옵시디언과 LLM WIKI로 만드는 연구 지식 베이스

AI를 잘 쓰는 연구자가 아니라, AI가 일할 수 있는 지식 구조.

쓰는 방식이 **축적되느냐, 매번 증발하느냐**가 갈림길입니다.
연구자의 지식 시스템(볼트 · LLM Wiki · 하네스) 위에 AI Agent를 올리는 법을 다룹니다.

WHO · 구요한 · YOHAN KOO

연구하면서 가르치는 실천가.

오늘 보여드리는 것은 이론이 아니라, 3년 · 10,000개 노트로 직접 운영해 온 실물입니다.

교육 · 연구

KIRD 객원교수

차의과학대 겸임교수 · 대학과 연구 현장에서 AI × 지식관리 교육.

기업 교육

LG 회장단·임원 교육

각사 CEO·본부장 1on1 AI 코칭 — 의사결정권자의 지식 운영체제 설계.

실천

10,000+ 노트 볼트

더배러 뉴스레터 · 옵시디언 × 생성형 AI 지식관리 시스템 운영.

출연연의 현재 · NST 정책연구 (헬로디디 2026-07 보도 · 단일 조사)

93%가 쓰는데,
영수증을 풀칠하고 있다.

93%

이미 연구에 활용 중 (출연연 종사자 242명 설문)

5%

조만간 활용 계획

2.1%

이용하지 않음

현장은 — AI 구독료 **영수증 증빙이 매월 반복**, 대부분 기관에 생성형 AI 보안규정 부재로 **책임은 연구자 개인에게 전가**, 일부 기관은 아예 차단.

문제는 "쓸 것이냐"가 아니라 "쓰는 방식이 축적되느냐, 증발하느냐"

2026 정부출연연구기관 업무보고

"모든 연구기관이 AI 전환을 하려면 AI가 잘 읽을 수 있는 형태의 데이터를 체계적으로 모아야 하는데 한계가 있다"

— 배경훈 부총리 겸 과기정통부 장관 · 출연연 업무보고

오늘 두 시간은 이 문장에 대한 **개인 수준의 실행 답안**입니다. 조직의 데이터 체계는 기다려야 하지만, **내 연구 지식의 구조는 오늘 밤부터** 바꿀 수 있습니다.

국가가 이미 움직이고 있다

주무 부처가 트랙을 열었고, 공모 마감은 닷새 뒤입니다.

과기정통부 · CO-SCIENTIST CHALLENGE

트랙 2개

① AI 기반 연구보고서 작성 ② 과학기술 AI 에이전트 개발 — 주무 부처의 공식 트랙 이름

NST · 국가과학AI연구센터 (NAIS)

해커톤 마감 D-5

〈2026 NAIS AI 해커톤〉 연구개발 특화 AI Agent — 9/7(일) 17시 마감 · 본선 9/30~10/1 (AI4Sci Korea 2026) · 누구나 최대 5인 팀

NAIS는 출연연 전용 온프레미스 AI 환경과 연구용 샌드박스를 준비 중이고, 'AI 과학자 플랫폼' 베타는 연내 예정입니다. "보안 때문에 못 쓴다"를 국가가 풀겠다고 나선 겁니다 — 다만 **그 사이 넉 달을 어떻게 쓰실 겁니까?**

출처: co-scientist.kr · nst.re.kr · ai4scikorea.org

NATURE 독자 폴 (2026-06 · N=1,907 · 확률 표본 아님)

쓰지만, 좋아하지 않는다.

66%

월 1회 이상 AI 사용 (매일 25 · 주 26 · 월 15)

48%

AI에 부정적 감정 (긍정 30%)

63%

"LLM 데이터·문헌 분석은 위험이 이익보다 크다"

이 불편함은 비합리가 아니라 **정확한 감각**입니다. 오늘은 그 불편함이 어디서 오는지 짚고, **구조로 푸는** 이야기를 합니다.

출처: Nature, "Scientists have a bad case of AI FOMO" (2026-06)

같은 질문, 상반된 두 연구

AI는 연구를 빠르게 하는가.

설계	관측 연구 (arXiv 2412.07727 · v3)	무작위 대조시험 (METR)
표본	자연과학 논문 4,130만 편 관측	n=16 · 실제 이슈 246건
결과	**논문 3.02배 · 인용 4.84배** · 팀리더 1.37년 단축	완료 시간 **19% 증가** (느려짐)
강점 / 약점	규모 / 선택 편향	인과 식별 / 소표본 · 코딩 한정

두 결과가 **동시에 참일 수 있다**는 것 — 여기서 오늘 이야기가 시작됩니다.

출처: arXiv 2412.07727 (v3) · arXiv 2507.09089 (METR)

METR의 백미 · 인식 격차

빨라졌다고 느꼈지만,
실제로는 느려졌다.

−24%

사전 예측 — "이만큼 단축될 것"

−20%

사후 추정 — "이만큼 단축됐다"

+19%

실측 — 완료 시간 증가

후속 측정은 METR 스스로 **선택 편향으로 해석 불가** 판정. **측정조차 어려워졌다** — 그래서 감각이 아니라 구조로 접근해야 합니다.

출처: METR (arXiv 2507.09089 · 2026-02 설계변경 공지)

연구자의 진짜 병목

병목은 생성이 아니라 검증이다.

11~57% 상용 배포 모델의 인용 조작률

14~95% GhostCite 벤치마크 (13개 LLM · 인용 37.5만 건)

최악 확장 추론에도 인용 정확도는 과제군 중 최하위

환각률 자체는 개선 추세입니다. 그러나 **인용**은 다릅니다 — 초안이 아무리 빨라져도 인용을 못 믿으면 **검증 비용이 생성 이득을 잡아먹습니다.**

출처: GhostCite (arXiv 2602.06718) 외 · 프리프린트 중심 — 수치는 범위로만

연구자에게 익숙한 질문

"무엇을 측정하지 않고 있는가."

기존 벤치마크가 재는 것

Support

인용이 주장을 뒷받침하는가는 잰다.

재지 않는 것

Existence

인용이 실재하는가는 재지 않는다 — 그래서 신뢰도를 체계적으로 과대평가.

실무 원칙은 하나 — 인용을 **생성하는 도구가 아니라 검색해서 가져오는 도구**를 쓸 것. 이것이 오늘의 설계 목표, **검증 가능한 구조**입니다.

PART 01 · 연구 지식 베이스의 뼈대

프롬프트가 아니라 구조가 차이를 만든다.

v1 프롬프트 → v2 컨텍스트 → v3 하네스. 3년간 무엇이 바뀌었는가.

3세대 진화

입력은 짧아지고, 답은 깊어진다.

V1 · PROMPT (2022–23)

"무엇을" 묻느냐

"논문 요약해줘" — 누가 물어도 같은 답.

V2 · CONTEXT (2024)

"어떤 상황에서"

역할·배경 패키징 — 좋지만 매번 처음부터.

V3 · HARNESS (2025–)

"어떤 환경에서"

지식·도구·권한·룰을 한 번 설계 → **한 줄로 내 맞춤 답.**

v1→v3에서 **사용자 입력은 다시 짧아집니다.** 그런데 출력은 더 풍부하고, 재현 가능하고, 축적됩니다. 차이를 만드는 건 프롬프트가 아니라 **하네스**.

권위 보강 · ANTHROPIC

컨텍스트 엔지니어링 = "원하는 결과를 낼 확률을 극대화하는 고신호(high-signal) 토큰의 최소 집합을 찾는 것"

— Anthropic · 하네스 = 컨트롤 플레인 · "두뇌와 손의 분리"

제가 '하네스'라 부르는 걸 Anthropic은 '컨트롤 플레인'이라 부릅니다. **이름만 다를 뿐 업계가 같은 곳을 가리킵니다.** 그리고 Claude Code SDK가 **Agent SDK로 개명**됐습니다 — 코딩에서 검증된 하네스가 연구로 일반화되는 중.

출처: Anthropic — Effective context engineering for AI agents

연구자의 하네스 = 이미 여러분 손에 있다

지식의 IDE, 공유 언어, 엔진.

Obsidian	지식의 IDE — 마크다운 볼트 · 연구노트 · 개념 위키를 한 화면에서
Markdown	인간과 AI의 공유 언어 — 모든 주요 LLM이 마크다운으로 학습됨
AI	그 위에서 도는 엔진 — 모델은 갈아끼워도 지식 구조는 남는다

코딩 에이전트의 하네스가 도구·권한·지식베이스·룰이라면, **연구자의 하네스는 볼트 + 레퍼런스 매니저 + 연구노트 + 개념 위키 + 시스템 파일**입니다.

왜 마크다운인가

포맷은 사고를 강제한다.

포맷	강제하는 사고
PPT	슬라이드 단위 사고 — 닫힌 포맷, AI에게도 닫힌 지식
표	비교 사고
MARKDOWN	**구조화 + 연결 사고** — 에이전트가 파싱 레이어 없이 그냥 읽는다
YAML	필드 수준 의미 선언 — 기계가 읽는 메타데이터

kepano의 "file over app" — **논문을 쓰기 전에, 어떤 포맷으로 지식을 쌓을지를 먼저 결정해야 합니다.**

같은 질문, 세 개의 답 + 그래프뷰

"내 연구 분야 최근 동향 정리하고 우리 과제와의 접점 찾아줘."

V1

일반론

교과서·위키 수준. 누가 물어도 같은 답.

V2

내 분야 특이적

좋아졌지만 이번 한 번만 — 매번 손으로 재입력.

V3

한 줄 + 축적

볼트의 연구노트·레퍼런스·위키가 하네스 — **내 선행연구가 반영된 답, 재현 가능.**

INTERMISSION

휴식, 15분.

2부는 오늘의 본론 — LLM Wiki, 지식이 복리로 쌓이는 구조입니다.

PART 02 · LLM WIKI

지식이 복리로 쌓이는 구조.

볼트가 일하는 공간이라면, 위키는 지식이 축적되는 코드베이스입니다.

LLM WIKI 패턴 · 2026

> "Obsidian is the IDE,
> the LLM is the programmer,
> the wiki is the codebase."
>
> — Andrej Karpathy · LLM Wiki (2026)

에이전트가 **불변 원문(raw sources)**을 지속·상호링크된 마크다운 위키로 **컴파일**합니다. 읽을 때마다 새로 요약하는 게 아니라 기존 페이지를 갱신 — 그래서 복리.

출처: Karpathy · LLM Wiki (GitHub gist)

아키텍처 대비

검색은 휘발되고, 위키는 쌓인다.

전통 RAG

영구 기억상실

매 쿼리마다 chunk를 재검색하고 결과는 폐기. `perpetually amnesiac`

LLM WIKI

복리 축적

한 번 합성한 결과가 페이지로 남고, 다음 질문은 그 페이지 위에서 시작.
`knowledge compounds`

어제의 답이 오늘의 출발점이 되는 구조 — 연구 지식에 필요한 건 이것입니다.

LLM WIKI의 3층 구조

원문과 해석을 분리한다.

Schema	CLAUDE.md · AGENTS.md — 해석의 질서를 선언하는 시스템 파일
Wiki	LLM이 관리하는 상호링크 마크다운 페이지 — 갱신되고 복리로 쌓임
Raw Sources	논문·기사·전사 원문 — 불변 보존, 위키가 여기를 인용

연구자에게 익숙한 감각입니다 — **사료와 논문을 분리하는 것.** 원문은 건드리지 않고, 해석만 계속 갱신합니다.

인용 조작 문제의 구조적 해법

내가 넣은 것 안에서만 말하게 하라.

도구가 검색한 코퍼스

조작 위험

ChatGPT·Gemini Deep Research — 인용 조작 위험(§검증 병목)을 안고 감 · 매번 휘발

내가 큐레이션한 코퍼스

내가 넣은 것만

NotebookLM · 내 볼트 · LLM Wiki — 준 소스에만 근거 · 복리로 축적

AI가 인용을 만들어내지 못하게 하는 가장 확실한 방법 — **내가 넣은 것 안에서만 말하게 하는 것.** 검증 병목과 LLM Wiki가 한 줄로 연결됩니다.

본인 논지 · 정직한 예방주사와 함께

위키보다 먼저 필요한 것은 스키마다.

문서를 쌓기 전에 **해석의 질서**를 설계하세요. 파일이 약 200개를 넘고 에이전트가 전체 그래프를 담지 못하면 패턴이 무너집니다 — 그때 필요한 게 경량 인덱스와 스키마.

Forte의 경고 — "노트가 엉망이면 AI가 구원하지 않는다." AI는 PKM을 대체하지 않고 leverage를 바꿉니다. 잘 쓰면 구조화된 memory로, 못 쓰면 문서를 high-speed drift로.

MCP · 표준이 된 연결 계층

벤더 락인이 아니라, 벤더 탈락인.

1

2024.11

Anthropic 공개

Model Context Protocol 발표

2

2025.03

OpenAI 채택

경쟁사가 4개월 만에 공식 채택

3

2025.12

Linux Foundation

Agentic AI Foundation에 기부

4

2026.02

Platinum 멤버

AWS·Google·Microsoft·OpenAI 등 8개사

5

2026.07

최대 규모 개정

구 버전 12개월 deprecation 창

모델 계층에서 싸우는 회사들이 인프라 계층에서는 협력합니다 — **Linux · Kubernetes가 걸어간 길과 같은 패턴.**

출처: modelcontextprotocol.io · 규모 수치는 방향성만 (18개월 만에 세 자릿수 배수 성장)

연구 도구 연계 실물

내 레퍼런스 매니저가 에이전트의 손에 연결된다.

zotero-mcp

Zotero 라이브러리 전체 유사도 검색 · DOI/ISBN/BibTeX로 논문 추가 · 오픈액세스 PDF 자동 수집

mcp-research

arXiv·Semantic Scholar 다중 소스 검색 · BibTeX/RIS 내보내기 · 중복 제거 · Zotero 자동 추가

Scholar Sidekick

10,000+ 인용 스타일 변환 · **철회(retraction) 확인 · 인용 조작 탐지** · 오픈액세스 확인

한계도 정직하게 — 대부분의 학술 MCP 서버는 검색·메타데이터에 집중하고 **인용 내보내기는 Zotero 계열이 그 간극을 메웁니다.**

이 강의에서 가장 강한 시연

에이전트가 인용을 검증한다.

LIVE

철회 논문 · 조작 인용 탐지

참고문헌 목록 → 에이전트 검증 파이프라인

AI가 초안을 쓰는 건 이제 놀랍지 않습니다. 놀라워야 할 건 — **AI가 자기가 쓴 인용을 스스로 검증하게 만드는 것.**

프론티어 랩 3사의 상이한 베팅

"새 모델이 아니라 워크플로로 과학자를 설득한다."

Anthropic

모델이 아니라 **워크플로** — Claude Science: 연구 도구 통합 · **감사 가능한 산출물** (Novo Nordisk · Allen Institute)

Google DeepMind

멀티 에이전트 — Co-Scientist: 특화 에이전트 연합의 가설 생성

OpenAI

범용 추론의 과학 적용 — OpenAI for Science · Prism 워크스페이스

성능이 아니라 **재현성 · 감사 가능성 · 기존 도구 통합**이 채택을 결정합니다 — 이 문장이 오늘 강의 전체의 요약이기도 합니다.

출처: TechCrunch (Claude Science) · Google DeepMind Blog

과장 없이, 궤적으로

회의에서 Nature까지, 1년. 검증은 아직 실험실 단계.

2025.03 · 발표 직후

실명 연구자들의 회의

"세부가 부족하다" · "이미 알려진 약물, 새로울 게 없다" · "가설 생성은 가장 즐거운 부분인데" (Sinapayen)

2026.05 · NATURE 게재

실험실 검증까지

cf-PICI 가설이 미발표 실험 결과와 일치 — 다만 임상은 아직. AlphaFold 없이 단백질 접힘을 예측하는 에이전트는 없다.

남는 질문 — **즐거운 노동을 자동화하는 방향이 맞는가.** 실행은 넘기되, 무엇을 지킬지는 연구자의 선택입니다.

규범이 곧 온다

공통 원칙 하나 — AI는 저자가 될 수 없다.

출판사	입장
SCIENCE (AAAS)	가장 엄격 — AI 생성 텍스트 사실상 전면 금지, 프롬프트·출력 요청 시 제공
NATURE / SPRINGER	AI 저자 금지 · 실질 사용은 Methods 기재 · **순수 copy editing 면제** · 심사자의 원고 업로드 금지
ELSEVIER	별도 AI declaration statement · 편집자·심사자의 AI 사용 금지

현 정책은 '생성형 AI 도구' 일반 — **자율 에이전트 규정은 아직 공백.** 이 공백이 지금 **감사 로그 남기는 구조**를 만들어야 하는 이유입니다. 국내도 AI연구윤리 가이드라인·국가연구데이터법 추진 중.

연구보안 실무 · 국가연구개발혁신법 제40조 + N2SF 3등급

등급을 나누는 게 답이다.

넣어도 되는 것	판단 유보 — 기관 확인	절대 금지
공개된 논문·특허·표준	미공개 실험 데이터 (등급 확인)	국가연구개발과제 미공개 정보
내가 쓴 초안 문장 (민감정보 제거)	과제 제안서 일부 표현	보안과제 자료 일체
공개 데이터셋	공동연구 파트너 자료	심사 중인 타인 원고·심사보고서
개념 질문·방법론 상담	내부 회의록	개인정보 포함 데이터

"AI 쓰지 마세요"도 "마음껏 쓰세요"도 답이 아닙니다. **등급을 나누는 게 답**이고 — 등급을 나누려면 **지식이 이미 구조화돼 있어야 합니다.**

출처: 한국연구재단 권고 (제40조) · N2SF 기밀(C)/민감(S)/공개(O) 차등 통제

영수증 이야기의 회수

제도가 바뀌기를 기다리는 동안, 지식이 증발하지 않게 하는 법.

제도는 "가능하다, 다만 명확히 관리되어야 한다"고 말하는데, 그 단서 조항이 현장에서는 "그냥 하지 말자"로 귀결되곤 합니다. PBS 전환기의 불확실성도 여러분이 직접 당사자입니다.

오늘 드리는 건 제도를 바꾸는 방법이 아니라, 그 사이 여러분 개인의 지식을 지키는 방법입니다.

AKM INDEX · 오늘 깐 사다리가 그대로 측정축

AI가 여러분을 분류하기 전에, 먼저 측정하세요.

P	**Prompt** 20 — 질문을 설계하는가 · v1
C	**Context** 25 — 맥락을 패키징하는가 · v2
H	**Harness** 20 — 환경(지식·도구·룰)을 설계하는가 · v3
L	**Loop** 20 — 루프가 **실제로 돌고** 산출물이 복리로 쌓이는가 · v4
X	**Interop & Governance** 15 — 이식·복구·보안·전수가 되는가

문서만 있으면 2점, 도구가 강제해야 3점, 시스템이 시스템을 고친 기록이 있어야 4점. 모든 점수는 아티팩트 증거 필수. **L(Loop)만은 최근 30일 가동 흔적**이 없으면 상한 2점 — akm.cmdspace.work 셀프 측정.

오늘부터 3단계

양이 아니라 연결이다.

STEP 1 · 오늘

.md 한 편

30분

마크다운 에디터 설치 → 논문 1편의 연구노트를 .md로.

STEP 2 · 이번 주

Zotero 10편

2시간

논문 10편 정리 → 볼트와 연결. 여기서 하네스가 생긴다.

STEP 3 · 다음 주

한 섹션 실전

반나절

진행 중인 논문·보고서 한 섹션을 볼트 기반으로 AI와 함께.

노트 50개면 그래프가 보이기 시작합니다. 단, **약 200개를 넘으면 스키마와 인덱스가 필요**해집니다 — 그때가 LLM Wiki로 갈 시점.

연구자의 하네스는 이미 여러분 손에 있습니다.

여러분이 읽고 쓰고 정리해온 모든 것 — 그게 재료입니다.
AI가 여러분을 분류하기 전에, **먼저 측정하세요.**

오늘 자료 labs.cmdspace.work/kird-0902 덱 · 링크 모음

AKM akm.cmdspace.work 셀프 측정

Starter Kit github.com/johnfkoo951 cmds-vault · cmds-llm-wiki

System Files system.cmdspace.work

더배러 뉴스레터

NAIS 해커톤 9/7 마감 ai4scikorea.org