

AI LITERACY FOR CLINICIANS & RESEARCHERS

Prompt, Context and Harness

Seoul National University Bundang Hospital
Cardiovascular Center · Ki-Hyun Jeon, M.D.

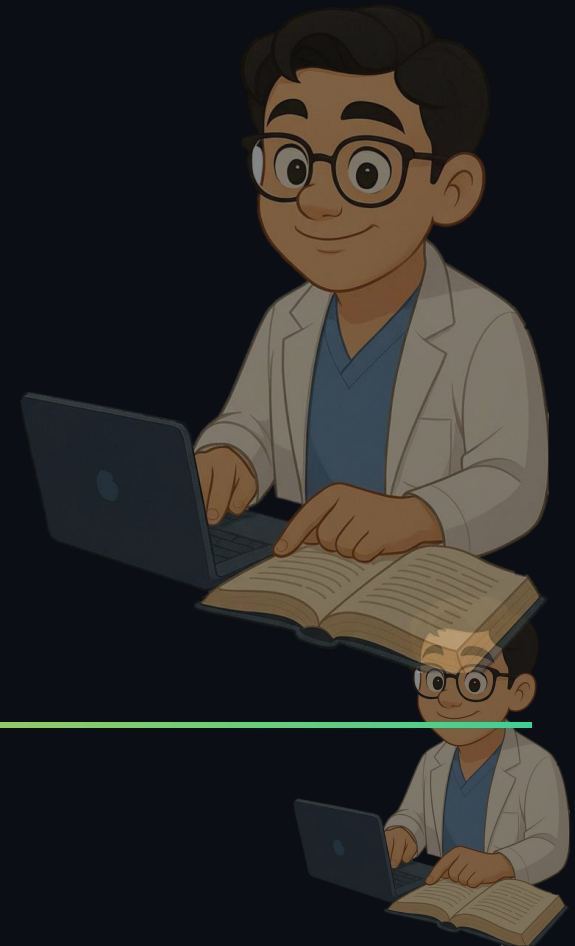
2026. 08. 대한부인종양학회



GUIDE TO BETTER ANSWERS

좋은 답변을 얻기 위한 방법

LLM 과의 대화에서 최고의 결과를 이끌어내는 핵심 전략



좋은 답변을 얻어내기 위한 방법

LLM 응용이 단일 호출에서 다중 턴 · 도구 사용 · 장기 자율 작업으로 진화하면서 “엔지니어링의 단위” 자체가 커졌다.



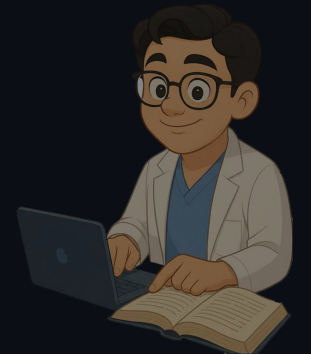
2022 - 2024

PROMPT ERA

단일 호출 최적화

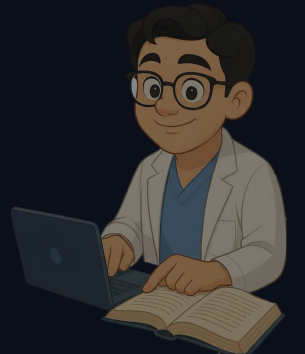
Single-turn classification · generation

한계 — 대화 · 도구 · 장기 작업에
서 무너짐



좋은 답변을 얻어내기 위한 방법

LLM 응용이 단일 호출에서 다중 턴 · 도구 사용 · 장기 자율 작업으로 진화하면서 “엔지니어링의 단위” 자체가 커졌다.



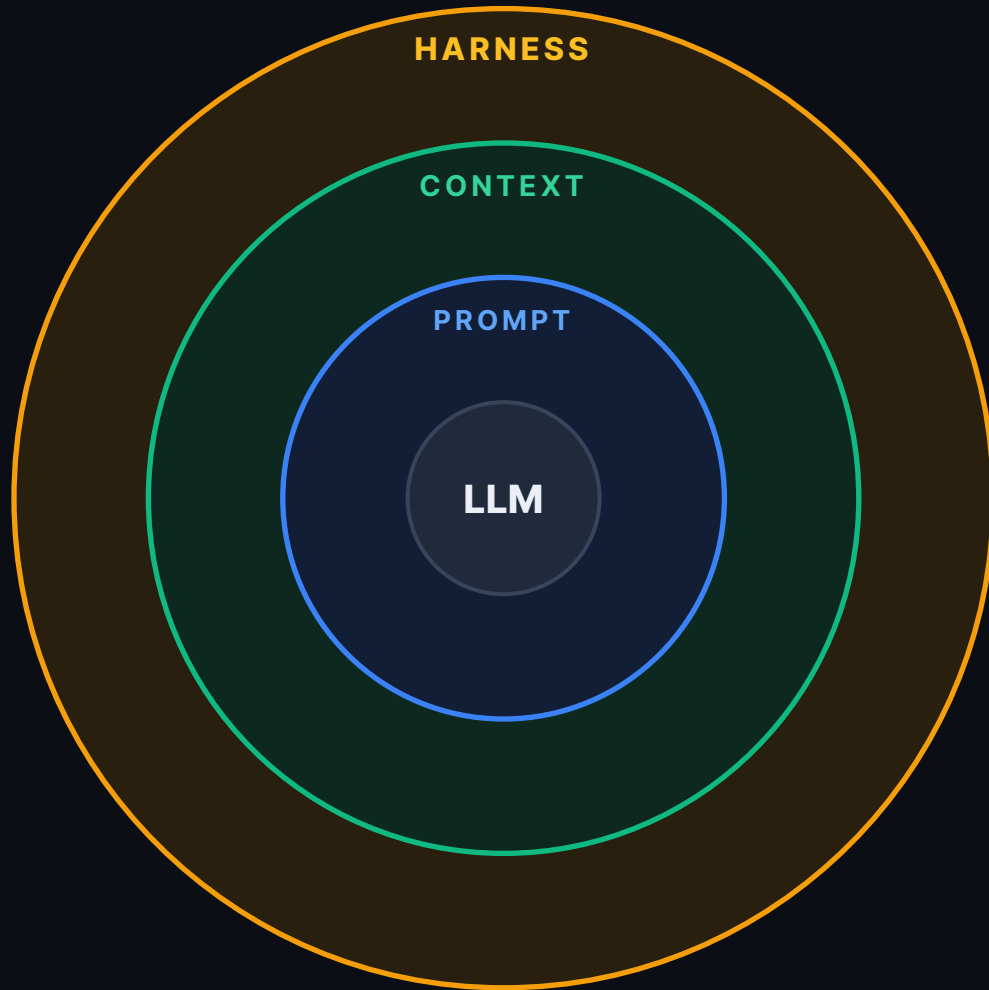
좋은 답변을 얻어내기 위한 방법

LLM 응용이 단일 호출에서 다중 턴 · 도구 사용 · 장기 자율 작업으로 진화하면서 “엔지니어링의 단위” 자체가 커졌다.



동심원 모델 — Prompt $\subset$ Context $\subset$ Harness

각 층이 다루는 것



LLM

모델 가중치 — 추론 · 생성을 수행하는 본체

Prompt

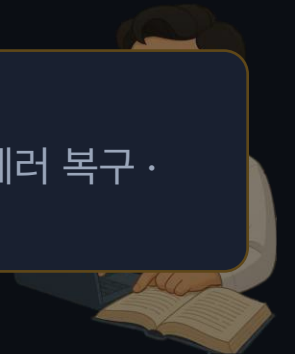
단일 메시지의 표현 · 구조 — 모델이 이번 턴에 받는 지시문 자체

Context

윈도우 전체의 구성 — 시스템 프롬프트 · 도구 정의 · RAG · 메모리 · 이력

Harness

모델을 둘러싼 모든 인프라 — 루프 · 영속성 · 검증 · 에러 복구 · 거버넌스



LINGUISTIC CONTROL

1st Wave: Prompt

01

Prompt Engineering 의 한계 (Limitations)



필요한 정보가 컨텍스트에 없으면 답할 수 없음

RAG · prior 검사 · 코드베이스 · 사용자 이력이 없으면 추측 · 환각.



다중 턴 / 세션 작업에서 정보가 휘발

Context rot — 대화가 길어질수록 정확도 저하.



“더 긴 프롬프트” 가 곧 해법은 아니다

모든 문제를 긴 프롬프트로 풀면 오히려 성능이 떨어진다.



Production 신뢰성 보장 어려움

단일 호출 결과의 양성 편향 · 검증 부재 · 업로드 데이터의 런타임 제한.



INFORMATION CONTROL

2nd Wave: Context

02

Context Engineering — 윈도우 전체의 큐레이션

Anthropic의 정의 · 2025-09-29

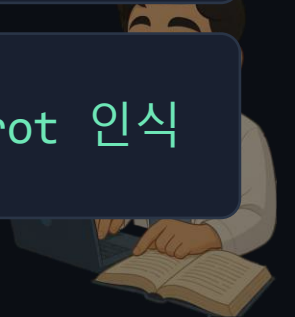
“프롬프트 엔지니어링의 자연스러운 진화. LLM 추론 시 사용 가능한 토큰 전체를 큐레이팅 · 유지하는 기법의 집합.”

관점의 전환

어떤 단어로 물을까 → 어떤 정보의 묶음을 보여줄까 단일 메시지 → 윈도우 전체

프롬프트 = 문자열 → 컨텍스트 = 시스템의 출력 static template → dynamic

정보가 많을수록 좋다 → 토큰은 한정 자원 context rot 인식



■ CONTEXT · WHAT GOES IN

Context — 윈도우에 들어가는 6 가지

“컨텍스트” 는 단순히 사용자 메시지가 아니다. 모델이 이번 호출에서 보는 토큰 전부를 가리킨다.



01

System instructions

역할 · 제약 · 출력 스키마 · 안전 규칙. 호출마다 최상단에 들어가는 영구 지시문.



02

Tool definitions (MCP)

어떤 함수 · API 를 호출할 수 있는지의 스펙. 이름 · 파라미터 · 반환 타입.



03

Retrieved data (RAG)

벡터 · 키워드 검색 · DB query 결과. 외부 지식이 컨텍스트로 들어오는 통로.



04

Message history

이전 턴의 user / assistant / tool result. 대화 · 에이전트 trajectory.



05

Memory / notes

세션 외부에 저장된 영속 정보. plan 파일 · progress 파일 · 사용자 프로필.

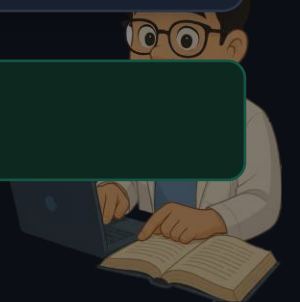


06

User message

이번 턴에 사용자가 보낸 입력 자체. 가장 작은 부분이지만 트리거.

TAKEAWAY 이 모두를 어떻게 채우고 비울지 결정하는 것이 곧 컨텍스트 엔지니어링의 핵심 작업이다.



SKILLS & MCP

Skill and MCP

02

Skill이란?

LLM이 외부 도구·API를 호출할 수 있도록
미리 정의된 **기능 단위**

예: 웹 검색, 이메일 발송, 데이터베이스 조회, 파일 읽기 — AI가 세상과 상호작용하는 손발



■ In Hospital · Agent = 인턴

LLM = 갓 들어온 인턴



의학 지식(knowledge)은 이미 다 갖췄습니다.
하지만 병원에 와서 막상 일하려면 —
지식만으로는 부족합니다. 실제로 **'할 줄'** 알아야 합니다.

Agent = 목표를 받아 '일하는 주체'. 단, 실전 능력은 이제부터 갖춘다.



핵심 지식을 갖춘 인턴 → 다음은 **'술기(Skill)'**를 익혀야 한다.



■ HOSPITAL MODEL · Skill = 술기

LLM+Skill = Agent 인턴이 익혀야 할 '술기'



L-tube insertion

L-tube 삽입



Foley catheter insertion

Foley 삽입

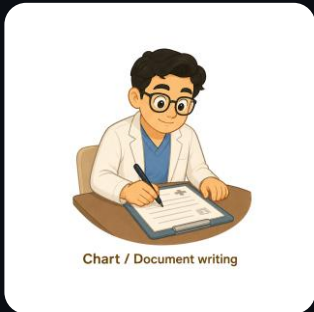
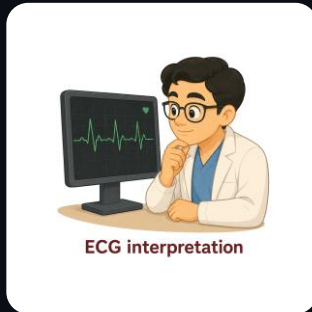


Chart / Document writing

Chart 작성



ECG interpretation

ECG 판독

지식을 '실행 가능한 능력'으로 바꾸는 것.
인턴은 다음을 직접 '할 줄' 알아야 합니다:
L-tube·Foley 삽입, chart 작성, ECG 판독 ...

LLM에서는 이것이 '**Skill**' 패키지
— 특정 작업을 수행하도록 묶어둔 능력.



핵심 지식(Knowledge) + 술기(Skill) → **AGENT**, 비로소 실제로 일할 수 있다.



`.skill`

Skill 안에 무엇이 있고, 어떻게 만들어지는가

폴더 구조

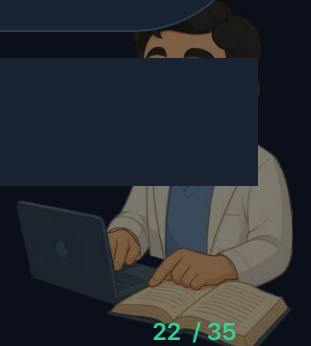
clinical-research-harness/

└ SKILL.md	# 언제 · 어떻게 실행할지 정의
└ references/	# 문서, 스키마, 예시
└ scripts/	# 실행 코드, 헬퍼
└ assets/	# 템플릿, 픽스처

SKILL Workflow

- 1 패턴 식별**
반복되는 워크플로를 발견하고 패키징 대상으로 인식.
- 2 SKILL.md 작성**
설명, 사용 시점, 단계를 기술.
- 3 배포**
플러그인, 저장소, 조직 라이브러리로 공유.
- 4 트리거 시 자동 재사용**
설명이 일치하면 모델이 자동으로 호출.

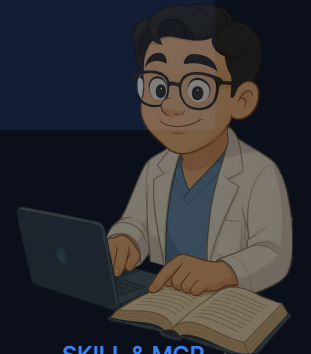
핵심 포인트 SKILL.md 앞부분의 트리거 설명이 핵심 — 모델은 이를 읽고 수동 import 없이 언제 실행할지 스스로 판단한다.



MCP란?

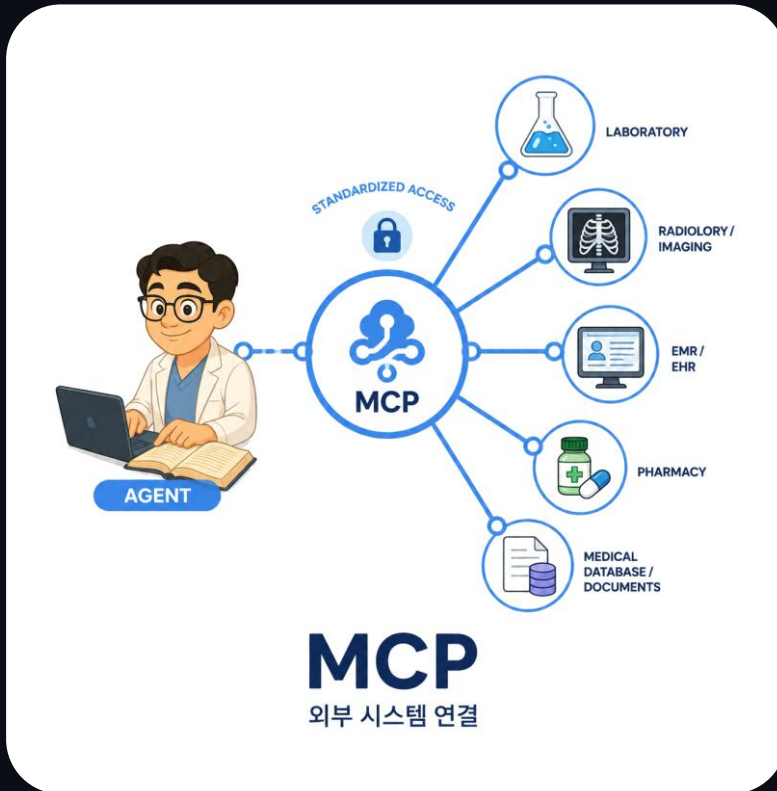
AI가 다양한 Skill을 일관된 방식으로 연결·실행하는 표준 통신 규약

Model Context Protocol — 어떤 LLM이든 동일한 방식으로 Skill을 호출할 수 있게 하는 **Anthropic** 주도의 개방형 표준



■ HOSPITAL MODEL · MCP(Model Context Protocol) = 검사실

MCP = 검사실과의 '표준 연결'



인턴이 모든 걸 혼자 하지 않습니다.

필요하면 검사실에 전화해 검사 절차를 묻고 의뢰합니다.

MCP는 **표준화된 방식(Standardized Access)**으로

외부 시스템 — 검사실·영상의학과·EMR·약제부 — 에 연결하는 통로.

혼자 다 하지 않는다. 필요할 때, 표준 규약으로 외부 조직에 연결.



핵심 Agent(인턴)가 MCP(검사실 연결)로 능력을 확장한다.



Model Context Protocol — 하나의 표준으로 어떤 도구든, 어떤 모델이든

LLM 호스트

Claude / GPT / Gemini

MCP 클라이언트

MCP 프로토콜로
서버를 탐색하고,
리소스를 나열하며,
도구를 호출

MCP 서버

PubMed Server

리소스: 논문 · 도구: 검색

MCP 서버

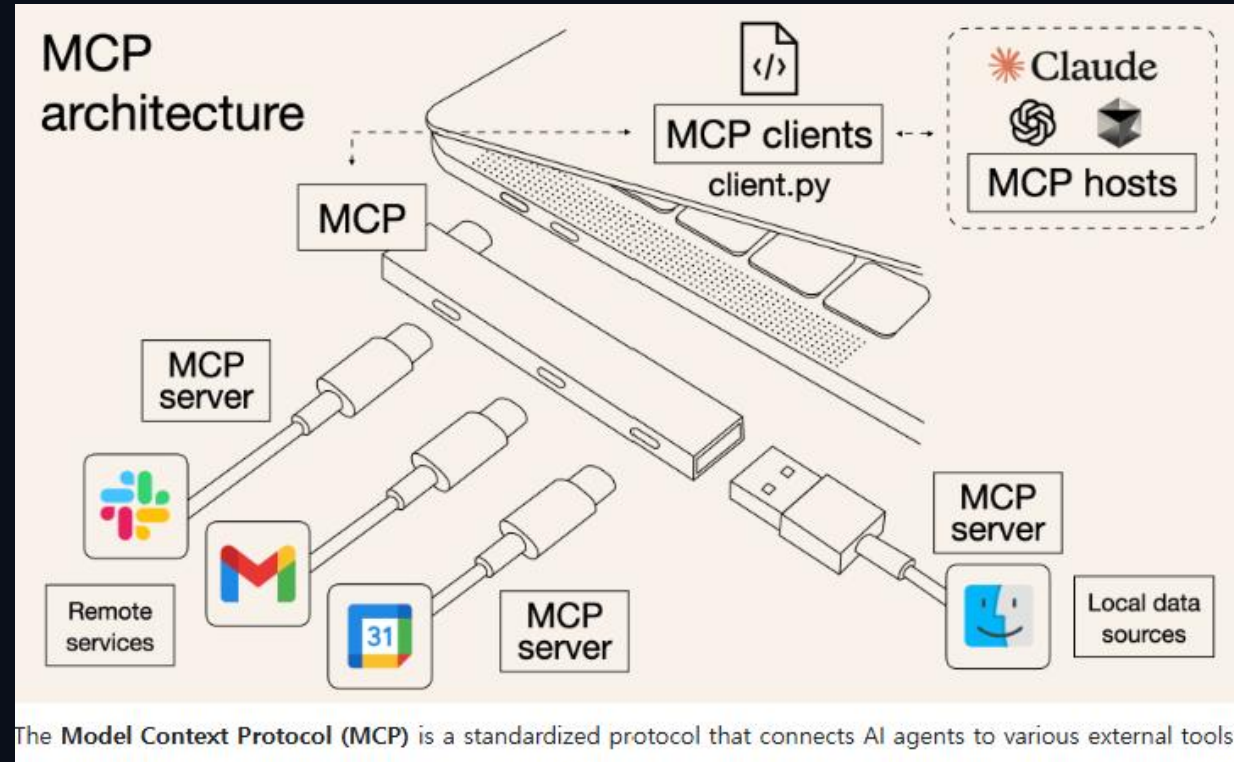
GitHub Server

리소스: 저장소 · 도구: PR 관리

MCP 서버

File System Server

리소스: 파일 · 도구: 읽기/쓰기



핵심 포인트 MCP 없이는 앱마다 별도 연동($N \times M$). MCP가 있으면 하나의 프로토콜로 모든 클라이언트가 모든 서버에 연결($N+M$).

MCP

EXECUTION & VERIFICATION CONTROL

3rd Wave: Harness

03

하네스란 무엇인가?

하네스(harness)'는 본래 말에 씌우는 마구 — 힘을 한 방향으로 모으고 통제하는 장치입니다.
LLM 하네스는 강력하지만 제멋대로인 AI의 능력을 정해진 절차·정책 안에 묶어 안전하게 일하게 만듭니다.



절차를 강제한다

단계(Phase)와 게이트를 정해, AI가 임의로 건너뛰지 못하게 한다.



사람을 끼워 넣는다

중요한 결정마다 연구자의 명시적 승인을 받고서 다음으로 넘어간다.

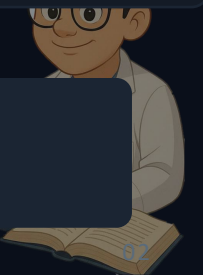


정책을 코드로 박는다

인용 근거·개인정보·통계 보고 규칙을 우회 불가능한 가드레일로 강제.

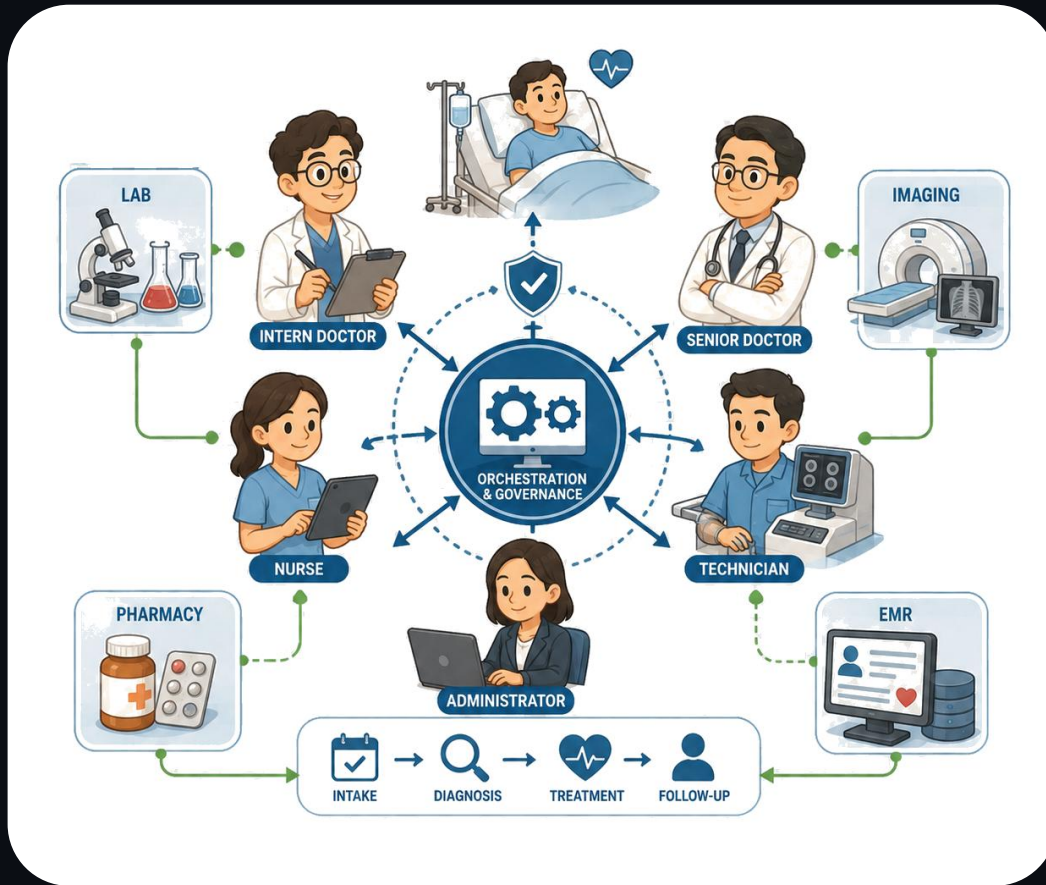


핵심: AI를 더 똑똑하게 만드는 게 아니라, **신뢰할 수 있게** 만드는 장치다.



■ HARNESS MODEL · Harness = 병원 전체

Harness = 병원 전체

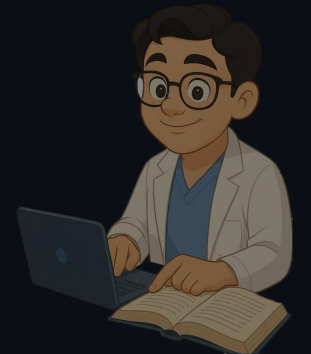


인턴 한 명이 아닙니다.

다양한 술기를 가진 여러 의사·간호사·기사·행정직
(여러 Agent)과 여러 검사실 같은 조직(MCP)이 함께.
이 모두를 총괄(Orchestration)·규제(Governance)하여
'환자 치료'라는 최종 목표를 이루는 것 = 병원 = Harness.



핵심 여러 Agent + 여러 MCP를 총괄·규제 → '환자 치료' 목표 달성. 그것이 Harness.



02 AI AGENTS ACROSS THE RESEARCH LIFECYCLE

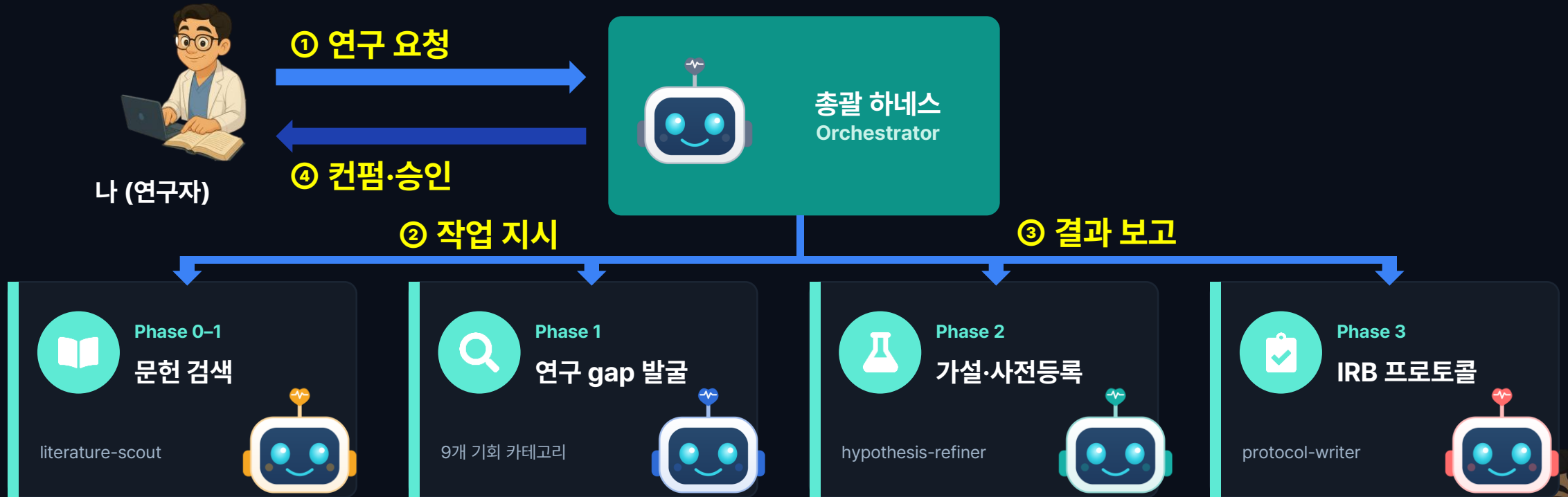
의학연구 전 과정의 AI 에이전트, 그리고 이를 총괄하는 하네스



각 단계는 전문 에이전트가 수행 — 하네스(orchestrator)가 전체를 라우팅·검증·연결한다. **경계는 IRB 승인 시점**, 두 하네스는 공유 workspace로 핸드오프.

02 AI AGENTS ACROSS THE RESEARCH LIFECYCLE

의학연구 전 과정의 AI 에이전트, 그리고 이를 총괄하는 하네스



나(연구자)가 요청하면 → 총괄 하네스가 각 전문 하네스에 작업을 지시·결과를 종합해 → 다시 나에게 컨펌을 받는 구조.

02 AI AGENTS ACROSS THE RESEARCH LIFECYCLE

연구계획 → 분석·논문작성, 하네스 인수인계

연구계획 하네스

Research Planning Harness

Phase 0-1 · 문헌 검색

Phase 1 · 연구 gap 발굴

Phase 2 · 가설·사전등록

Phase 3 · IRB 프로토콜



두 하네스의 경계

인수인계

공유 workspace 핸드오프

핸드오프 산출물

- 잠긴 사전등록
- IRB 승인서
- 데이터 명세
- 분석 계획

분석·논문작성 하네스

Analysis & Manuscript Harness

Phase 4 · 데이터 검정

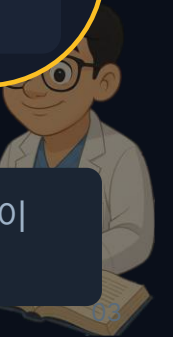
Phase 5 · 통계 분석

Phase 6 · 논문 초안

Phase 7 · 자체 동료검토



정확한 인수인계가 workflow를 완성한다 — 잠긴 사전등록·데이터 명세·IRB 승인을 넘기면, 분석·논문작성 하네스가 **사전등록대로 분석**해 p-hacking 없이 결론에 도달한다.



02 AI AGENTS ACROSS THE RESEARCH LIFECYCLE

의학연구 전 과정의 AI 에이전트, 그리고 이를 총괄하는 하네스



분석·논문작성 하네스가 각 전문 에이전트를 지시·종합 — 데이터 검정→통계 분석→논문 초안→자체 동료검토를 거쳐, 완성된 논문 최종본을 연구자(나)에게 전달한다.

https://github.com

/revfactory/harness

revfactory

Overview

Repositories286

Projects

Packages

Stars93



Minho Hwang

revfactory · he/him

Follow

Improving the world better

680 followers · 9 following

@kakao

South Korea

revfactory@gmail.com

rev.minho

in/hwang-minho

http://revf.tistory.com

Achievements



/Aperivue/medsci-skills

Yoojin-nam

Overview

Repositories7

Projects

Packages

Stars7



Yoojin Nam

Yoojin-nam · he/him

Unfollow

Radiologist & medical AI researcher

12 followers · 4 following

eugene8998@gmail.com

https://scholar.google.com/citations?user=HDrzyXAAAAAJ&hl=ko

https://orcid.org/0000-0001-8565-1360

Achievements



/JeonKH81/clinical-research-harness

JeonKH81

Overview

Repositories13

Projects

Packages

Stars6



DoctorGPT

JeonKH81

Edit profile

Interventional Cardiologist AI Researcher

Medical AI communicator

9 followers · 4 following

Seoul National University Bundang Hospital

kihyun.jeon

@DoctorGPT516



Structural heart disease screening

revfactory (Minho Hwang)

github.com/revfactory

Gemini에게 물어보기

revfactory

Type to search

Overview Repositories 288 Projects Packages Stars 93



Minho Hwang

revfactory · he/him

Follow

Improving the world better

638 followers · 9 following

@kakao

South Korea

revfactory@gmail.com

rev.minho

in/hwang-minho

http://revf.tistory.com

revfactory / README.md

반갑습니다! 🙌

저는 카카오 AI Native 전략 팀의 리더 황민호(robin.hwang) 입니다. 2013년부터 카카오에 재직하며 Daum 검색 서비스, 광고 플랫폼 Moment, 오픈소스 관리 플랫폼 OLIVE 등 주요 프로젝트 개발에 참여했습니다. 현재는 AI Agent와 같은 최신 AI 트렌드와 기술을 선제적으로 탐구하고 구현하여 카카오의 AI 전략 수립에 기여하고 있습니다.

새로운 기술을 배우고 최신 산업 동향을 파악하는 것을 즐깁니다. 또한, 오픈소스 프로젝트에 기여하고 다른 개발자들과 협력하여 어려운 문제를 해결하는 데에도 적극적으로 참여하고 있습니다.

Pinned

harness-100 Public

1k 373

harness Public

A meta-skill that designs domain-specific agent teams, defines specialized agents, and generates the skills they use.

HTML 7.1k 970

1,244 contributions in the last year

2026

2025

2024

2023

2022

2021

Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun

Mon

Wed

Fri

Learn how we count contributions

Less More

@koi2026 @chatgpt-agit

SNUBH Cardiovascular Center

Prompt · Context · Harness

CLINICAL RESEARCH HARNESS

SNUBH Cardiovascular Center

Clinical Research Harness



Thank you.

for your kind attention



EMAIL

imcardio@gmail.com



TEL

+82 31 787 7018



YOUTUBE

@DoctorGPT516



FACEBOOK

kihyun.jeon



GITHUB

github.com/JeonKH81



NOTION

imcardio.notion.site/ai-literacy-2026