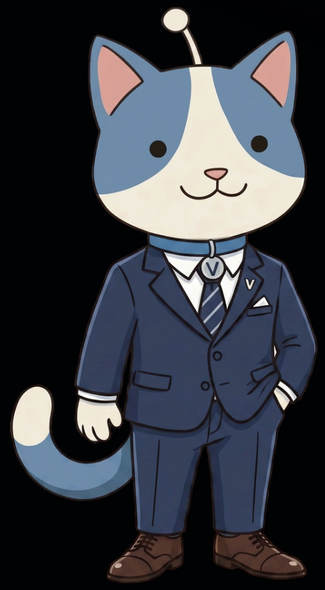
LV6 · ADVANCED

Agents & Skills

여러 단계 작업을 자율 처리하는 AI 에이전트 설계

75분 · 12 슬라이드 · 아키텍트·시니어 개발자 · Anthropic Building Effective Agents 백서



ROADMAP · 75 MINUTES

오늘의 4 블록

"한 번의 응답"이 아닌 "끝까지 일을 끝내는" AI를 설계한다

BLOCK A · 15 min

철학 + 정의

Workflow vs Agent / Anthropic 5 패턴 / 언제 Agent 쓰지 말아야 하나

BLOCK B · 25 min

5 패턴 풀 커버

Prompt Chaining / Routing / Parallelization / Orchestrator / Evaluator-Optimizer

BLOCK C · 25 min

Skills + 4 시나리오

Skills 정의 / Customer Support / Coder / Research / Operations Agent

BLOCK D · 10 min

안전 + 모범

관찰 가능성 · 권한 · 비용 통제 · Lv7 로드맵

핵심 약속: 오늘 끝나면 **"우리 회사 업무 1개에 어떤 Agent 패턴이 맞는지"** 판별할 수 있다.

PRINCIPLE 01 · DEFINITION

W

Workflow vs Agent

Anthropic Definition

Anthropic 정의: **Workflow = 사람이 코드로 흐름 고정. Agent = LLM이 흐름을 결정.**

Workflow

"이메일 받으면 → 분류 → 답장 초안 → 사람 확인" 같은 **고정된 순서**. 예측 가능·디버그 쉬움.

Agent

"이 고객 문제 해결해줘" → LLM이 **어떤 도구를 언제 쓸지 스스로 결정**. 강력·예측 어려움.

⚖ Anthropic 권장

"**Workflow로 가능하면 Workflow**로. Agent는 흐름이 예측 안 될 때만."

비용·복잡도

Agent는 토큰 수배~수십배. 잘못 설계하면 **무한 루프·비용 폭탄**.

PRINCIPLE 02 · 5 PATTERNS

5

5 패턴 요약

Anthropic Building Effective Agents

Anthropic 백서 (2024년 12월): 이 5개만 알면 90% 케이스 커버

① Prompt Chaining

A → B → C 순차. **단계별 검증 가능**. 가장 흔함.

② Routing

입력 분류 → 적절한 처리기로. **비용 효율** (간단한 건 Haiku).

③ Parallelization

여러 LLM 동시 호출. **속도** 또는 **다중 의견 합의**.

④ Orchestrator-Workers

중앙 LLM이 작업 분해 → 워커들에게 위임. **복잡한 작업**용.

⑤ Evaluator-Optimizer

생성 LLM + 평가 LLM 반복. **품질이 중요**한 작업 (번역·코드).

PATTERN 1 & 2 · CHAINING + ROUTING

Chaining + Routing 흐름

가장 흔한 패턴 — 입력을 단계별로 변형. 첫 단계에서 분기 가능.

STEP 1
입력
고객 이메일

→

STEP 2
분류 (Route)
긴급 / 일반 / 스팸

→

STEP 3
처리
분류별 응답 생성

핵심: Chaining은 "단계마다 게이트(Gate) 검증" 가능 — 한 단계 실패 시 즉시 중단·재시도. 디버깅 쉬움.

PATTERN 3 & 4 · PARALLEL + ORCHESTRATOR

Parallel + Orchestrator

속도 또는 다중 관점 — 동시 호출. 또는 마스터 LLM이 분해·위임.

PARALLEL
입력
긴 문서 1개

→

3 LLM 동시
요약 · 키워드 · 위험점
동시 처리 → 합치기

→

RESULT
종합 결과
속도 3배 ↑

Orchestrator: "여행 계획 짜줘" → 마스터 LLM이 [숙소·항공·일정] 3 서브태스크 분배 → 워커 LLM들 처리 → 통합. 복잡한 자동화에 사용.

PATTERN 5 · EVALUATOR-OPTIMIZER

품질을 위한 반복

생성 LLM이 만든 결과를 평가 LLM이 검토 → 피드백 → 재생성. 번역·코드·중요 콘텐츠에 강력.

STEP 1
생성
Sonnet이 초안

→

STEP 2
평가
Sonnet 평가자가 점수·피드백

→

STEP 3
재생성
피드백 반영 → 반복

주의: 반복 횟수 제한 필수 (보통 3회). 무한 루프 방지. 비용 폭탄 1순위.

PRINCIPLE 03 · SKILLS

Skills

Reusable Capabilities

Skills = **Claude에 추가하는 재사용 가능한 능력 단위** — 함수 + 시스템 프롬프트 + 예시 묶음

무엇

"PDF 분석", "회사 정책 답변", "코드 리뷰" 같은 단위 능력을 **한 번 정의 → 무한 재사용**.

어디서

Claude Code, Claude Desktop, API 등 어디서든 동일하게 호출.

MCP와 차이

MCP = 외부 시스템과 연결. Skills = **특정 작업 수행 방법** 캡슐화.

회사 활용

"우리 회사 보고서 작성 Skill", "법무 검토 Skill" 같이 **회사 표준 작업**을 Skill로.

SCENARIO · 4 AGENT PATTERNS

4 실전 Agent

Anthropic 백서가 권장하는 4가지 실전 패턴 — 회사에 즉시 적용 가능.

01 · Customer Support Agent

고객 문의 끝까지 처리

흐름: 분류 → 검색 → 답변 → 검증 → 만족도 측정. 못 풀면 사람에게.

02 · Coder Agent

이슈 → PR 자동 생성

흐름: 이슈 분석 → 영향 파일 찾기 → 코드 수정 → 테스트 → PR. Claude Code 기반.

03 · Research Agent

주제 → 보고서

흐름: 검색 → 1차 요약 → 누락 영역 검색 → 보강 → 최종 보고서. Anthropic Deep Research 패턴.

04 · Operations Agent

운영 자동화

흐름: 모니터링 → 이상 감지 → 진단 → 1차 조치 → 사람 알림. 야간 당직 보조.

공통: 모두 **사람 confirm 게이트** 포함. 자율도 100%는 위험.

PRINCIPLE 04 · OBSERVABILITY

0

관찰 가능성

Observability

Agent는 **예측이 어렵다** — 무엇을 했는지·왜 했는지 추적 필수

Trace 로깅

모든 도구 호출·LLM 응답·결정 이유를 trace_id로 묶어 저장.

비용 추적

trace 단위 토큰·비용 집계. 한도 초과 시 자동 중단.

무한 루프 차단

최대 반복 횟수 + 최대 비용 + 최대 시간 3중 가드.

사람 확인 게이트

중요한 결정 전 사람 confirm 필수. 자동 실행하지 마라.

PRINCIPLE 05 · WHEN NOT

N

Agent를 쓰지 않을 때

Anti-patterns

Anthropic 권장: **"가능하면 Workflow"** — Agent 안 쓰는 게 옳을 때

□ 작업이 단순할 때

"PDF 요약"은 Agent 필요 X. 단일 API 호출이면 끝.

□ 단계가 고정될 때

매번 같은 순서면 Workflow. Agent는 **예측 불가능한 분기**가 있을 때.

□ 실패 비용이 클 때

금융 거래, 법무 발송 등은 사람이 매 단계 검증. Agent 자율 X.

□ 디버깅이 안 될 때

Trace 인프라 없이 Agent 도입 X. 사고 시 원인 추적 불가.

NEXT STEPS

다음은 Cloud

LV7 · CLOUD

클라우드 AI 엔터프라이즈

AWS Bedrock · Vertex · Azure에서 Claude 대규모 운영. 한국 인프라 보강.

LV8 · EDU

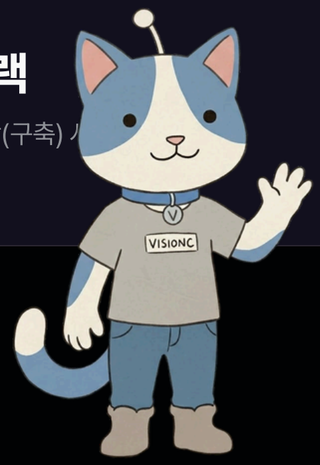
교사·강사를 위한 AI

교육 현장 적용 + 학습 효과 분석.

DEPT·BUILD

기업 도입 트랙

15강(부서별) + 30강(구축) 시



 오늘의 약속: 우리 회사 업무 1개에 **어떤 Agent 패턴이 맞는지** 결정한다.