Part 0 오리엔테이션 — AI 와 함께 연구한다는 것

AI 산업보건 석사논문

1주차

이 Part 에서 답할 질문

- 생성형 AI 는 연구의 무엇을 바꾸었고 무엇을 바꾸지 않았는가
- 어떤 일을 AI 에 맡길 수 있고 어떤 일은 맡길 수 없는가 — 그 기준은 무엇인가
- 이 과목은 열여섯 주 동안 무엇을 하고, 무엇을 평가하는가

0. 문제로 시작합니다

한 대학원생이 논문 초고를 가져왔습니다. 검색식은 AI 가 짰고, 통계 코드도 AI 가 짰고, 방법 절과 고찰도 AI 가 초안을 썼습니다. 학생은 전부 읽고 고쳤다고 합니다. 표는 깔끔하고 문장은 매끈합니다.

그런데 방법 절에 "SPSS 26.0 을 사용하였다"고 되어 있습니다. 학생은 R 을 썼습니다. 검색식의 MeSH 용어 하나는 실존하지 않습니다. 회귀표의 참조범주는 학생이 의도한 것과 반대입니다. 세 오류 모두 **형식이 완전해서** 읽는 동안 아무도 멈추지 않았습니다. 이 초고는 어디서부터 잘못됐습니까.

1. 무엇이 바뀌었는가 — 1.1 수행 비용은 낮아졌고 책임은 그대로입니다

생성형 AI 는 검색식·코드·초안을 즉시 만듭니다. 예전에 연구자를 가르던 "구현 능력"의 희소성이 사라졌습니다. 그러나 잘못된 검색식으로 누락된 문헌, 조용히 삭제된 결측치, 실재하지 않는 인용의 책임은 전부 연구자에게 있고, "AI 가 그렇게 했다"는 어떤 국면에서도 항변이 되지 않습니다.

1. 무엇이 바뀌었는가 — 1.2 실패의 모양이 바뀌었습니다

예전	지금
못 해서 멈춘다 — 코드가 안 돌아가고, 문장이 안 써진다	**완료되었으나 틀린 채** 제출된다
오류가 형식으로 드러난다	오류가 형식으로 드러나지 않는다 — 가짜 MeSH 는 진짜처럼 생겼고, 결측을 지운 회귀표도 정상으로 출력된다
형식을 검사하면 방어가 된다	형식 검사는 방어가 되지 않는다

1. 무엇이 바뀌었는가 — 1.2 실패의 모양이 바뀌었습니다

"돌아간다"와 "맞다"는 다릅니다

형식이 완전하다는 것은 내용이 맞다는 뜻이 아닙니다

1. 무엇이 바뀌었는가 — 1.3 그래서 남는 능력

구현 능력이 보편화될수록 희소해지는 것은 **"무엇이 옳은지 아는 능력"** 과 **"틀렸음을 알아채는 능력"** 입니다. 건강근로자효과를 떠올리는 역학 직관, 어떤 노출기준이 현행인지 아는 실무 지식, 이 답이 현장에서 위험한지 판별하는 감각 — AI 시대에 가치가 떨어지는 것이 아니라 유일하게 대체 불가능한 것으로 남습니다.

2. 기본 구조 — 위임의 기준과 수준 — 2.1 검증 비대칭

어떤 과업을 AI 에 맡길지는 "AI 가 잘하는가"가 아니라 **"틀렸을 때 내가 알아챌 수 있는가"** 로 정합니다. AI 가 가장 유창한 영역(사실 진술·인용·요약)이 정확히 검증이 가장 어려운 영역이므로, 능력을 기준으로 삼으면 반대 방향으로 위임하게 됩니다.

과업	AI 가 잘하는가	틀리면 알아챌 수 있는가	위임
통계 코드	예	예 — 참값을 심은 합성자료로 검증	가능
고시의 노출기준 값	예 (유창하게)	**아니오** — 원문을 보지 않으면	불가 (원문을 붙이면 가능)
검증된 척도의 문항	예 (유창하게)	**아니오** — 한 글자 바뀌어도 형식은 완전	불가
방법 절 초안	예	사실 목록이 먼저 있어야만	목록 뒤에만

2. 기본 구조 — 위임의 기준과 수준 — 2.2 자율성 수준 L0~L4

위임 여부는 "잘하는가"가 아니라 "틀렸을 때 내가 알아챌 수 있는가"가 정한다

L0	미사용	타인 원고 심사 · 검증된 척도 문항
L1	보조 — 전부 검토	방법·결과 문장화
L2	초안 — 전면 재검토	서론·고찰 · 검색식 · 계획서 문장화
L3	위임 — 표본 검수	코드 · 추출 · 선별 · 표 정리
L4	자동 — 테스트 통과 시 채택	논리 검사 코드 · 합성자료 생성

2. 기본 구조 — 위임의 기준과 수준 — 2.2 자율성 수준 L0~L4

수준	뜻	검증
L0	쓰지 않는다	—
L1	보조 — 산출물 전부를 사람이 검토	100%
L2	초안 — AI 가 만들고 사람이 전면 재검토	100%, 원문·사실 대조
L3	위임 — 실행 뒤 표본 검수	20~30% + 표적 검수
L4	자동 — 테스트 통과 시 채택	사람이 만든 테스트

한 연구 안에서 과업마다 수준이 다릅니다. "AI 를 활용하였다"는 그래서 아무 정보도 담지 않습니다.

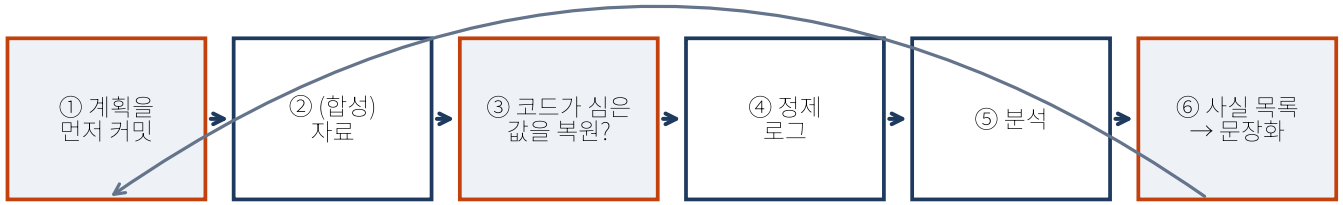
2. 기본 구조 — 위임의 기준과 수준 — 2.3 세 원칙

원칙	내용
검증 비대칭	알아챌 수 없는 것은 맡기지 않는다
근거 접지	사실이 필요한 과업에는 원문을 붙이고 "이 문서에서만 답하고 없으면 없다고 하라"
기록	모델·버전·프롬프트·시점·검수 비율·오류율. 기록하지 않은 프롬프트는 폐기한 원자료와 같다

2. 기본 구조 — 위임의 기준과 수준 — 2.4 해서는 안 되는 것

근로자 식별 정보·건강정보를 외부 서비스에 입력 · 사업장 영업비밀 입력 · **심사 의뢰받은 타인 원고 입력** · 검증하지 않은 인용 · **수행하지 않은 분석의 기술** · AI 를 저자로 · 원자료 없이 결과 표 생성 · 사후 가설을 사전 가설로.

3. 이 과목은 어떻게 굴러가는가 — 3.1 다섯 번 도는 하나의 루프



같은 루프를 자료만 바꿔 다섯 번 돈다 — 방법 다섯 개가 아니라 습관 하나



3. 이 과목은 어떻게 굴러가는가 — 3.1 다섯 번 도는 하나의 루프

논문 유형 다섯 가지를 차례로 실습합니다. 자료와 통계는 다르지만 뼈대는 같습니다 — 계획을 먼저 커밋하고, 참값을 심은 자료로 코드를 검증하고, 정제 로그를 남기고, 사실 목록에서 문장을 만듭니다. 방법 다섯 개를 배우는 것이 아니라 **검증하는 습관 하나를 다섯 번** 몸에 붙이는 것입니다.

Part	주	유형	자료
I	2–3	2차자료 연관 분석	근로환경조사 (hitlpaper)
II	4–5	문헌고찰·메타분석	출판 메타분석 재현
III	6–8	설문조사	각자 설문지 → 페르소나 응답
IV	9–11	노출평가	상황별 가상 측정자료
V	12–14	AI 개발	예측모형 · 미니 RAG
VI	15–16	학위논문 계획서 발표	다섯 중 하나

3. 이 과목은 어떻게 굴러가는가 — 3.2 매주 세 시간

구간	내용
개념 강의 30~40분 (Part 첫 주)	그 유형이 무엇이고 어떻게 하는지. 끝까지 한 번 보여 준다
제출물 진단 → 핀포인트 (둘째 주부터)	지난 제출물에서 교수가 붙인 오류 태그의 분포 → 상위 1~2개
판정 투표	산출물 둘을 띄우고 "어느 쪽이 검증 가능한가" 1차 투표 → 팀 토론 → 2차 투표
스튜디오 90분	실습. 순회하며 묻는 것은 하나 — **"이게 틀렸다면 어떻게 알아채나"**
오늘 잡은 오류 5분	한 명 1개씩

AI 튜터는 없습니다. AI 는 여러분의 연구 도구이지 교사가 아닙니다.

3. 이 과목은 어떻게 굴러가는가 — 3.3 무엇을 평가하는가

모든 제출물 머리에 **AI 사용 기술 4줄** — 어느 과업에 · 어느 수준으로 · 무엇으로 검증했고 · 오류율은(숫자). 없으면 받지 않습니다. 코드·계획서·프롬프트 기록은 개인 저장소에 두고 커밋 해시를 적습니다.

제출	배점 (제안)
P0 위임 배정표	5
P1~P5 결과 보고	각 13
P6 계획서·발표·심사	30

3. 이 과목은 어떻게 굴러가는가 — 3.3 무엇을 평가하는가

결과가 유의한가는 점수와 무관합니다

과정을 검증할 수 있는가를 채점합니다

4. 시범 — 위임 배정표 한 장

교대근무–수면장애 설문연구의 공정 10단계를 배정한 예입니다.

#	단계	배정	수준	틀렸을 때 무엇으로 알아채는가
1	연구 질문 확정	인간 배타	L0	—
2	측정도구 **선정**	인간 배타	L0	원 출처 원문과 글자 단위 대조. AI 가 재생한 문항이면 **알아챌 방법이 없다** → 원문 복사로만
3	표본·모집 경로	인간 배타	L0	현장 확인 — 누가 배포·회수하는가
4	IRB 문서 문장화	공동	L2	사실 목록과 1:1 대조. 목록에 없는 문장 = 창작
5	합성자료 생성	AI 위임	L3	큰 n 에서 심은 OR 복원
6	정제 코드	AI 위임	L3	규칙별 전후 n 출력

4. 시범 — 위임 배정표 한 장

#	단계	배정	수준	틀렸을 때 무엇으로 알아채는가
7	분석 코드	공동	L3	합성자료 복원 · 참조범주 로그 · R 재현
8	결과 표 정리	AI 위임 (원 출력에서만)	L3	전 셀 ↔ 출력 대조
9	서론·고찰 초고	공동	L2	인용 실존 DOI + 수치 인용 원문 대조
10	방법·결과 문장화	공동	L1	문장별 사실 목록 번호 대조, 수치 전수

2번과 4번에 "알아챌 방법이 없다"고 쓸 수 있는 학생이 이 표를 이해한 학생입니다. 빈칸이 답인 경우가 있습니다.

5. 주의점과 흔한 오류

오류	무엇이 잘못인가	대신
"AI 가 잘하는 일을 맡긴다"	유창한 영역 = 검증 어려운 영역	틀리면 알아챌 수 있는 일을 맡긴다
배정만 고치고 "알아채는 법" 칸을 비움	검증 없는 위임	그 칸이 표의 전부
"AI 를 활용하였고 충분히 검토하였다"	정보 없음	과업별 수준 · 검수 비율 · 오류율(숫자)
제목만 주고 논문 요약을 시킴	창작물	원문을 붙인다
코드가 돌아가니 맞다	돌아가지만 틀린다	합성자료로 검증
프롬프트를 기록하지 않음	사후 복원 불가	첫 주부터 log.csv

6. AI 와 함께 — 이 과목에서의 약속

하는 것	하지 않는 것
프롬프트를 파일로 쓰고 기록한다 (`00_공통/프롬프트_기록양식.md`)	대화창에 즉흥으로 치고 잊는다
원문을 붙이고 "없으면 없다고 하라"	기억에 의존한 사실 질의
근거 문장 병기를 요구한다	자유 서술 요약
산출물을 채택·미채택 모두 보관한다	채택본만 남긴다
자기 원고는 AI 에 비판시킨다	**타인 원고를 AI 에 넣는다**

프롬프트의 여섯 요소와 안티패턴은 `00_공통/프롬프트_원칙.md` .

6-1. 약어

이 과목의 약어(L0~L4 · PECO · OR · DAG · SEG · GM · UCL · RAG · AUC …)는 `00_공통/용어집.md` 에 전부 풀어 두었습니다. 낯선 낱말이 나오면 먼저 거기서 찾으십시오.

7. 실습 — 위임 배정표

순서	하는 일
1	워크숍 「위임 배정표 부검」 (팀, 45분) — 결함이 심어진 배정표를 고친다
2	자기 관심 주제(없으면 설문연구 전형)의 공정 10단계를 세 영역에 배정하고 "틀리면 어떻게 알아채는가"를 쓴다
3	개인 저장소를 만들고 `prompts/log.csv` 를 시작한다
제출	P0 위임 배정표 (루브릭 `30_실습.md`)

8. 되돌아보기

- 내 배정표에서 "알아챌 방법이 없다"고 쓴 단계는 몇 개인가
- 오늘 AI 산출물 중 내가 검증할 수 없었던 것은 무엇인가
- 다음 Part 로 가져갈 것: **"AI 답을 보기 전에 내 답을 먼저 건다"**

가져갈 다섯 줄

1. AI 는 수행 비용을 낮췄지만 **책임 구조는 바꾸지 않았다**
2. 실패는 "못 해서"가 아니라 **"완료되었으나 틀린 채"** 온다. 형식 검사는 방어가 아니다
3. 위임의 기준은 **"틀렸을 때 내가 알아챌 수 있는가"** — 과업마다 다른 수준(L0~L4)을 붙인다
4. 원문을 붙이고, 기록하고, 타인 원고는 넣지 않는다
5. 이 과목은 **검증 루프 하나를 다섯 번** 돈다. 채점은 검증 가능성이다