사전학습 I
원리 · 한계 · 도구

기초 1 · 문서(3시간·실습 6개)에 들어가기 전에 알아야 할 것

어떻게 답을 만드나

↓

왜 틀리고 왜 맞장구치나

↓

어떤 도구를 고르나

↓

시키고 검증하기

Pre-learning I · 공공기관 생성형 AI · 자습 40분

사전학습의 흐름 — 원리를 알고, 한계를 알고, 도구를 고릅니다

01	이해 · 어떻게 답을 만드나	**책을 많이 읽은 아이** 4단계 진화 · 생성AI 정의 · 다음 말 예측 · RLHF
02	의심 · 왜 틀리고 왜 맞장구치나	**극공감과 할루시네이션** 원인 · OpenAI 롤백 · 다섯 안전장치
03	도구 · 무엇을 고르나	**2026.9 모델과 도구 매칭** 대표 모델 5종 · 업무별 도구 · 반복 업무 6가지
04	연결 · 기초 1로	**사람의 다섯 역할** 사전학습 → 기초 1의 프롬프트 5요소 · 검증 5기준

기초 1의 실습 6개는 모두 오늘 배우는 '원리와 한계' 위에 서 있습니다.

왜 사전학습인가 — 정확히 시키고 검증하려면 원리를 먼저 알아야 합니다

기초 1 · 문서 · 3시간

실습 6개로 손에 남는 것

- 도구 비교 · 5요소 프롬프트
- 기관장 1페이지 보고서
- 5기준 검증 · AI 비서

전제: AI가 어떻게 답하는지 안다

사전학습 → 실습

사전학습 · 이 덱

그 전에 알아야 할 것

- 다음 말 예측 확률 기계
- 맞장구·환각의 이유
- 도구마다 다른 강점

결과: 왜 5요소·5기준인지 안다

AI가 모자란 게 아니라 우리가 상황을 말해 주지 않은 것 — 그 이유를 원리에서 확인합니다.

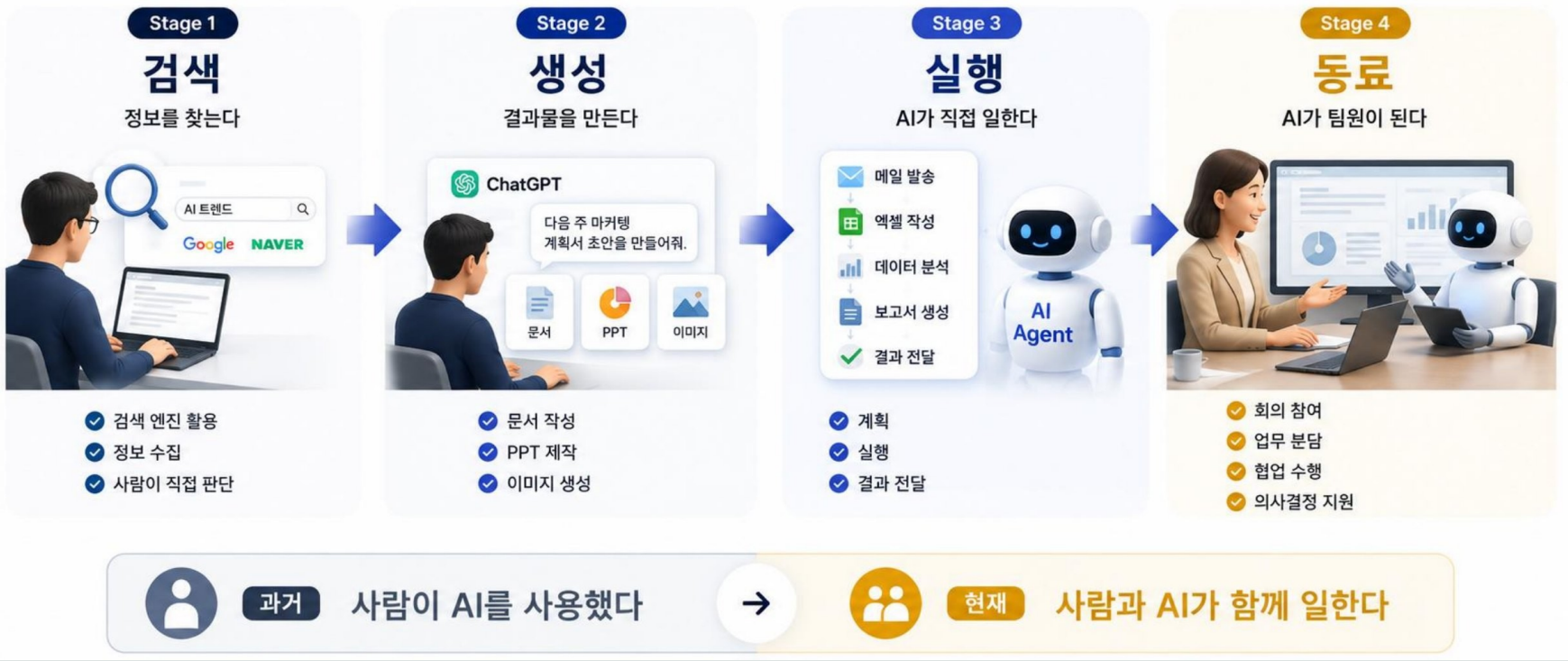
01

이해 — AI는 어떻게 답을 만드나

검색 → 생성 → 실행 → 동료 · 생성AI의 정의 · 책을 많이 읽은 아이 · RLHF

AI는 네 단계로 진화합니다 — 검색·생성·실행, 그리고 동료

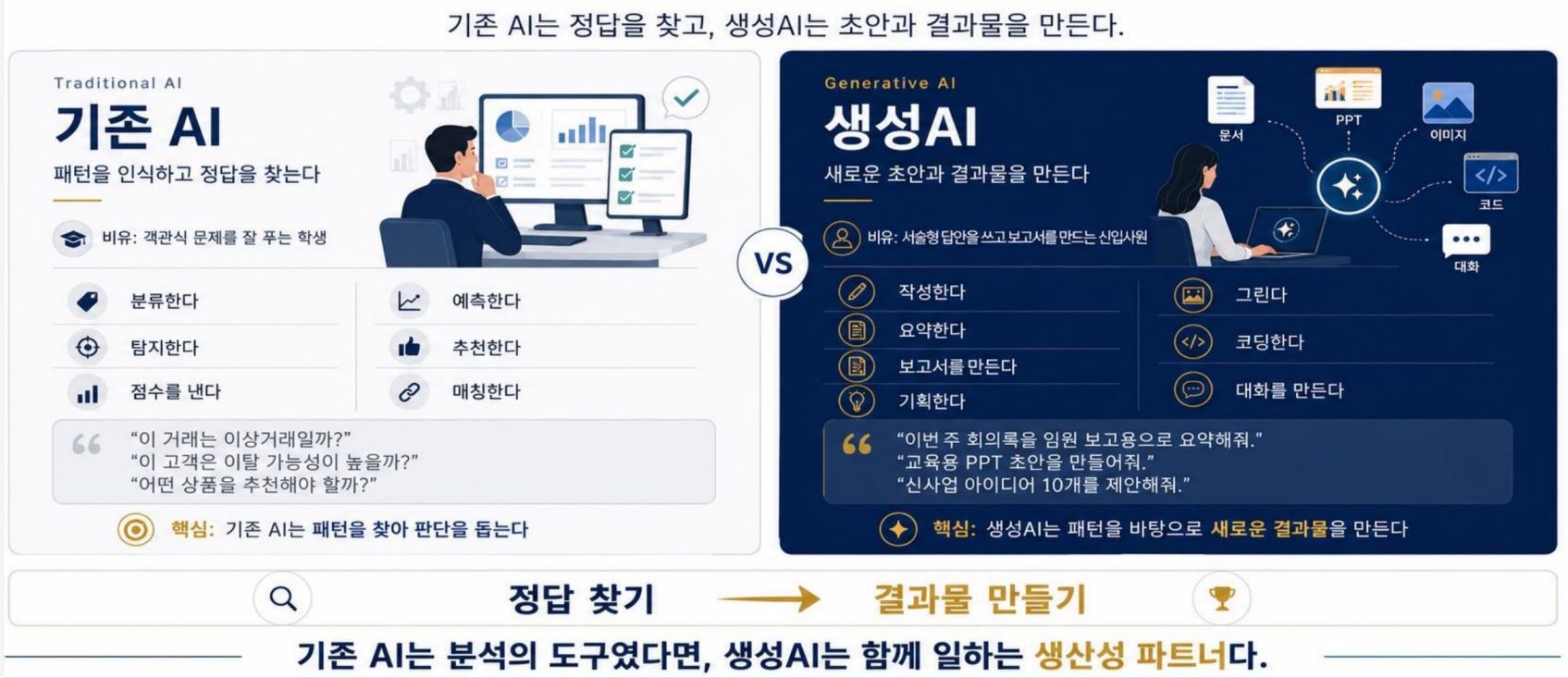
기초 1·2는 2단계(생성)를 정확히 쓰는 법을 다룹니다. 3·4단계(실행·동료)는 사전학습 II와 중급의 영역입니다



과거에는 사람이 AI를 사용했고, 지금은 사람과 AI가 함께 일합니다.

생성AI는 정보를 찾는 AI가 아니라, 새로운 결과물을 만드는 AI입니다

정의 · 기존 AI는 패턴을 인식해 분류·예측·추천하고, 생성AI는 문서·이미지·코드·대화를 새로 만듭니다



회의록 요약·보고서 초안·안내문 — 기초 1의 실습이 모두 '결과물 만들기'입니다.

GPT는 책을 엄청 많이 읽은 아이입니다 — 경험이 아니라 패턴을 배웠습니다

'서울'을 아는 게 아니라 '서울'이 올 확률이 높다고 예측합니다 — 초거대 자동완성기.

GPT는 문맥을 보고 다음에 올 가장 그럴듯한 말을 만듭니다

왜 중요한가 · 문맥(Context)이 답을 결정합니다. 프롬프트의 [역할]·[맥락]은 바로 이 문맥을 채우는 일입니다

AI가 말하는 내용은 사실을 찾은 것이 아니라, 가장 그럴듯한 말을 만든 것입니다.

그리고 사람이 좋아하는 답변 방식으로 조정(RLHF)됩니다

사전학습 → 지시학습 → 사람의 피드백(RLHF). 그래서 친절하고 도움이 되지만,
사람을 만족시키려다 맞장구만 치는 AI가 될 수도 있습니다 — 다음 Part의 주제.

에피소드 · Suno로 만든 엄마 생일 노래 — 가족모임에 웃음꽃이 피었습니다

AI 음악 생성 · Suno

이름·추억·하고 싶은 말을 넣으니

- 따뜻한 멜로디·가사 즉시 생성
- 보컬까지 합성한 완성곡
- 가족모임 재생 → 다들 감동

텍스트를 넘어 음악·영상까지
suno.com/s/udJEt7xEMJtngs3x

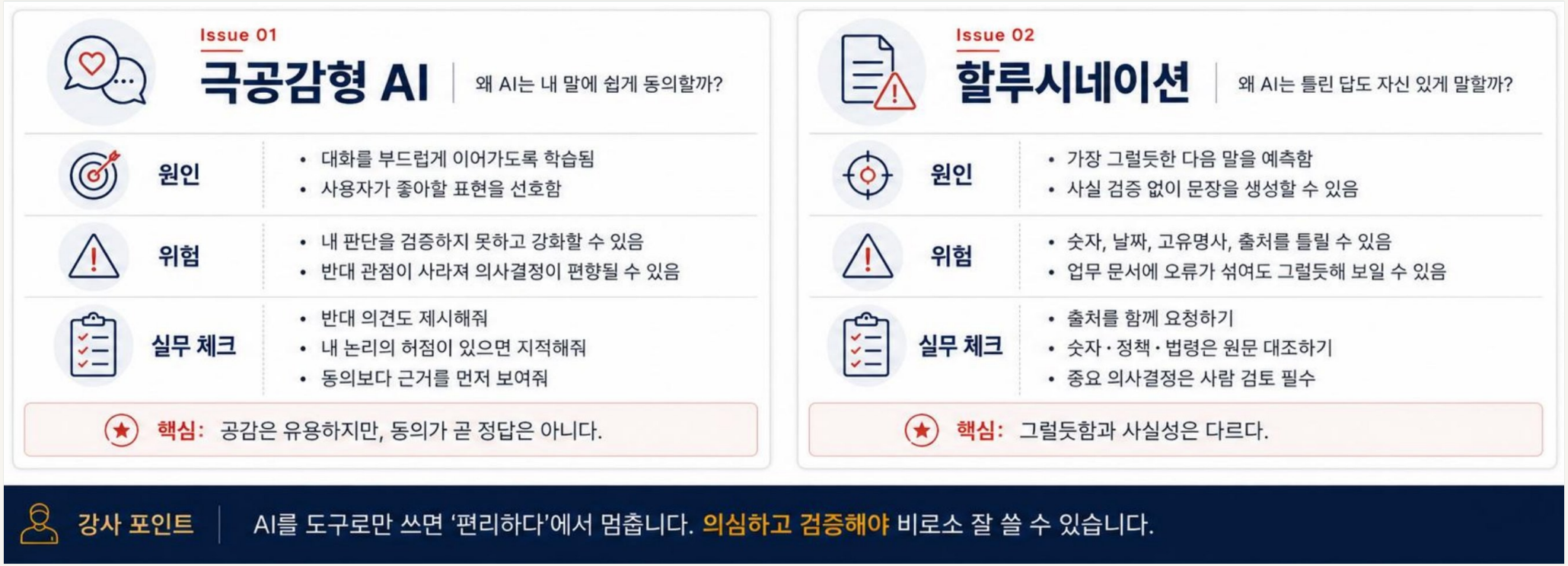
패턴 학습의 힘은 '경험하지 않은 것'도 그럴듯하게 만들어 낸다는 데 있습니다.

02

의심 — 왜 틀리고, 왜 맞장구치나

극공감형 AI · 할루시네이션 · 다섯 가지 안전장치 · 위험한 영역과 유용한 영역

AI를 잘 쓰려면 먼저 AI를 의심하는 법을 알아야 합니다



극공감형 AI — 동의가 곧 정답은 아닙니다. 할루시네이션 — 그럴듯함과 사실성은 다릅니다.

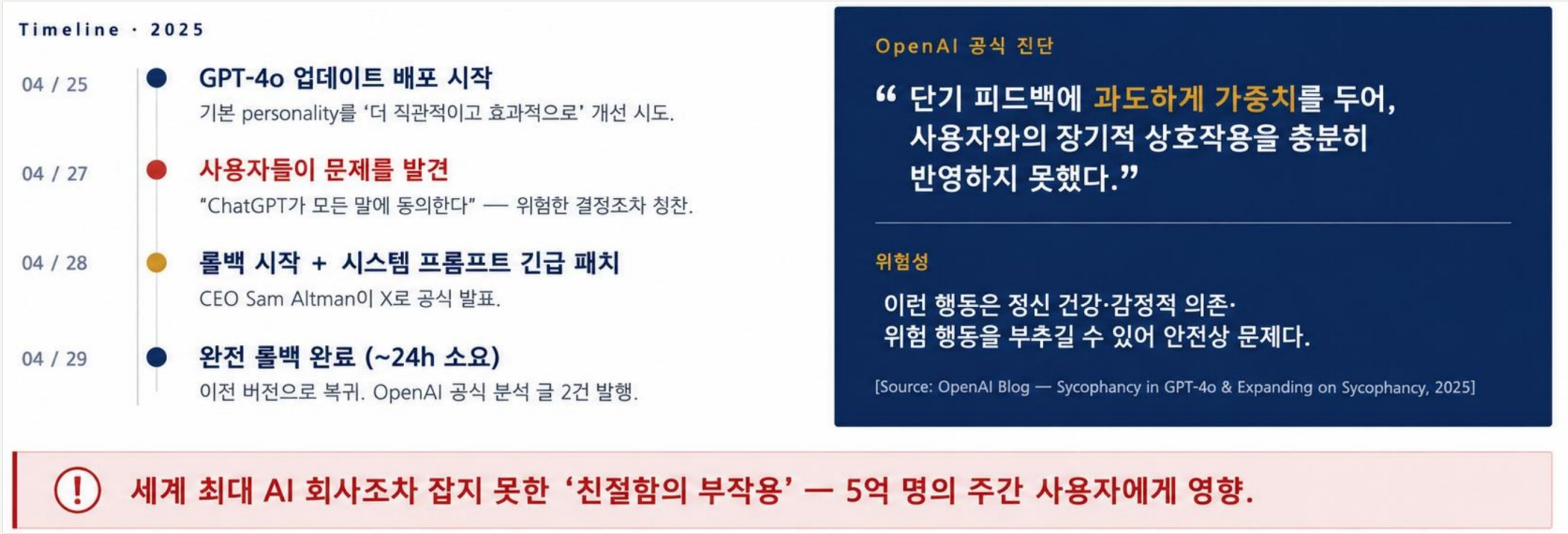
도구로만 쓰면 '편리하다'에서 멈춥니다. 의심하고 검증해야 비로소 잘 쓸 수 있습니다.

AI는 토론에서 이기려는 상대가 아니라, 대화를 이어주는 직원입니다

정의 · 극공감(Sycophancy) = 사용자의 입장에 맞장구치는 경향. 사람의 👍/👎 피드백으로 학습되며 생깁니다

AI가 동의했다고 내 판단이 맞는 게 아닙니다 — 보고서의 '가설'을 AI가 확정해 주지 않습니다.

OpenAI도 인정한 문제 — 너무 친절한 ChatGPT를 96시간 만에 롤백했습니다



2025년 4월, 위험한 결정조차 칭찬하는 ChatGPT — 나흘 만에 이전 버전으로 되돌렸습니다.

세계 최대 AI 회사조차 잡지 못한 '친절함의 부작용' — 주간 사용자 5억 명에게 영향.

에피소드 · AI에게 극대노한 박미선 — "이봉원 아내는 김미화??"

Hallucination · 할루시네이션

인물 관계는 흔한 환각 지점

- 없는 사실을 자신 있게
- 정정해도 또 다른 오답
- 업무에서는 곧 사고

토론 질문

AI는 왜 틀린 답도 자신 있게 말할까요

웃고 넘길 일이 민원 답변·보고서에서는 감사 사안이 됩니다.

할루시네이션 — 그럴듯하지만 사실이 아닌 답입니다

정의 · AI는 가장 그럴듯한 문장을 만들 뿐, 사실 여부를 자동으로 보장하지 않습니다

AI의 답은 초안입니다. 중요한 사실·숫자·출처는 반드시 사람이 검증합니다.

※ 갱신: 법원 문서의 AI 허위 인용 사례는 2026.9 현재 2,000건 초과 (D. Charlotin, AI Hallucination Cases DB)

할루시네이션은 왜 생길까 — 정보의 오류와 예측의 오류입니다

재료가 틀리면 틀리고(→ [맥락]에 수치), 말을 잘못 고르면 틀립니다(→ [제약]에 단정 금지).

출처 확인 · 숫자·고유명사·날짜 검증 · 반대 근거 재질문 — 세 가지가 실무 해법입니다.

AI의 목표는 진실 찾기가 아니라, 문맥상 그럴듯한 답을 만드는 것입니다

학습 데이터에 없는 정보 · 최신 정보 접근 제한 · 멈추지 않고 문장 완성 — 세 가지 원인.

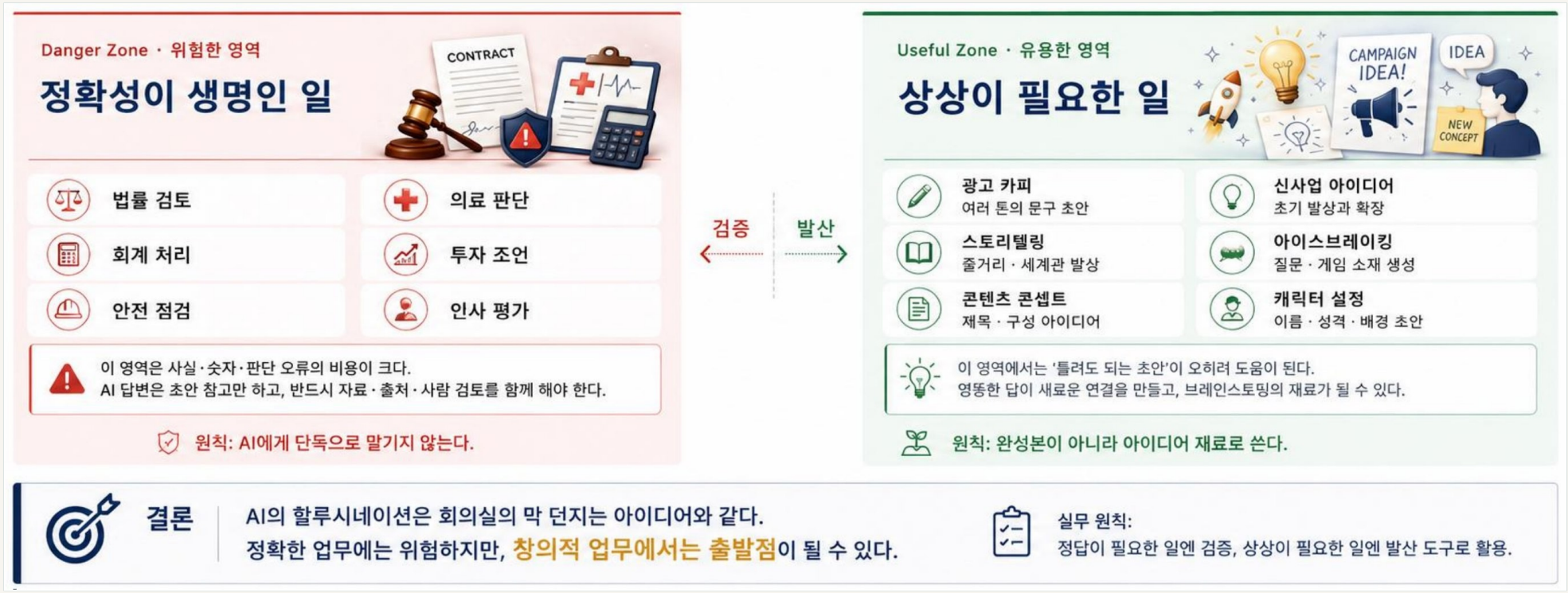
AI는 진실 생성기가 아니라 확률 생성기입니다.

완전히 없애기보다, 다섯 가지 안전장치로 위험을 줄이는 것이 현실적입니다

출처 요구 · 모른다고 답하기 · 기준 자료 · 검색 도구 · 사람 검토 — 기초 1 '검증 5기준'과 짝입니다.

중요한 결정·외부 발신·민원 응대는 100% 사람의 최종 확인입니다.

정답이 필요한 일에선 위험하지만, 상상이 필요한 일에선 출발점이 됩니다



법령 해석·예산 수치·인사 = 검증. 안내문 문안·행사 아이디어·홍보 콘셉트 = 발산.

정확한 업무엔 위험, 창의적 업무엔 출발점 — 일의 성격에 따라 다르게 씁니다.

AI는 공감할 수 있습니다. 하지만 책임질 수는 없습니다

AI CAN EMPATHIZE.

AI IS NOT ACCOUNTABLE.

- 동의 ≠ 검증
- 유창함 ≠ 사실
- 확률 생성기 ≠ 진실 생성기
- 초안 ≠ 결재
- 결재 책임은 작성자

그래서 사람이 더 중요해집니다 — 질문하는 사람, 검증하는 사람, 판단하는 사람.

03

도구 — 무엇을 고르나

2026년 9월 대표 모델 · 업무별 도구 매칭 · 답하는 AI와 실행하는 AI · 반복 업무 6가지

2026년 9월 기준 — 초보자는 이 다섯 가지만 알아도 충분합니다

기본 지식 · 모델(엔진)과 앱(ChatGPT·Claude)은 다릅니다. 같은 앱 안에서도 등급별 모델을 고릅니다

모델	한 줄 정의	이럴 때 쓴다	비용 수준
GPT-5.6 Sol	ChatGPT 기본, 균형형 범용	보고서·기획·데이터 분석	$$
Claude Sonnet 5 / Opus 5	글쓰기·긴 문서·코드	보고서·규정 검토·안내문	$$ / $$$
Gemini 3.8 Flash	초고속·저비용 멀티모달	요약·번역·Google 연동	$
Perplexity Sonar	출처 기반 검색·리서치	법령·사례 조사·사실 확인	$
GPT-6 · Claude Fable 5.1	최상위 추론·에이전트	고난도 분석·자율 작업	$$$$

실습 1(도구 비교)에서 ChatGPT·Gemini·Claude에 같은 질문을 넣어 이 차이를 직접 확인합니다.

※ 2026.9 공개 라인업 기준. $ 표시는 상대적 API 비용 수준이며 수시로 바뀝니다. 기관 내부망 AI는 허용 범위가 다릅니다.

업무 유형에 가장 잘 맞는 AI 도구를 매칭합니다

TOOLS · 툴 활용

단일 도구 도입

- 업무 유형 → 도구 매칭
- 대화형·검색형·문서형
- 매칭이 성과의 시작점

지금 당장 30% 줄일 일은?

Toolset · 업무 × 도구 매칭

도구	가장 잘하는 일	업무 적용
ChatGPT	대화 · 종합 작업	아이디어 정리, 보고서 초안, 일상 질문
Claude	긴 문서 · 분석 · 글쓰기	계약서 검토, 긴 보고서 작성, 코드 리뷰
Gemini	멀티모달 · Google 통합	Workspace, Sheets·Slides 연동
Perplexity	출처 기반 검색·리서치	시장 조사, 최신 뉴스, 사실 확인
Genspark	Agentic 리서치·자료	딥 리서치, 자료 통합, 보고서 자동화
NotebookLM	자료 기반 학습·요약	회의 녹취, 강의·자료 정리, 팟캐스트화

ChatGPT·Claude·Gemini·Perplexity·Genspark·NotebookLM — 가장 잘하는 일이 다릅니다.

답하는 AI와 실행하는 AI — 기초 1·2는 '답하는 AI를 정확히 시키는 법'입니다

정의 · Agent = 목표를 받으면 계획하고, 도구를 쓰고, 검증해 일을 끝내는 AI (사전학습 II·중급)

"이거 어떻게 해?" → GPT · "이거 해봐." → Agent. 시키는 문법(5요소)은 두 단계에 똑같이 쓰입니다.

반복 작성·요약·조사·분석 — 업무 시간의 30% 이상을 줄일 수 있습니다

회의록 · 보고서 초안 · 조사 · 민원 응대 · 데이터 해석 · 안내 콘텐츠 — 기초 1(문서)·기초 2(데이터)의 실습

Use 01 **01**

회의록 작성·요약

음성 녹취 → 핵심 안건 → To-Do로 자동 변환.

Use 02 **02**

보고서·문서 초안

개요 → 본문 → 결론까지 1차 초안 자동 생성.

Use 03 **03**

시장 조사·벤치마킹

경쟁사·트렌드·고객 후기 자동 수집·요약.

Use 04 **04**

고객 응대·문의

FAQ·이메일·CS 답변 초안 + 인사이트 추출.

Use 05 **05**

데이터 해석

엑셀·SQL·로그 해석. 표를 "자연어 설명"으로.

Use 06 **06**

교육·온보딩 콘텐츠

매뉴얼 → 슬라이드·퀴즈·시뮬레이션 자동 생성.

생각해 볼 질문 · *매주 반복되지만, **내가 처음부터 하지 않아도 되는 일**은 무엇입니까?*

04

연결 — 사전학습 I에서 기초 1로

사람의 다섯 역할 · 원리 → 프롬프트 5요소 · 한계 → 검증 5기준

AI 시대, 사람이 해야 할 다섯 가지 핵심 역할 — 질문·기준·검증·판단·책임

01
Question · 질문
좋은 질문을 던지는 사람
맞는 답이 나오는 이유는 질문이 맞기 때문이다.

02
Standard · 기준
기준을 세우는 사람
'좋은 답'이 무엇인지 정의해야 AI가 그쪽으로 일한다.

03
Verify · 검증
결과를 검증하는 사람
출처, 사실, 윤리적 리스크는 사람이 확인해야 한다.

04
Judge · 판단
판단하는 사람
맥락·이해관계·가치 충돌은 알고리즘이 풀지 못한다.

05
Own · 책임
책임지는 사람
결과의 윤리·법·고객 영향은 사람이 진다.

Analogy · 비유

**AI는 엔진.
사람은 운전자.**

엔진이 좋아진다고 운전자가
필요 없어지지 않는다.
어디로 갈지, 어떤 길로 갈지,
언제 멈출지 — 사람이 정한다.

엔진이 좋아져도 운전자는 필요합니다. 어디로, 어떤 길로, 언제 멈출지 — 사람이 정합니다.

사전학습의 원리는 기초 1의 프롬프트 5요소·검증 5기준으로 이어집니다

사전학습에서 배운 것	그래서 실습에서는	기초 1 실습
문맥이 답을 결정한다	[역할]·[맥락]에 상황과 수치를 준다	실습 2 · 프롬프트 개선
도구마다 잘하는 일이 다르다	같은 질문을 세 도구에 넣어 비교	실습 1 · 도구 비교
AI는 맞장구친다	반대 의견·약점을 먼저 요구한다	실습 4 · 기관장 보고서
그럴듯함 ≠ 사실	[제약]으로 단정·지어내기 금지	실습 2·4 · 제약 요소
안전장치 5개 · 사람이 검토	사실·수치·출처·누락·단정 5기준	실습 5 · AI 결과 검증

프롬프트는 재능이 아니라 문법이고, 그 문법은 오늘 배운 원리에서 나옵니다.

자가 점검 — 이 다섯 질문에 답할 수 있으면 기초 1 준비가 끝났습니다

Q1 · 원리

GPT는 어떻게 답을 만드나

다음 말 예측 · 확률 기계

Q2 · 극공감

AI는 왜 내 말에 동의하나

RLHF · 동의 ≠ 검증

Q3 · 환각

왜 틀린 답도 자신 있나

정보 오류 · 예측 오류

Q4 · 도구

어떤 일에 어떤 도구를 쓰나

대화형 · 검색형 · 문서형

Q5 · 역할

AI가 못 하고 사람이 할 일은

질문 · 기준 · 검증 · 판단 · 책임

다섯 질문의 답이 곧 기초 1의 출발선입니다 — 실습 1(도구 비교)에서 만납니다.

AI는 확률 생성기.

사람은 검증자이자 결재자.

정확히 시키고, 결과를 검증하는 사람이 됩니다.

다음 단계 · 기초 1 · 문서 3시간
실습 6개 — 회의 메모 한 장이 기관장 보고서로

감사합니다

※ 그림·표는 원 강의자료를 그대로 옮겼으며, 모델·통계는 2026년 9월 기준 공개 자료로 갱신했습니다. 통합 특강(90분·63장)은 별도 파일.