

Dantelabs

Hermes × Codex 로 세우는 나만의 AI 가상 오피스

- 8주 마스터 과정

Section2. > Unit.1

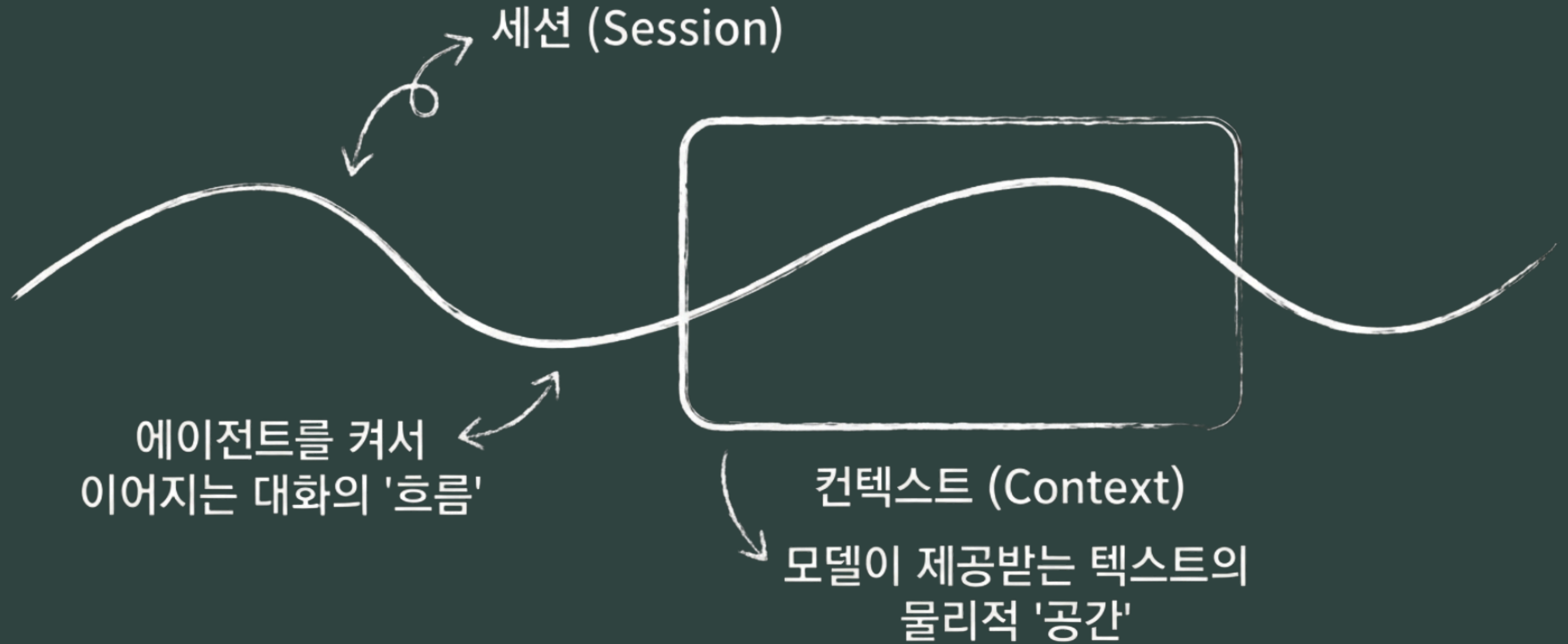


우리의 첫 AI 직원, 어떻게 채용할까?

- 페르소나 (성격)
- 메모리 (기억력)
- 스킬 (업무 능력)

이 모든 파편화된 요소들은 결국 '하나의 틀' 안에서 움직입니다.
그 틀의 원리를 이해하는 것이 채용의 첫걸음입니다.





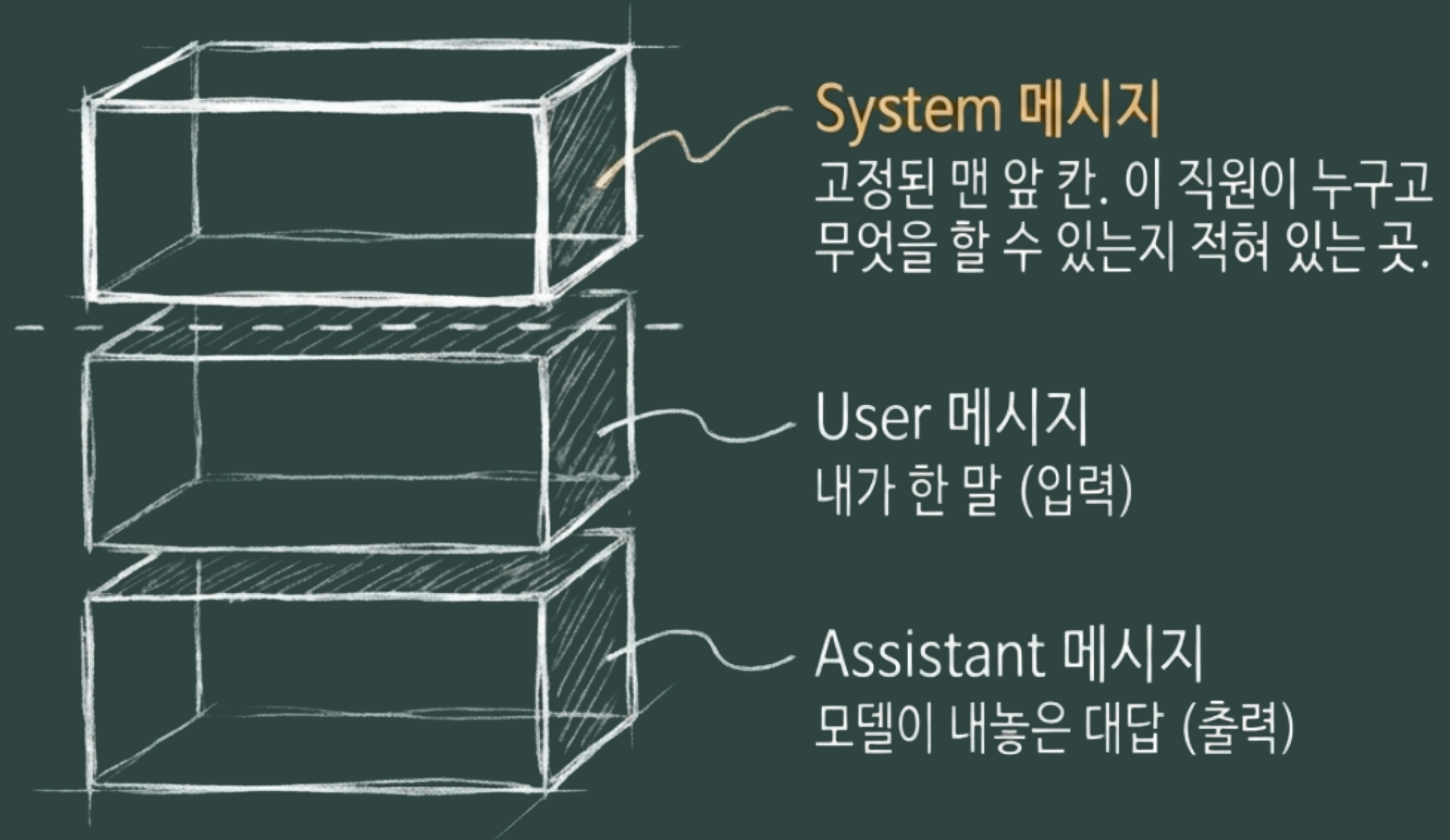
단 하나의 절대 규칙: 공간은 무한하지 않다.

컨텍스트는 끝없이 넓은 운동장이 아닙니다. 크기가 엄격하게 정해져 있는 유한한 작업 공간입니다.



작업 공간의 해부도

컨텍스트 공간은 메시지들이 차곡차곡 쌓이는 배열로 이루어집니다. 딱 세 종류의 칸만 존재합니다.



대화가 이어질 때 일어나는 일



- 한 번 오갈 때 새 문장 하나만 썩 처리된다.
- 매번 그동안 쌓인 전체가 통째로 다시 들어간다!

대화가 길어질수록, 매번 읽어야 할 내용이 기하급수적으로 누적됩니다.

필연적인 충돌, 그리고 복선



무한정 쌓을 수 없다

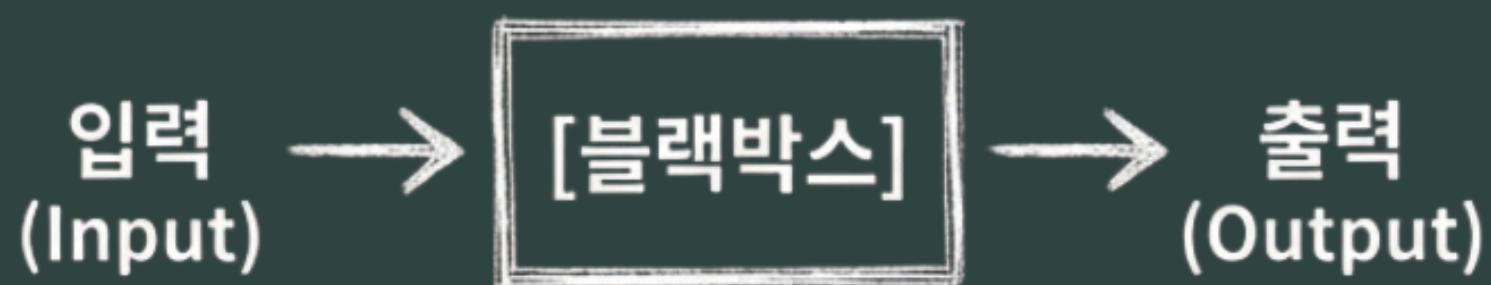
공간은 정해져 있는데, 읽어야 할 내용은 기하급수적으로 늘어납니다.
이것이 바로 우리가 나중에 ‘메모리’와 ‘압축’을 따로 배워야 하는 이유입니다.

LLM vs. 에이전트



이 둘을 가르는 결정적 차이는 무엇일까요?

LLM (텍스트 기계)

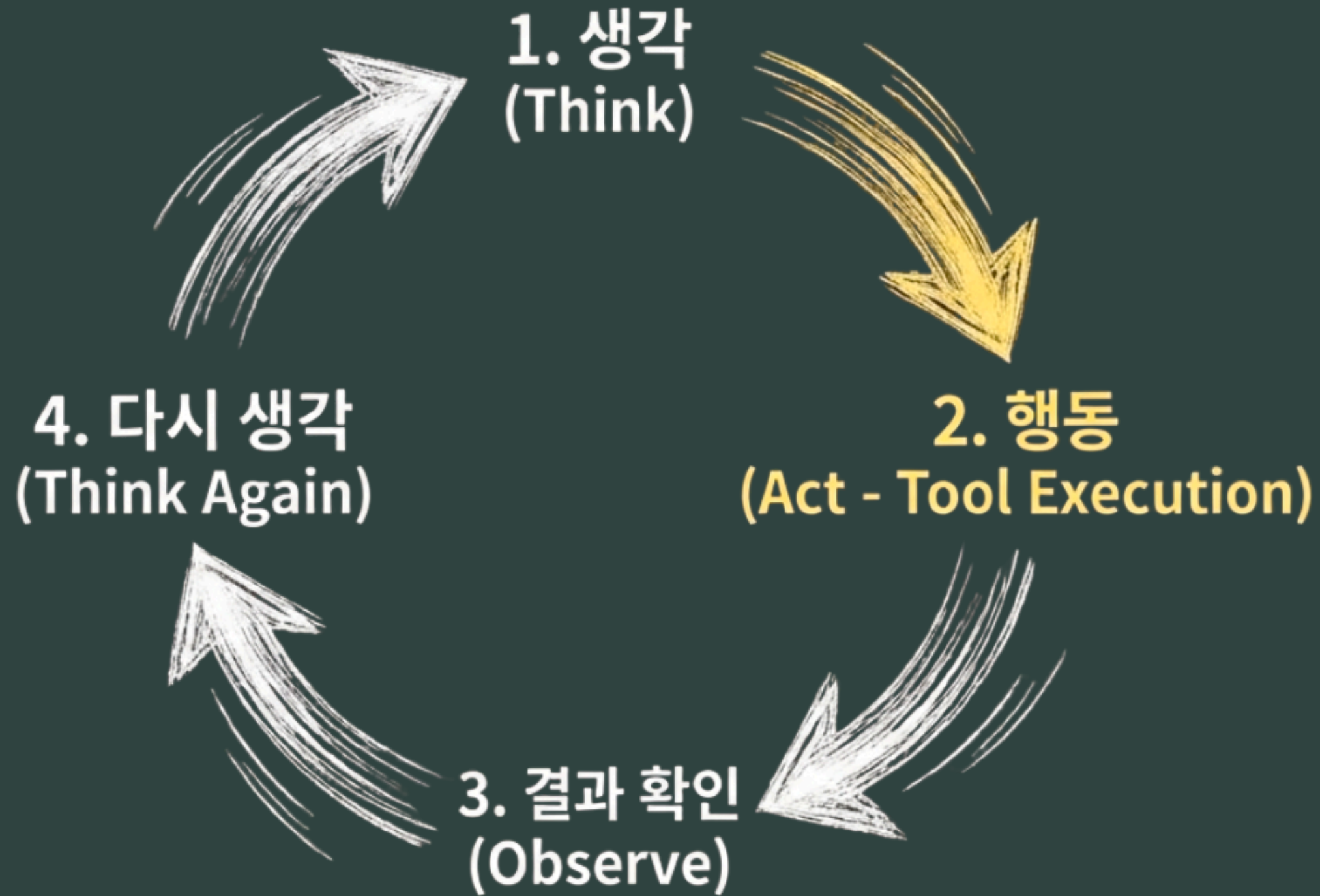


한 번 입력하면 한 번 출력하고 끝납니다.

에이전트 (행동가)



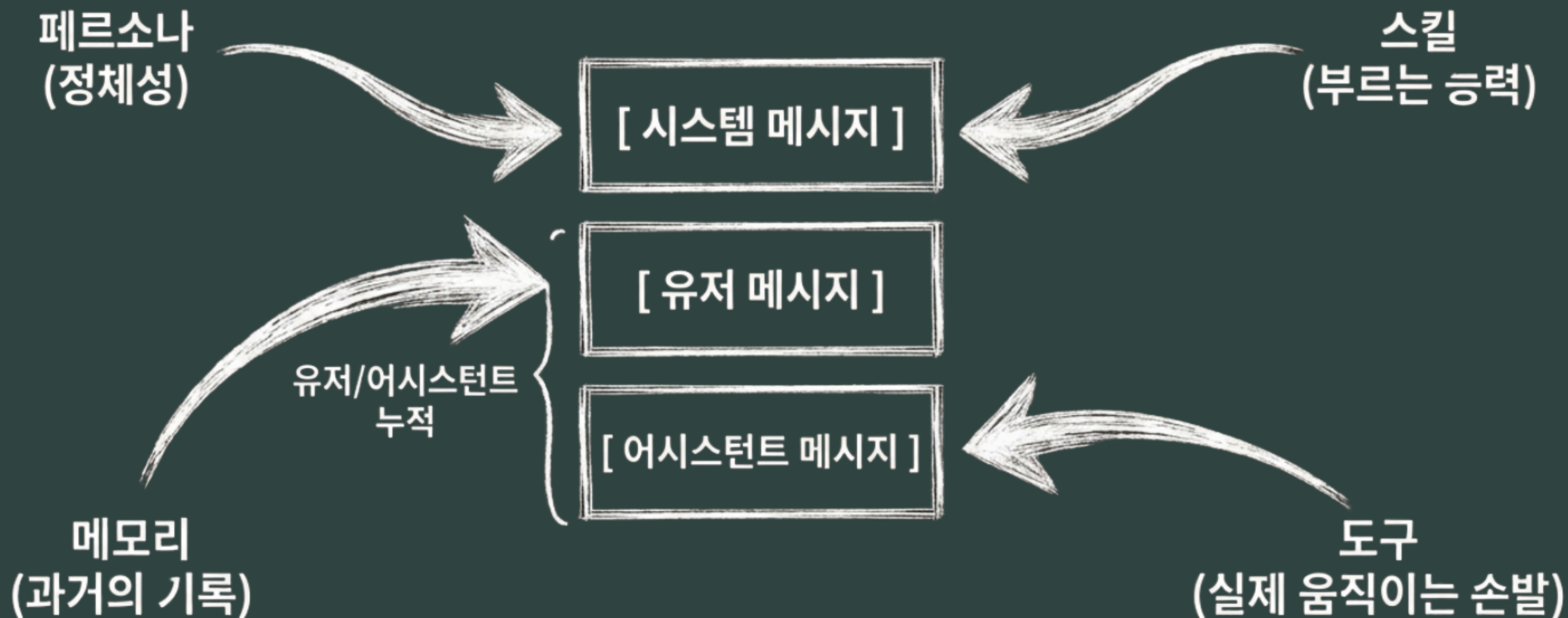
출력의 결과가 '도구 사용'의 판단이라면,
도구를 실행하고 그 결과를
다시 컨텍스트로 집어넣습니다.



이 되먹임 구조, **피드백 루프(Feedback Loop)**.

이 고리 덕분에 '뉴스 좀 정리해 줘' 한마디에 **검색, 읽기, 추리기, 초안 작성**이 스스로 굴러갑니다.

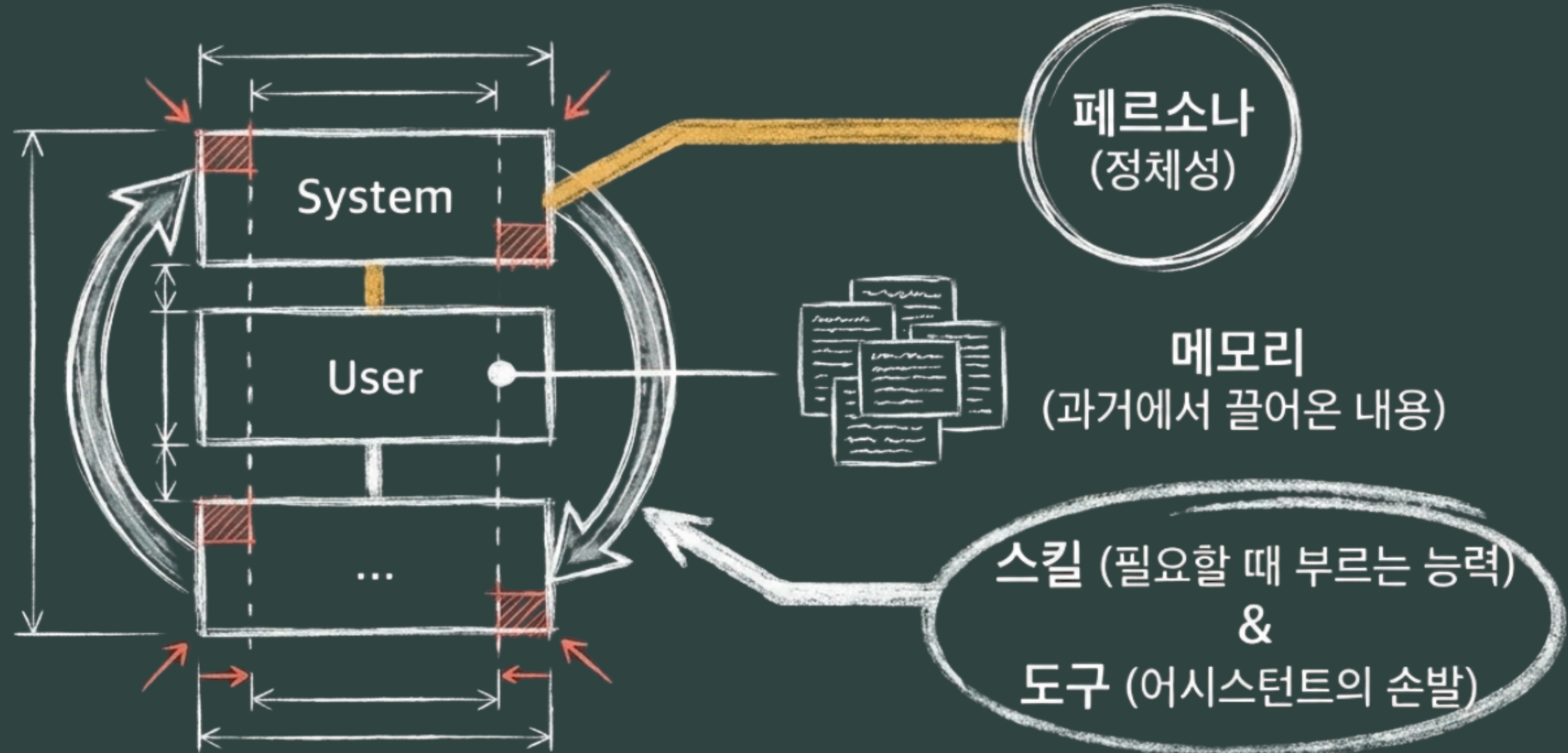
컨텍스트, 무엇을 어디에 채울 것인가?



모든 요소는 따로 놀지 않습니다. **정해진 칸에 정확히 꽂히는 부품들입니다.**

총합: 하네스(Harness) 설계도

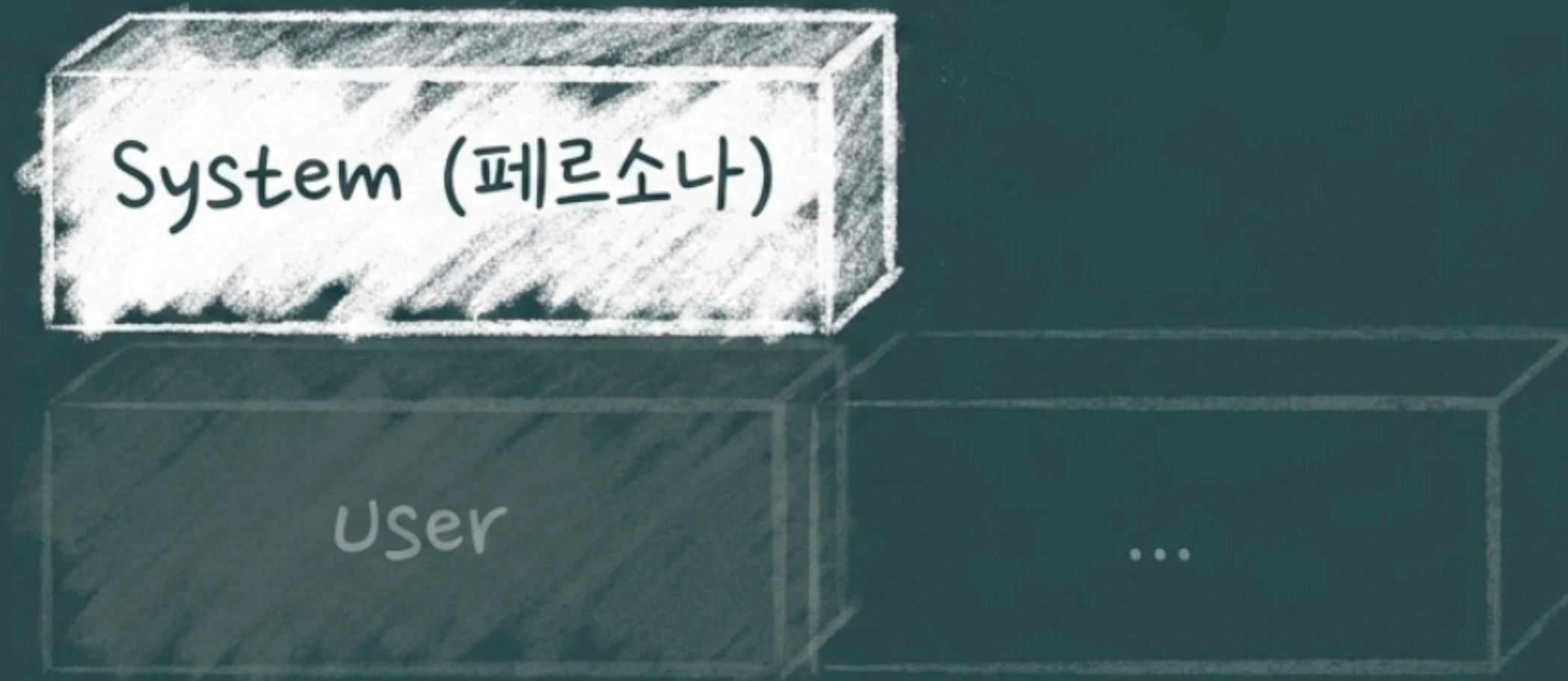
이 컨텍스트의 어느 칸을 어떻게 채울지 설계하는 일, 이것이 바로 '하네스'입니다.
우리가 앞으로 배울 모든 것은 이 구조도 위에 자리 잡습니다.



본격적인 하네스 설계는 자회사 설립 챕터에서 깊게 다룹니다.

이제, 채용하러 가보겠습니다.

컨텍스트를 채우는 첫 번째 작업은 시스템의 맨 앞칸, 직원의 정체성을 만드는 일입니다.
다음 시간, 우리의 첫 직원 '소피'에게 영혼을 불어넣는 작업부터 시작하겠습니다.



To be continued...

Dantelabs

Hermes × Codex 로 세우는 나만의 AI 가상 오피스

- 8주 마스터 과정

Section1. > Unit.2



첫 번째 AI 직원, '소피' 채용하기

헤르메스 에이전트
페르소나 생성 워크플로우



💡 Windows / macOS 환경별 숨긴 파일 보는 방법

Windows에서 숨김 파일 보기

1. 파일 탐색기에서 보기

Windows 11 기준:

1. 파일 탐색기 열기
2. 상단 메뉴에서 보기
3. 표시
4. 숨긴 항목 체크

그러면 숨김 파일/폴더가 보입니다.

Windows 10 기준:

1. 파일 탐색기 열기
2. 상단의 보기 탭 클릭
3. 숨긴 항목 체크

macOS에서 숨김 파일 보기

1. Finder에서 단축키로 보기

Finder에서 해당 폴더를 연 다음:

Command + Shift + .

즉,

⌘ + ⌘ + .

을 누르면 숨김 파일이 보입니다.

다시 누르면 숨김 파일이 다시 숨겨집니다.

헤르메스의 해부도 (핵심 3대 파일)

config.yaml



에이전트 설정 및
모델 정보.
(들여쓰기와 Key-Value 구조)

.env



API 키와 토큰.
외부 연동의 핵심.
보안 철저!

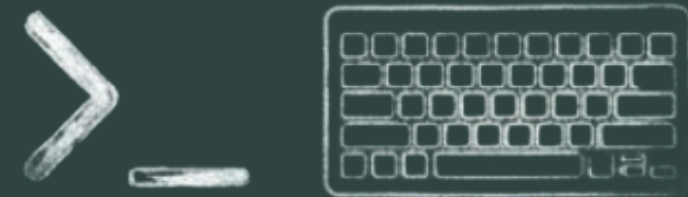
soul.md



페르소나의 본체.
오늘 우리가 다룰
핵심 파일!

두 가지 작업 방식 (CLI vs Web)

터미널 (CLI)



빠르고 프로페셔널함.

```
hermes profile create SOPHIE
```

```
hermes profile use SOPHIE
```

대시보드 (Web)



직관적이고 쉬움.



어느 쪽이든 결과는 같습니다!
나에게 편한 도구를 선택하세요.

터미널과 친해지기 (CLI Flow)

Step 1: 탄생

hermes profile create SOPHIE

(프로필 생성. clone 옵션으로 복사 가능)

Step 2: 확인

hermes profile list

(현재 등록된 프로필 목록 확인)

Step 3: 빙의

hermes profile use SOPHIE

(디폴트 대신 소피로 활성 에이전트 전환)

Step 4: 이별 (필요시)

hermes profile delete SOPHIE

(폴더와 함께 완전히 삭제)

성격 부여의 두 가지 레벨



[임시 성격] /personality 명령어

- /personality catgirl (고양이 말투)
- /personality noir (차분한 탐정)
- 특징: 현재 대화 세션에서만 유지되는 가벼운 스타일 전환.



[영구 페르소나] soul.md

- 에이전트의 영구적인 정체성 (프로필).
- 특징: 시스템 프롬프트 레벨에서 고정되는 근본적인 성격.

영혼의 구조 (Soul.md 4분할 캔버스)

Identity (정체성)

당신은 누구인가?

(예: 꼼꼼한 총무 비서)

Style (어투/스타일)

어떻게 말할 것인가?

(예: 단테님이라고 부르기, 간결한 비서형)

Avoid (금지사항)

절대 하지 말아야 할 것!

팩트 체크, 민감한 개인정보 외부 유출 금지

Default (기본 행동)

모호할 땐 어떻게 할까?

한 번만 짧게 되묻기, 3줄 요약 우선하기

소피(Sophie) 온보딩 성공!



빈 텍스트 파일에서 시작해 완벽한 맞춤형 AI 어시스턴트가 탄생했습니다.

Dantelabs

Hermes × Codex 로 세우는 나만의 AI 가상 오피스

- 8주 마스터 과정

Section1. > Unit.3



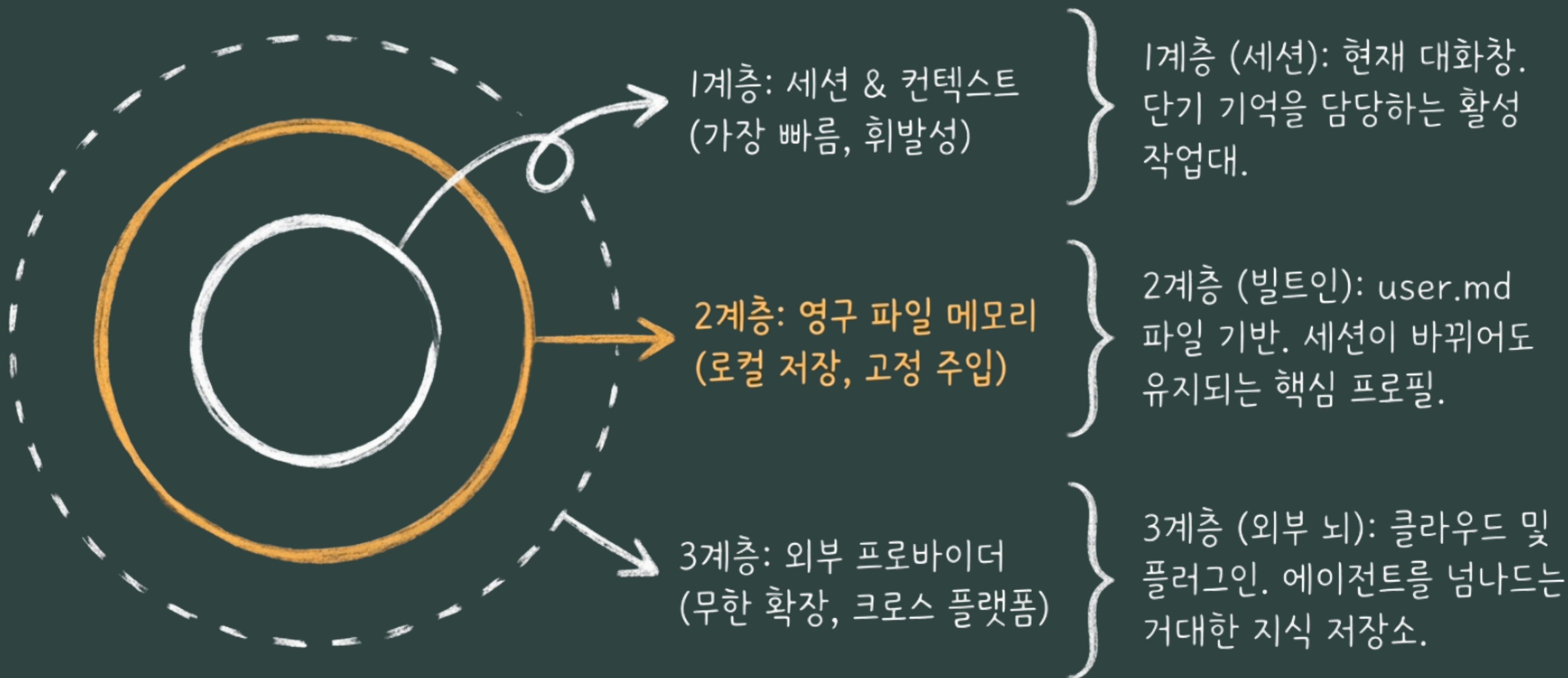
헤르메스 메모리 관리법

3계층 기억 시스템

Part 1



헤르메스의 3중 기억 구조



1계층: 세션과 컨텍스트

세션 = 현재 대화창 / 컨텍스트 = LLM 입력창에 쌓이는 메시지 배열

Context Window



~~Slack~~

~~CLI~~

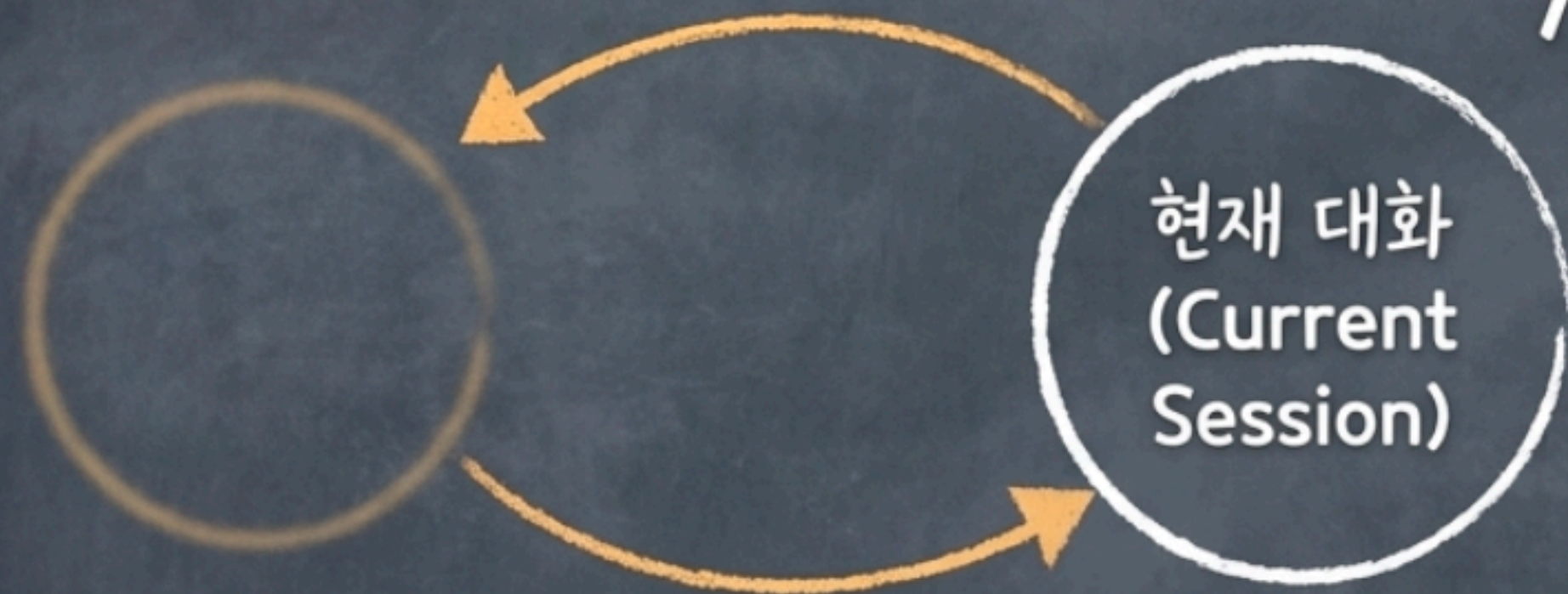
채널별 완전한 분리
(Slack, Telegram, CLI 세션은
서로 컨텍스트를 공유하지 않음)



모든 대화는 로컬 SQLite
(sessions.json 및 .db 파일)에
안전하게 자동 저장됨

기억의 방 이동하기

/resume [세션명]



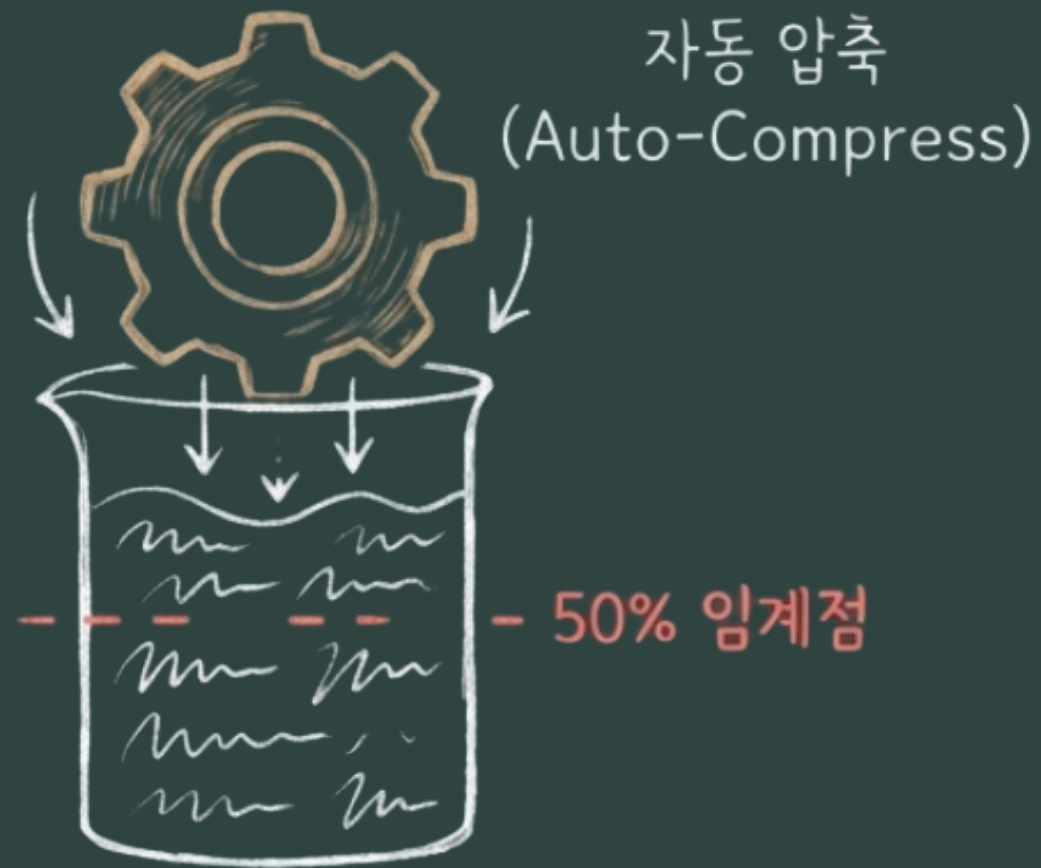
과거에 나누었던 특정 대화 세션으로
즉시 복귀하여 맥락 이어가기

/new 또는 /reset



현재 단기 기억을 비우고
완전히 새로운 컨텍스트 시작

단기 기억의 자체 정화 매커니즘



용량 관리: 대화가 50~70% 차오르면 속도 저하 방지를 위해 헤르메스가 자동으로 대화를 압축합니다.



시간 관리: 24시간 동안 아무런 대화가 없으면, 쾌적한 환경을 위해 세션이 자동 리셋됩니다.

수동 제어 커맨드

압축하기: /compress [주제]

새 출발: /new 또는 /reset

검색은 기억이 아닙니다.

- 세션 서치(Session Search)는 과거 세션을 텍스트로 조회하는 '결정론적 검색' 도구입니다.
- ✓ 키텍스트(기억) 공간을 차지하지 않으므로 토큰 소모가 매우 적습니다.
- 대화 세션이 바뀌어도 과거 의제나 표를 불러오는 데 완벽하게 작동합니다.

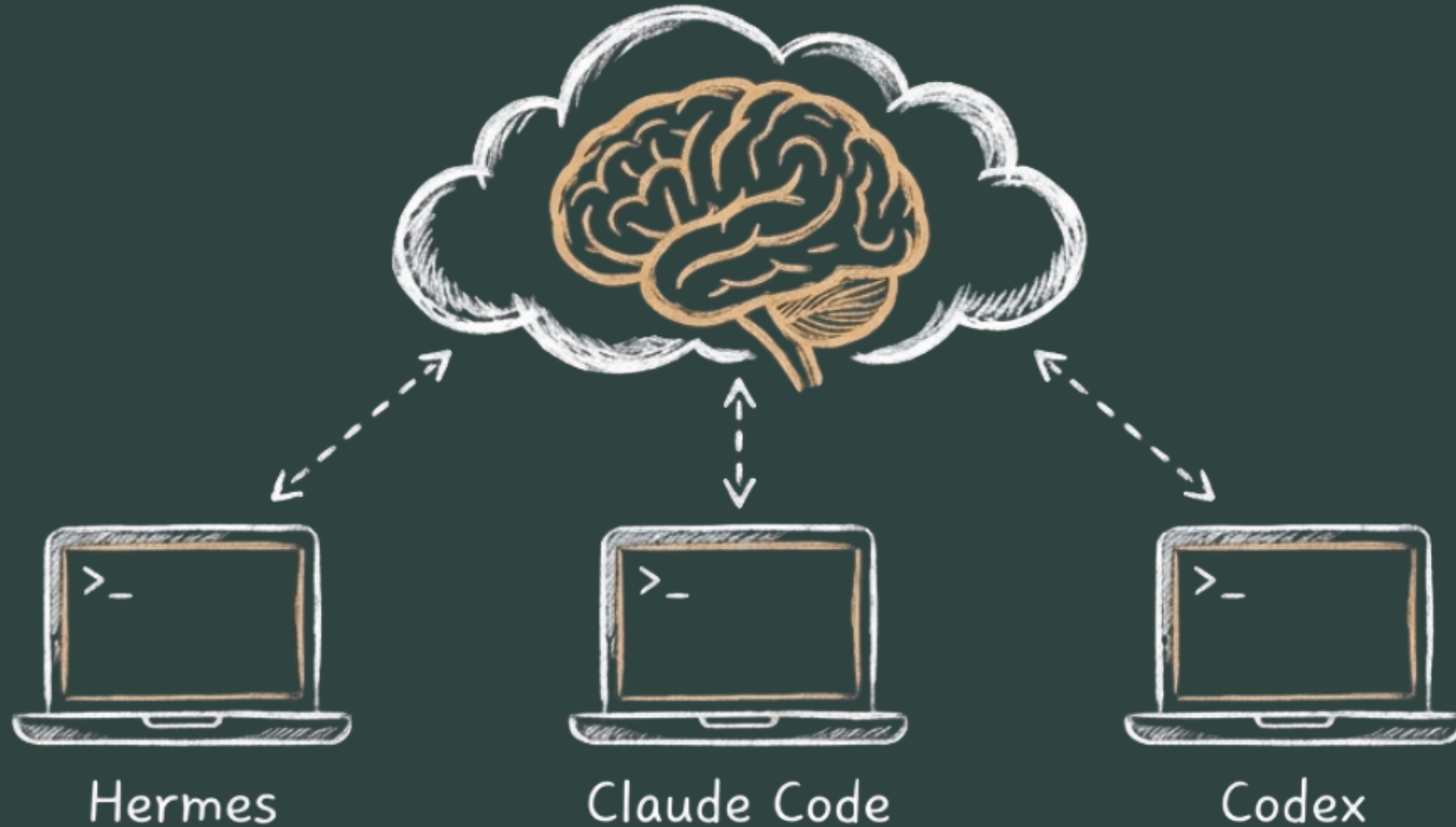


2계층 메모리 파일의 구조와 위치

| | | |
|-------|------------------------|----------------------------|
| 파일 | user.md | memory.md (or agent.md) |
| 용도 | 사용자 프로필 & 성향 | 프로젝트 환경 & 컨벤션 |
| 글자 제한 | 최대 1,375자 | 최대 2,200자 |
| 특징 | 자연어 대화로 즉시 수정/삭제 가능 | 유니버설 메모리로 타 도구와 공유 가능 |

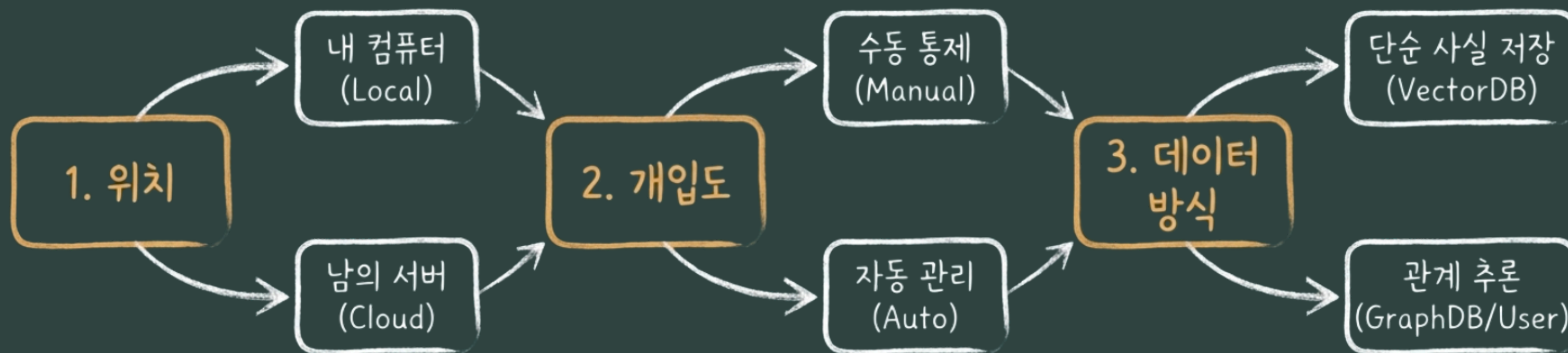


3계층: 외부 메모리 프로바이더 (선택 사항)



- **크로스 플랫폼:** 헤르메스뿐만 아니라 Claude Code, Codex 등 다양한 에이전트를 넘나들며 하나의 기억을 공유합니다.
- **무한한 확장:** 로컬 텍스트 한계를 넘어, 지속적으로 누적되는 당신만의 거대한 지식 베이스를 구축합니다.

나에게 맞는 외부 메모리 찾기 (3가지 질문)



로컬: 무료/비공개/키 불필요

VS

클라우드: 관리 편의성/어디서든 접근

수동: 섬세한 제어 가능

VS

자동: 알아서 수집하고 검색

VectorDB: 단순 사실과 데이터 검색

VS

Graph/추론: 관계성 파악 및 사용자 분석

주요 외부 프로바이더 비교표

| 도구명 | 위치 | 자동화 여부 | DB 방식 / 특징 |
|-------------------------|------|--------|------------------------------|
| 추천 : 입문용 Holographic | 로컬 | 수동 | Vector DB (무료, 기본 제공으로 추천) |
| Mem0 ★ 추천 : 자동화 | 클라우드 | 자동 | Vector DB (편리한 자동화, 초보자 추천) |
| Hindsight | 클라우드 | 자동 | Graph DB (Reflect 기능, 정보 엮기) |
| Honcho | 클라우드 | 자동 | 사용자 모델링 (멀티 에이전트, 성향 추론) |

Q1. 기억을 어디에 저장할까요?



로컬 (Local)

- 특징: 완벽한 비공개, 무료
- 조건: 가입 불필요, 키(Key) 발급 불필요
- 내 기기에 단단히 묶어두고 싶을 때
- ↳ 대표 도구: Holographic



클라우드 (Cloud)

- 특징: 관리가 편함, 어디서든 접근 가능
- 조건: 가입 필수, 키(Key) 발급 필수
- 여러 기기에서 유연하게 사용하고 싶을 때
- ↳ 대표 도구: MEMO

Q2. 얼마나 개입하고 싶으신가요?

자동 관리

- '이것 좀 기억해!' 요청 없이도 알아서 검색하고 저장
- 장점: 관리에 에너지가 들지 않음
- ↳ 도구: MEMO, 혼초(Honcho)

수동 통제

- 직접 도구를 사용해 넣고 빼며 명시적으로 통제
- 장점: 내 스타일에 맞춘 섬세한 제어 가능
- ↳ 도구: Holographic, OpenBiking



Q3. 어떤 스마트함이 필요한가요?



Vector DB

- 단순 사실 저장 및 검색 (가장 기본적인 방식)
- 데이터베이스에서 꺼내어 컨텍스트에 삽입

도구: Holographic, MEMO



Graph DB

- 지식 그래프 (사람과 프로젝트의 관계성 파악)
- Reflect 기능으로 흩어진 기억을 엮어 종합

도구: Hindsight



사용자 모델링

- 대화를 분석해 '사용자가 어떤 사람인지' 추론하여 정제 저장
- 오랜 시간 쓸수록 나를 깊이 이해함
- **유료 서비스**

도구: 혼초 (Honcho)

Dantelabs

Hermes × Codex 로 세우는 나만의 AI 가상 오피스

- 8주 마스터 과정

Section1. > Unit.4





AI 에이전트를 위한 외부 메모리 관리법

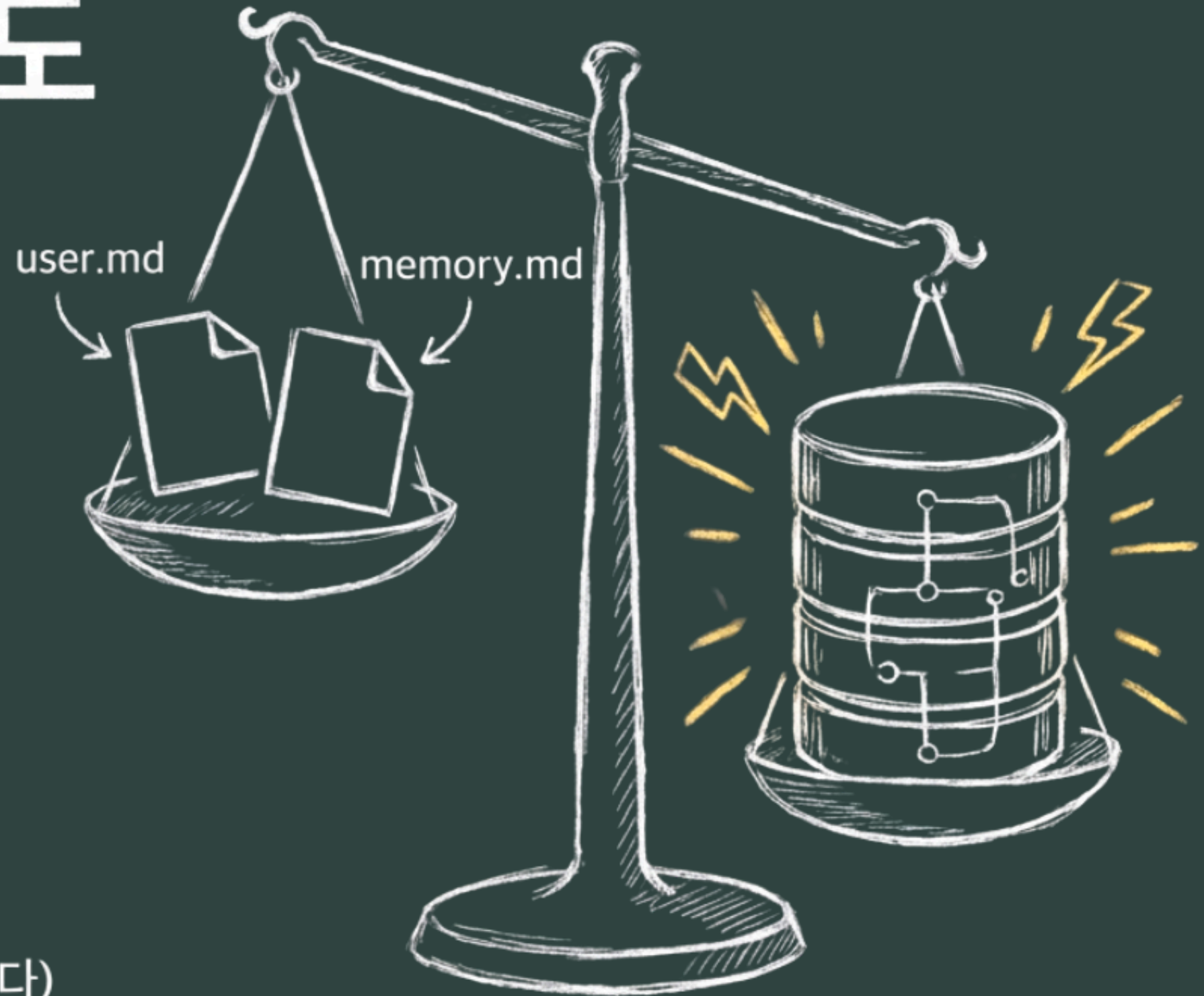
프로바이더 레시피

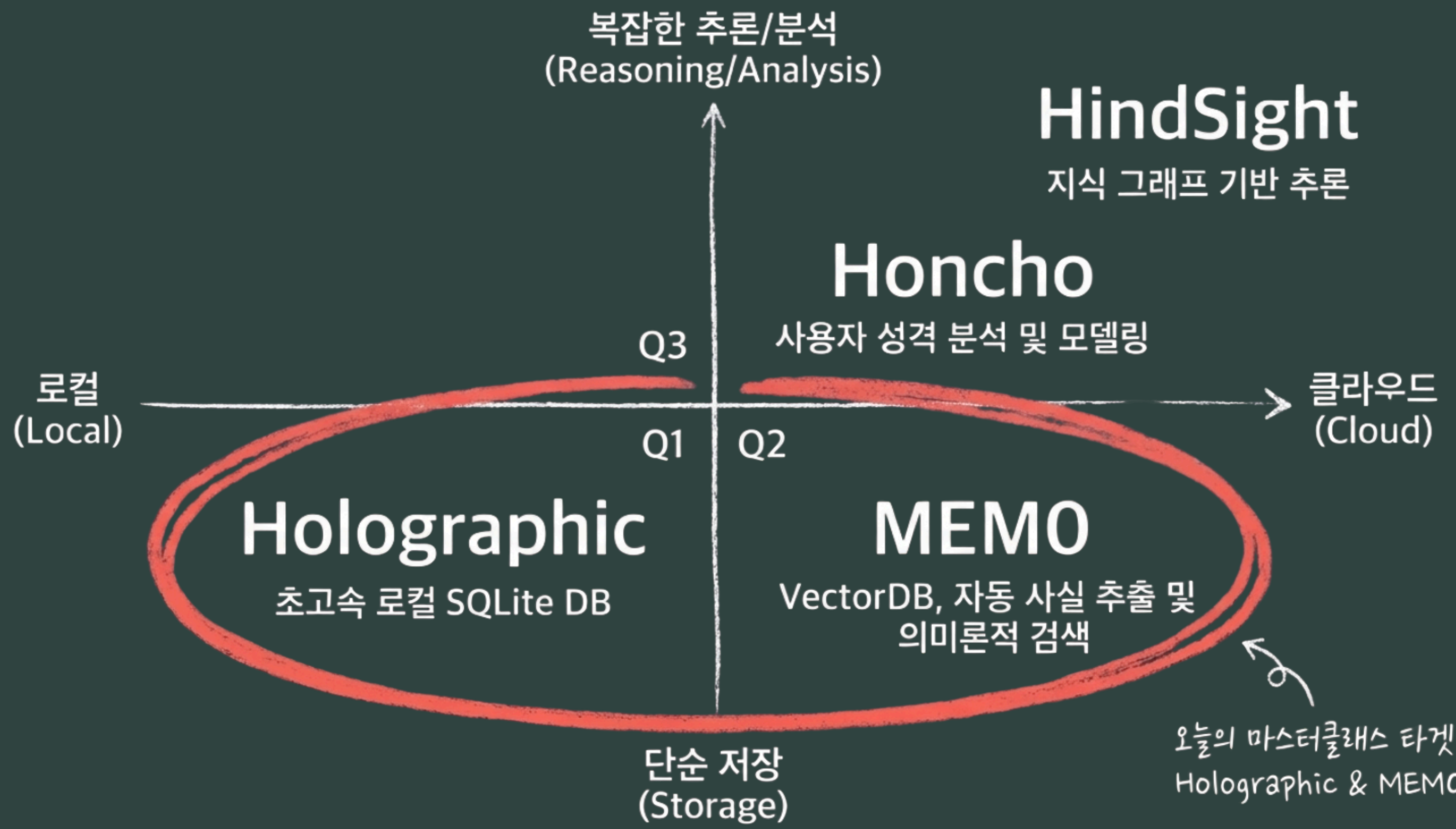


빌트인 메모리로도 충분할까요?

대화 기반의 user.md와 자동 생성되는 memory.md는 가볍고 훌륭합니다.

하지만 외부 프로바이더를 연결하면?
더 강력한 지능, 크로스 플랫폼
중앙 관리의 문이 열립니다.
(동시 사용은 불가, 단 하나의 무기만 선택해야 합니다)





내 작업실엔 어떤 도구가 맞을까? (결정의 시간)

| 구분 | Holographic | MEMO |
|---------|------------------------|--------------------------------|
| 환경 & 비용 | 로컬 SQLite ★ (완전 무료) | 클라우드 VectorDB (무료 크레딧 + 유료) |
| 속도 & 관리 | 초고속 읽기·쓰기 | 중앙화된 대시보드 관리 ★ |
| 핵심 기술 | Trust Score 알고리즘 | Semantic Search & Reranking |

빠른 로컬 테스트는 Holographic, 고도의 문맥 파악과 서비스화에는 MEMO를 선택하세요.