-type: none"> <li>Assistant Agent: 외부 도구 접근 가능 (Semantic Scholar API)</li> </ul> </li> <li><b>상세 문서 출력:</b> PDF, CSV 형식</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>새로운 과학 개념 탐색 및 혁신적 아이디어 생산</li> <li>협력적 프레임워크: 혁신적이면서 잘 짜여진 과학 가설 생성</li> <li>유사 아이디어 제거: 연구 아이디어 참신성 평가</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li><b>LLM의 내재적 한계:</b> 환각 등 부정합한 응답을 생성하며 유해한 콘텐츠를 생성할 수 있음</li> <li><b>특정 분야 집중:</b> biologically inspired materials 분야에서만 활용됨. 타 분야에 적용될 잠재력이 있으나 본 연구에서는 확인되지 않음</li> <li><b>실증 미흡:</b> 문헌 검색 및 기존 과학 데이터의 분석, 합성을 활용했으며 실제 실험은 아님.</li> </ul>									

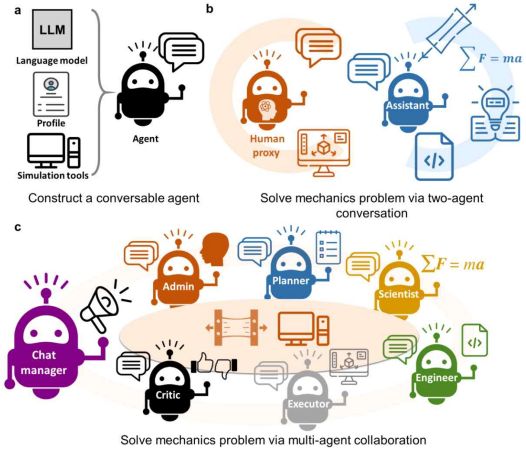
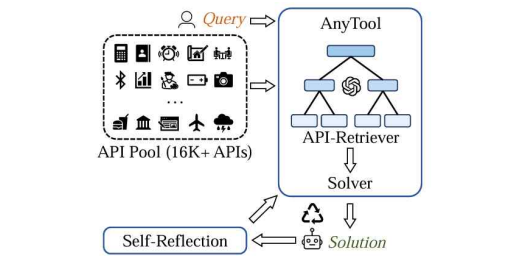
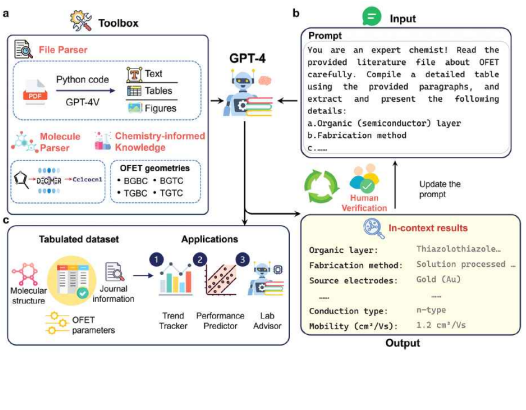
<p><b>MARG</b> (Multi-Agent Generation) <a href="#">D'Arcy et al., 2024</a></p>	<p><b>Review</b></p>  <p>Figure 1: Overview of our multi-agent architecture.</p> 	<p><b>설명</b> · LLM활용 논문 리뷰 생성</p> <p><b>독창성</b> · Multi-agent 구조: 논문 전체를 처리 가능하게 하는 입력 분산/병렬 처리 구조 · Specialized Agents (MARG-S): 실험, 명확성, 영향력(임팩트) 평가에 특화된 전문가 에이전트를 추가하여 구체적, 맞춤형 피드백 생성 · Refinement Stage: 1차 생성 코멘트를 별도의 에이전트 그룹이 재검토·수정하여 구체성, 정확성, 유효성을 향상</p> <p><b>단점</b> · 높은 비용: 기존 방식 대비 토큰 사용량이 많아 API 비용 및 실행 시간이 증가 · 내부 통신 오류: 에이전트 간 메시지 전달 시 일부 정보 누락·오해 발생 · 비주요 정보 부재: 도표·그림·수식 배열 정보 미활용으로 특정 유형의 명확성 문제 포착 못함.</p>
<p><b>PaperQA2</b> <a href="#">Skarlinski et al., 2024</a></p>		<p><b>설명</b> · 과학 정보 처리 에이전트: 인간 전문가를 능가하는 과학 문헌 검색, 요약, 모순 탐지 등</p> <p><b>독창성</b> · 실제 작업 기반 평가: 현실적인 문헌 검색, 요약, 모순 탐지 작업 중심 시 성능 평가 · 인간과의 직접 비교: 박사급 전문가와 직접 성능 비교 · 에이전트 기반 RAG 구성: 단순 검색-생성 방식이 아닌 다단계 도구 기반 접근으로 정확도 향상 · 모순 탐지 자동화(ContraCrow): 인간의 한계를 넘는 대규모 문헌 간 모순 탐지 자동화</p> <p><b>단점</b> · 모델 비용: 검색 및 요약이 많아 실행 비용이 높음 · 데이터셋 편중: 주로 생명과학 분야 논문을 기반으로 실험되어 타 분야 일반화 미흡</p>
<p><b>AgentRxiv</b> <a href="#">Schmidgall et al., 2025</a></p>		<p><b>설명</b> · 자율 에이전트 기반 협업 연구 환경 구축: 연구 계획 수립부터 논문 작성까지 전체 연구 가속화</p> <p><b>독창성</b> · end-to-end 과학 연구 자동화: 계획·실험·분석·작성까지 전 과정 자동화 · 멀티도메인 적용성: 다양한 과학 분야에서의 프로토타입 설정</p> <p><b>단점</b> · LLM 의존성: LLM의 지식 정확성과 추론 한계가 전체 성능에 영향을 줌 · 분야별 도구 및 DB 연동 성능 편차: - 잘 되는 분야: 화학, 소재, 생명과학, 제약, 문헌, 약물탐색 - 안 되는 분야: 사회과학, 인문학, 특허, 산업 표준 문서, 실험 장비 제어 및 수집</p>



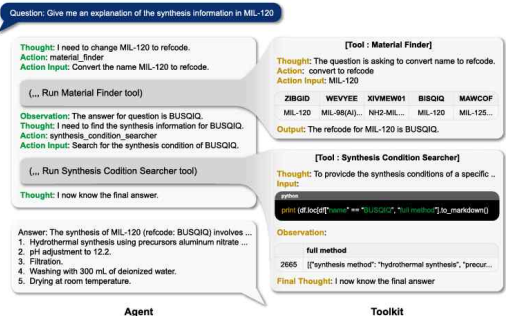
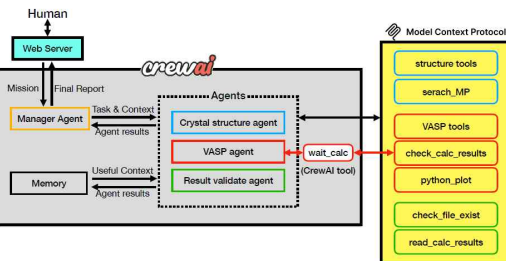
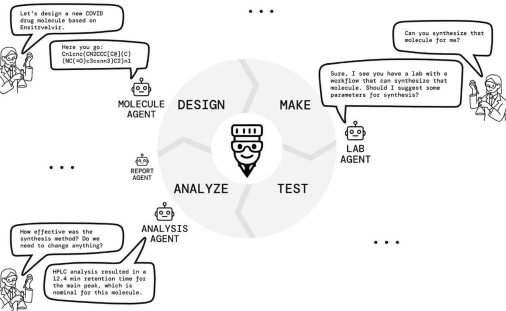
<p><b>MOFGen</b> Inizan et al., 2025</p>		<p><b>설명</b></p> <ul style="list-style-type: none"> <li>금속-유기 골격체(Metal-Organic Framework) 탐색 및 합성: 합성 가능성 안정성 동시 고려</li> <li>- 대규모 화학 공간에서 새로운 구조를 신속히 탐색하고 실제 탐색까지 연결</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>Agentic AI통합: MOF 설계·검증·합성을 한 시스템 내에서 처리</li> <li>대규모 신규 링커 생성: 80,000+ 합성 가능 링커 발굴</li> <li>실험 검증 성과: Diffusion 모델 생성물 중 MOF 5종을 실제 합성 및 결정 구조 확인</li> <li>화학 공간 확장: 기존 실험 데이터에 없는 조성과 구조 발굴</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li>합성 가능성 예측 불확실: ML 예측이 다소 낙관적으로 실험 평가와 불일치</li> <li>금속 SBU 다양성 한정: 주로 Zn 기반 MOF에 집중하여 다른 금속 확장 제한적</li> <li>DFT 비용 문제: r2SCAN-D4 등 고준위 계산의 연산 비용이 큼</li> <li>LLM 조성 편향: 주어진 프롬프트에 기반한 조성 재현율 경향</li> </ul>
<p><b>Coscientist</b> Boiko et al., 2023</p>		<p><b>설명</b></p> <ul style="list-style-type: none"> <li>LLM기반 인간 개입을 최소화한 페루프 연구: 화학 지식 탐색·설계·실험 계획·데이터 분석</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>LLM중심 페루프: 아이디어~실험 계획~데이터 분석까지 LLM이 주도</li> <li>도구 확장성: API를 통한 화학 데이터베이스 및 계산 플랫폼 직접 호출</li> <li>인간-에이전트 비교 평가: 실험 계획 품질을 인간 전문가와 비교하여 성능 측정</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li>화학 지식 편향: 학습 데이터 지식 불균형</li> <li>실험 안정성: 제안된 일부 반응 조건이 안정성 검증 없이 출력될 가능성</li> <li>계산 비용: LLM과 외부 도구 간 연동시 처리 지연</li> </ul>
<p><b>ChemCrow</b> Bran et al., 2024</p>		<p><b>설명</b></p> <ul style="list-style-type: none"> <li>LLM 화학 분야 활용성 극대화: 전문 도구(계산화학, DB, 시각화)와의 체계적 결합 구현</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>LLM 도구 생태계 정의: 화학 관련 외부 도구를 기능적, 체계적으로 분류</li> <li>프롬프트-API 변환 패턴: 자연어 명령을 API 호출로 변환하는 표준 절차 제안</li> <li>정량적 성능 비교: LLM 단독 vs LLM+도구 조합 성능 <ul style="list-style-type: none"> <li>정확도(accuracy): 50~70% → 85~95%</li> <li>시간 효율(time efficiency): LLM+도구 조합시 평균 처리 시간이 30~50% 단축</li> <li>에러율(Error Rate): 화학 구조 표현 오류, 반응식 불균형 등 20% → 5% 이하</li> </ul> </li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li>도구 의존성: 외부 API접근성, 라이선스, 속도 제한에 따른 제약</li> <li>지식 일관성: LLM과 도구 간 결과 불일치 발생 가능</li> <li>프롬프트 민감성: 동일 요청이라도 문구에 따라 호출 결과 품질 차이</li> </ul>

<div>SciMON</div> <div>Wang et al., 2024</div>	<div><div>Background Context</div><div>Problems, motivations, focus points...</div></div> <div><div>Inspiration Retrieval</div><div></div></div> <div><div>Idea Generation</div><div>"Given [context], a [new idea], Δ vs. prior work..."</div></div> <div><div>Prior Literature</div><div>{{(Background_i, idea_i)}</div></div> <div><div>Iterative Novelty Boosting</div><div></div></div> <tr><td><div>MatPilot</div><div>Ni et al., 2024</div></td><td><div><div>(a) Dispensing</div><div>(b) Ball Milling</div><div>(c) Sintering</div><div>(d) Molding</div><div>(e) DMS</div><div>(f) DHM</div></div></td><td><div><div>설명</div><div>· LLM 기반 AI재료과학자: 재료 탐구, 설계, 실험 수행 인간-기계 협업 지능형 과학 연구 플랫폼</div></div><div><div>독창성</div><div>· 인간-기계 협업 중심 설계: AI가 제안하고 인간이 검증·조율</div><div>· 지속 업데이트 지식베이스: 최신 연구 반영 가능 RAG 구조</div><div>· 다중 에이전트 혁신 프레임워크: 탐색·평가·통합 단계별 역할 분담</div><div>· 재료 실험 전 주기 자동화: 수작업 대기 일관성과 재현성 확보</div></div><div><div>단점</div><div>· 임베디드 인텔리전트 미완성: 전면 적용까지 1~2년 소요 예상</div><div>· 분야 편향 가능성: 특정 재료 분야에 특화된 데이터셋 한정</div><div>· 복잡한 인과관계 해석 한계: 상관관계 중심의 기존 AI 한계 잔존</div></div></td></tr> <tr><td><div>Agent Laboratory</div><div>Schmidgall et al., 2025</div></td><td><div><div>Phases</div><div>Literature ReviewExperimentationReport Writing</div><div>Subtasks</div><div>Literature ReviewPlan ExperimentationRun ExperimentsReport Writing</div><div>Human</div><div>INSTRUCTORPHD StudentPHD StudentML EngineerPHD StudentPH</div></div></td></tr>	<div>MatPilot</div> <div>Ni et al., 2024</div>	<div><div>(a) Dispensing</div><div>(b) Ball Milling</div><div>(c) Sintering</div><div>(d) Molding</div><div>(e) DMS</div><div>(f) DHM</div></div>	<div><div>설명</div><div>· LLM 기반 AI재료과학자: 재료 탐구, 설계, 실험 수행 인간-기계 협업 지능형 과학 연구 플랫폼</div></div> <div><div>독창성</div><div>· 인간-기계 협업 중심 설계: AI가 제안하고 인간이 검증·조율</div><div>· 지속 업데이트 지식베이스: 최신 연구 반영 가능 RAG 구조</div><div>· 다중 에이전트 혁신 프레임워크: 탐색·평가·통합 단계별 역할 분담</div><div>· 재료 실험 전 주기 자동화: 수작업 대기 일관성과 재현성 확보</div></div> <div><div>단점</div><div>· 임베디드 인텔리전트 미완성: 전면 적용까지 1~2년 소요 예상</div><div>· 분야 편향 가능성: 특정 재료 분야에 특화된 데이터셋 한정</div><div>· 복잡한 인과관계 해석 한계: 상관관계 중심의 기존 AI 한계 잔존</div></div>	<div>Agent Laboratory</div> <div>Schmidgall et al., 2025</div>	<div><div>Phases</div><div>Literature ReviewExperimentationReport Writing</div><div>Subtasks</div><div>Literature ReviewPlan ExperimentationRun ExperimentsReport Writing</div><div>Human</div><div>INSTRUCTORPHD StudentPHD StudentML EngineerPHD StudentPH</div></div>
<div>MatPilot</div> <div>Ni et al., 2024</div>	<div><div>(a) Dispensing</div><div>(b) Ball Milling</div><div>(c) Sintering</div><div>(d) Molding</div><div>(e) DMS</div><div>(f) DHM</div></div>	<div><div>설명</div><div>· LLM 기반 AI재료과학자: 재료 탐구, 설계, 실험 수행 인간-기계 협업 지능형 과학 연구 플랫폼</div></div> <div><div>독창성</div><div>· 인간-기계 협업 중심 설계: AI가 제안하고 인간이 검증·조율</div><div>· 지속 업데이트 지식베이스: 최신 연구 반영 가능 RAG 구조</div><div>· 다중 에이전트 혁신 프레임워크: 탐색·평가·통합 단계별 역할 분담</div><div>· 재료 실험 전 주기 자동화: 수작업 대기 일관성과 재현성 확보</div></div> <div><div>단점</div><div>· 임베디드 인텔리전트 미완성: 전면 적용까지 1~2년 소요 예상</div><div>· 분야 편향 가능성: 특정 재료 분야에 특화된 데이터셋 한정</div><div>· 복잡한 인과관계 해석 한계: 상관관계 중심의 기존 AI 한계 잔존</div></div>				
<div>Agent Laboratory</div> <div>Schmidgall et al., 2025</div>	<div><div>Phases</div><div>Literature ReviewExperimentationReport Writing</div><div>Subtasks</div><div>Literature ReviewPlan ExperimentationRun ExperimentsReport Writing</div><div>Human</div><div>INSTRUCTORPHD StudentPHD StudentML EngineerPHD StudentPH</div></div>					

<p><b>OpenFOAMGPT</b>  <a href="#">Feng et al., 2025</a></p>		<p><b>설명</b> · OpenFOAMGPT: LLM 기반 다중 에이전트 시스템을 활용한 CFD 연구 자동화 시스템</p> <p><b>독창성</b> · <b>Multi-agent End-to-End CFD 자동화</b>: 전처리부터 후처리까지 무인화  · <b>Prompt Pool 기법</b>: 케이스별 완전 지시문 생성으로 구성 파일 수정 방식의 비효율 제거  · <b>에러 구동 반복 개선</b>: 에러 로그를 통한 설정 파일 자동 수정 및 재시도  · <b>고신뢰성 및 고재현성</b>: 450건 이상의 시뮬레이션에서 100% 성공률 달성</p> <p><b>단점</b> · <b>대상 제한</b>: OpenFOAM 환경에 최적화되어 다른 CFD 플랫폼에는 적용이 어려움  · <b>실시간·대규모 환경 실증 미흡</b>: 실제 산업 환경에서의 실시간·대규모 시뮬레이션 부하 및 하드웨어 제약 조건에 대한 검증 부족</p>
<p><b>An Automatic End-to-End Chemical Synthesis Development Platform</b>  <b>Powered by Large Language Models</b>  <a href="#">Feng et al., 2025</a></p>		<p><b>설명</b> · 합성 경로 설계, 실험 실행, 결과 분석까지 통합적으로 수행하는 LLM 기반의 자동화된 종단간(end-to-end) 화학 합성 개발 플랫폼</p> <p><b>독창성</b> · <b>합성 경로 계획</b>: LLM이 화학 반응 DB와 문헌을 기반으로 최적 합성 경로 설계  · <b>로봇 실험 제어</b>: 로봇 팔 자동화 장비와 연동, 반응물 투입, 반응 조건 제어, 샘플링 수행  · <b>실험 데이터 분석</b>: 실시간 센서 데이터와 분석 장비 결과를 기반으로 반응 진행 상황 평가  · <b>반응 최적화 루프</b>: 결과를 바탕으로 LLM이 조건을 재설정하여 최적 수율 및 선택성 달성  · <b>다중 모듈 복합</b>: 계획-실행-분석 모듈간 실시간 통신 및 데이터 교환</p> <p><b>단점</b> · <b>범용성 제한</b>: 특정 화학 반응 유형에 최적화  · <b>재현성 문제</b>: 실험 장비, 환경 차이 반영 알 수 없음</p>
<p><b>OWL: Optimized Workforce Learning for General Multi-Agent Assistance in Real-World Task Automation</b>  <a href="#">Feng et al., 2025</a></p>		<p><b>설명</b> · <b>다중 협력 학습 프레임워크 (OWL)</b>: 복잡한 다단계 작업을 효율적이고 유연하게 처리할 수 있는 범용 다중 에이전트 지원 시스템</p> <p><b>독창성</b> · <b>OWL 프레임워크 제안</b>: 다중 에이전트 학습에서 역할 최적화, 지식 공유, 자원 관리 통합  · <b>시뮬레이션 + 실세계 평가</b>: 가상 환경 뿐 아니라 실제 로봇 실험으로 효과 검증  · <b>범용성 확보</b>: 산업별 맞춤형 적용이 가능한 구조  · <b>적응형 협력</b>: 작업 중 발생하는 예외 상황에 즉각 대응 가능</p> <p><b>단점</b> · <b>실증 규모 한계</b>: 검증 실험의 규모가 제한적이며 특정 산업 시나리오에 최적화된 부분 존재  - 고도로 동적인 환경에서의 장기적 성능 안정성 추가 검증 필요</p>

<p><b>MechAgents</b>  <a href="#">Ni et al., 2023</a></p>		<p><b>설명</b> • MechAgents: LLM 기반 역학 분야의 문제 해결, 신규 데이터 생성, 지식 통합 프레임워크</p> <p><b>독창성</b> • 역학 전용 다중 에이전트 협력 구조          • 데이터 생성+문제 해결 통합: 해답 도출을 넘어 데이터 생성·보강          • 다중 데이터 소스 결합: 실험·계산·이론의 상호 보완적 활용          • 타 분야 확장성: 복잡한 다단계 작업을 효율적이고 유연하게 처리할 수 있는</p> <p><b>단점</b> • 데이터 품질이 생성 데이터의 정확성에 의존          • 실제 산업 환경의 대규모, 고정밀 문제에 대한 성능 검증 부족</p>
<p><b>AnyTool</b>  <a href="#">Du et al., 2024</a></p>		<p><b>설명</b> • AnyTool: 16,000개 이상의 실세계 API를 사용하는 GPT-4 기반 에이전트</p> <p><b>독창성</b> • Plug-and-Play 구조: 외부 모듈 학습 없이 GPT-4의 함수 호출 기능만으로 구현          • 계층적 API검색 구조: LLM 최대 컨텍스트 길이 한계 극복          • 자기반성 메커니즘: 실패 원인 후 반복적 재시도를 통한 성능 향상          • 평가 프로토콜 개선: 기존의 비현실적 성능 과장 문제 해결</p> <p><b>단점</b> • 복잡 시나리오 검증 부족: 매우 복잡한 환경에서의 성능 검증 데이터 부재          • GPT-4 의존성: GPT-4 성능 한계가 AnyTool의 결과에 직접적인 영향을 미침</p>
<p><b>Organic Semiconductor Device Research Agent</b>  <a href="#">Zhang et al., 2024</a></p>		<p><b>설명</b> • OFET 연구용 Agent: GPT-4 기반 LLM과 ML 알고리즘을 결합한 AI 에이전트</p> <p><b>독창성</b> • 멀티모달 LLM활용: 이미지 포함 비정형 데이터까지 파라미터 추출          • 고정밀 성능 해석: SHAP 기반, 분자·공정 요인의 정량적 영향 분석          • 실험 검증: 제안된 최적화 방안(DP-DTT OFET)으로 mobility 향상 (0.42→1.10 cm<sup>2</sup>/Vs)</p> <p><b>단점</b> • 특수 상황 검증 부족: 보고된 최적화 전략이 모든 물질과 공정에 일반화 가능하다는 보장 없음          • GPT-4 성능 의존성: 모델의 환각 가능성과 이미지 OCR 정확도 한계</p>



<b>MOF-LLM</b> <a href="#">Kang et al., 2025</a>	 <p>The diagram illustrates the MOF-LLM workflow. It shows an <b>Agent</b> (left) and a <b>Toolkit</b> (right). The Agent receives a question: "Give me an explanation of the synthesis information in MIL-120." The Agent's thought process involves identifying the need to change MIL-120 to refcode, using a material finder tool, and then using a synthesis condition searcher tool. The Toolkit provides the necessary tools and information, including a list of materials (ZIF40, WUYEE, XWWEW01, BISOQ, MAWQF) and a specific synthesis method for MIL-120 (hydrothermal synthesis using precursors aluminum nitrate).</p>	<p><b>설명</b></p> <ul style="list-style-type: none"> <li>MOF-LLM: LLM을 활용해 금속-유기 골격체(MOF) 문헌으로부터 합성·물성 데이터를 대규모로 자동 추출, 정리하고 새로운 MOF 재료 설계, 물성 예측 연구 가속화</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>멀티모달 데이터 처리: 텍스트, 표, 이미지 모두에서 데이터 추출</li> <li>자동+수동 혼합 검증: LLM 자동화 장점과 전문가 검증 결합</li> <li>대규모 MOF 데이터셋 구축: 1만건 이상의 문헌에서 데이터 수집</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li>데이터 품질 의존성: 원문 보고 형식 불일치에 따른 누락 및 오류 가능성</li> <li>LLM 성능 의존성: GPT-4의 특정 화학 전문용어 처리 능력 제한</li> </ul>
<b>VASPIlot</b> <a href="#">Liu et al., 2025</a>	 <p>The diagram shows the VASPIlot architecture. A <b>Human</b> interacts with a <b>Web Server</b>, which connects to a <b>Manager Agent</b>. The Manager Agent interacts with a <b>Crew</b> of Agents (Crystal structure agent, VASP agent, Result validate agent) and a <b>Model Context Protocol (MCP)</b>. The MCP contains tools like structure tools, serach_MP, VASP tools, python_plot, check_file_exist, and read_calc_results. The Crew agents interact with the MCP tools to perform tasks like wait_calc, check_calc_results, and read_calc_results.</p>	<p><b>설명</b></p> <ul style="list-style-type: none"> <li>VASPIlot: CrewAI 기반 멀티에이전트 플랫폼과 MCP를 활용해 VASP DFT 워크플로우 자동화</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>VASP 특화 멀티에이전트 시스템: 기존 멀티에이전트 프레임워크(MAF)와 차별화</li> <li>MCP 활용: 다양한 도구와의 호환성을 확보</li> <li>Automatic Error Handling: 계산 중 발생하는 VASP 오류를 자동으로 파악, 파라미터 수정</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li>VASP 종속적 구조: 타 DFT 확장시 추가 작업 필요</li> <li>LLM 선택 제한: 현재 DeepSeek-V3-0324 모델에 의존. 타 LLM 사용시 호환성 검증 필요</li> <li>사람 검토 필요: 중간 결과 및 파라미터 선택 등 최종 선택은 인간의 몫</li> </ul>
<b>Tippy</b> <a href="#">Fehlis et al., 2025</a>	 <p>The diagram illustrates the Tippy workflow, which is a cycle of DESIGN, MAKE, ANALYZE, and TEST. A <b>MOLECULE AGENT</b> designs a molecule, a <b>LAB AGENT</b> makes it, and an <b>ANALYSIS AGENT</b> analyzes it. The cycle repeats until the molecule is synthesized successfully. The diagram shows a conversation between the agents, with the MOLECULE AGENT asking for a new COVID drug molecule, the LAB AGENT suggesting a workflow, and the ANALYSIS AGENT reporting the results of the synthesis.</p>	<p><b>설명</b></p> <ul style="list-style-type: none"> <li>Tippy: 제약 산업의 핵심 프로세스 DMTA(Design-Make-Test-Analyze) 사이클 자동화</li> </ul> <p><b>독창성</b></p> <ul style="list-style-type: none"> <li>DMTA 전 주기 에이전트 구현: 각 단계를 독립적 AI가 처리</li> <li>Safety Guardrail Agent 도입: 합성 중 위험 반응, 불법 물질, 보안 침해 사전 탐지</li> <li>실제 화합물 적용: Ensitrelvir 기반 COVID 약물 후보 설계 및 분석</li> </ul> <p><b>단점</b></p> <ul style="list-style-type: none"> <li>재현성 이슈: 보안상 이유로 합성 파라미터 비공개.</li> <li>HPLC 외 분석 방법 범용성 미검증: 타 분석기기 호환성 명시되지 않음</li> <li>실험 플로우 고정: 동적 재설계 기능 미포함</li> <li>초기 단계 적용 사례 중심: 실제 신약개발 파이프라인 전체에 대한 적용은 아직 제한적</li> </ul>