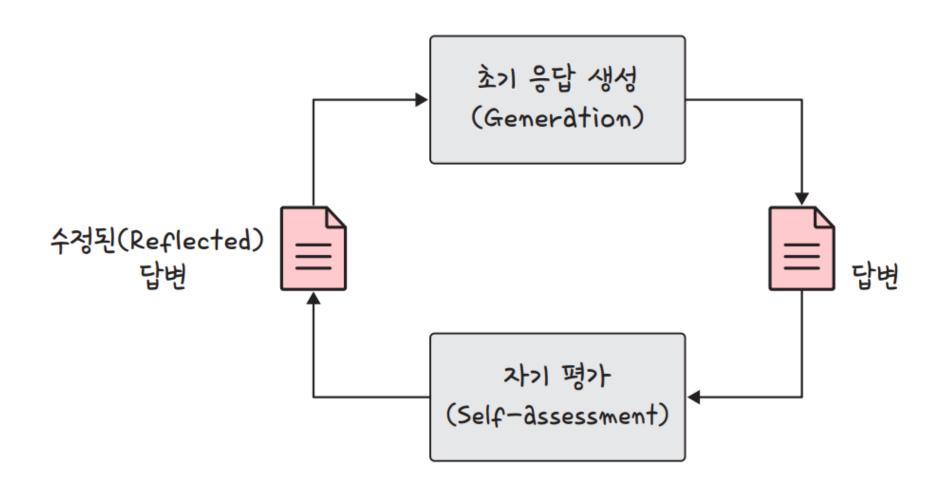
# Al agent Pattern Design

Jae-ik Shin

- Building effective agents
- <a href="https://www.anthropic.com/engineering/building-effective-agents">https://www.anthropic.com/engineering/building-effective-agents</a>

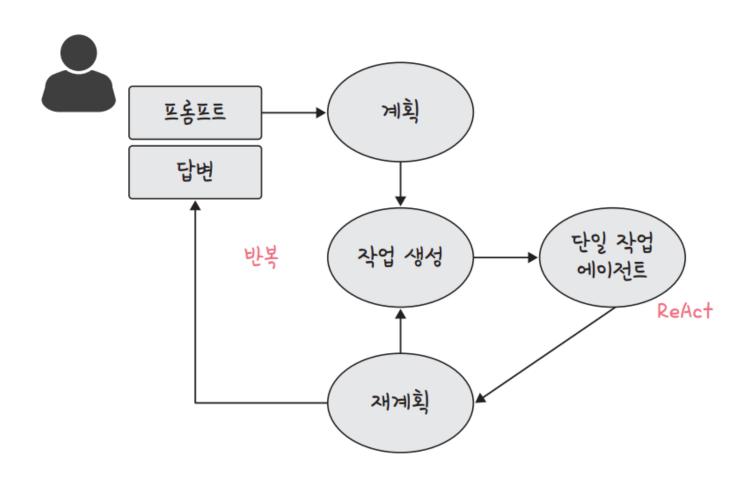
- OpenAI의 에이전트 구축을 위한 실용 가이드
- https://news.hada.io/topic?id=20570

## Reflection Pattern 반응 패턴



https://yozm.wishket.com/magazine/detail/3078/

# Planning Pattern 계획패턴



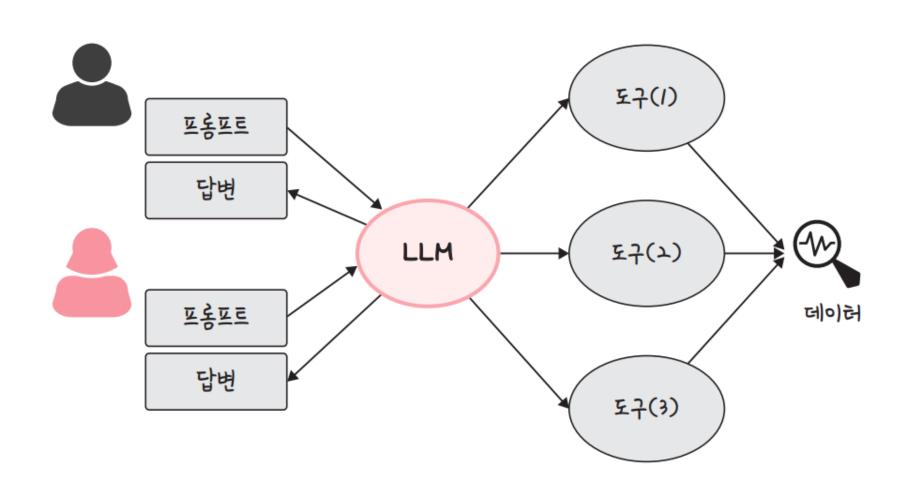
## 계획과실행을 분리된 모델을 사용

- <a href="https://github.com/magnitudedev/magnitude">https://github.com/magnitudedev/magnitude</a>
- 웹앱을 위한 AI 기반 오픈소스 E2E 테스팅 프레임워크
- **자연어**로 쉽게 테스트 케이스를 작성하고, <u>추론 AI 에이전트가 테스트 계획을 세우고, 비주얼 AI 에이전트가 UI 변화를 인식하고 적응하여 테스트를 실행</u>
- 멀티모달 LLM과 비전 모델을 조합해 빠르고 정확한 UI 테스트를 수행하고, 계획은 저장되어 항상 동일한 방식으로 실행 가능
- 문제가 발생하면 추론 에이전트가 자동으로 문제를 해결함
- CI/CD 통합이 가능하며, Playwright와 유사한 방식으로 테스트 실행을 지원. CI 환경(GitHub Actions 등)에서 테스트 실행 가능
- 왜 OpenAl Operator나 Claude Computer Use를 사용하지 않는가?
  - Magnitude는 **계획과 실행을 분리**하여 설계함
  - 일반적인 데스크탑/웹 작업용 API보다 속도, 신뢰성, 비용 면에서 테스트 실행에 최적화되어 있음
  - 별도의 자체 런너를 제공하여 신뢰성, 속도, 비용 효율성을 극대화
- LLM 추천
  - 플래너 모델: 일반 멀티모달 LLM 사용 (추천: Gemini 2.5 Pro)
  - 익스큐터 모델: 빠른 비전 LLM인 Moondream 사용

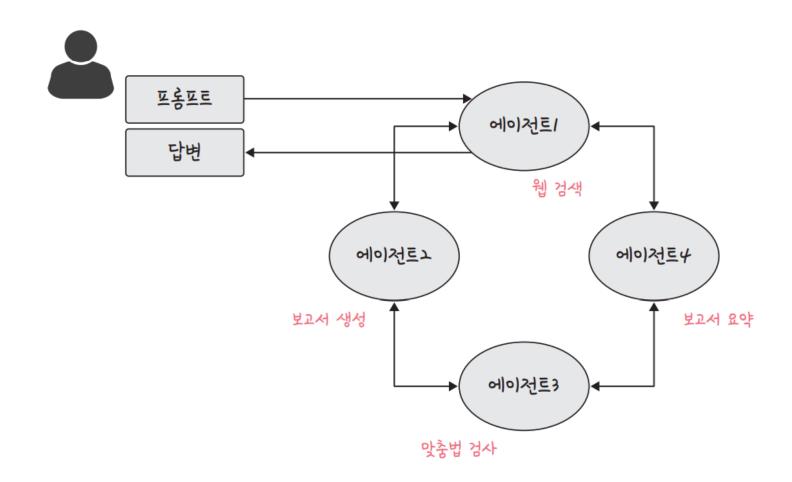
#### Cline Plan-Act mode

- https://wikidocs.net/279368
- Cline에 자연어로 개발 작업을 지시하면 먼저 프로젝트 파일 구조와 코드 AST를 분석하고 정규식 검색을 통해 관련 파일을 읽어들이며, 대용량 프로젝트도 중요한 맥락만 선별하여 파악합니다. 그런 다음 계획(Plan) 단계를 거쳐 작업 계획을 수립하고, 실행(Act) 단계에서 이를 수행하는 이중 모드로 동작합니다. Cline은 매 단계마다 사용자의 승인을 받아 휴먼-인-더-루프 형태로 안전하게 동작하며, 필요한 경우 모델 컨텍스트 프로토콜(MCP)을 통해 자체 기능을 확장할 수도 있습니다
- https://docs.cline.bot/getting-started/model-selection-guide
- Planning LLM model + Acting LLM model
- SPARC: Specification, Pseudocode, Architecture, Refinement, and Completion
- https://github.com/ruvnet/sparc

## Tool User Pattern 도구사용 패턴



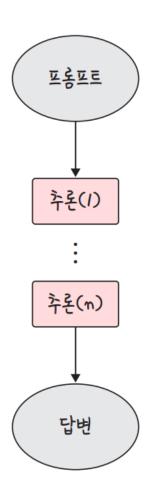
# Multi-Agent Pattern 멀티에이전트 패턴



#### Zed Agent Panel

- https://news.hada.io/topic?id=20775
- https://zed.dev/docs/ai/tools
- 7개의 도구만 사용하는 최소한의 에이전트 프레임워크:
  - 읽기, 쓰기, 차이, 탐색, 명령, 질문, 생각
- https://github.com/aperoc/toolkami

### Chain of Thought Prompting Pattern



#### Qwen3 CoT On/Off

- Qwen3 의 Hybrid thinking mode
- https://devocean.sk.com/blog/techBoardDetail.do?ID=167448&b oardType=techBlog

Token: think / no\_think

Template: enable\_thinking=true / false