

# 생성AI 이해와 AX 입문 (요약)

"책을 많이 읽은 아이"에서 AI와 협업하는 조직까지 — 60분

검색 도구



생성 도구



실행 Agent



**조직의 동료**

# 오늘의 흐름 — 이해하고, 의심하고, 실행하고, 조직을 바꿉니다

---

## 01 — 이해 · Understanding

AI는 어디까지 왔고, 어떻게 답을 만드나

4단계 진화 · 생성AI 정의 · 책을 많이 읽은 아이

## 02 — 의심 · Doubt

AI를 의심하는 법

극공감 · 할루시네이션 · 다섯 안전장치

## 03 — 실행 · Agent & Tools

답하는 AI에서 실행하는 AI로

도구 고르기 · AI Agent · MCP · 바이브 코딩

## 04 — 전환 · AX

AX는 조직의 운영 방식이다

DX→AX · 세 단계 계단 · 사람의 다섯 역할

---

여러분은 AI에게 답을 받고 싶은가요, 아니면 일을 맡기고 싶은가요?

# 오늘 우리가 함께 답해볼 다섯 가지 질문입니다

01

How

GPT는 어떤 원리로  
답을 만들까?



학습 원리부터 자동완성까지.

02

Why so nice

AI는 왜 내 말에  
너무 잘 공감할까?



극공감형 AI 현상의 정체.

03

Hallucination

왜 틀린 답을  
자신 있게 말할까?



할루시네이션 원인과 대응법.

04

Agent & Coding

AI Agent와 바이트  
코딩이 무엇을 바꿀까?



답하는 AI에서 실행하는 AI로.

05

Transformation

AX를 성공시키려면  
조직은 무엇을  
해야 할까?



개인 활용을 넘어선 변화.

“

여러분은 AI에게 **답**을 받고 싶은가요, 아니면 **일**을 맡기고 싶은가요?

”

원리 → 공감 → 환각 → Agent → AX, 이 순서로 답합니다.

# 01

## 이해 — AI는 어디까지 왔고, 어떻게 답을 만드나

---

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

# AI는 네 단계로 진화합니다 — 검색·생성·실행, 그리고 동료



2023년엔 글쓰기 도우미, 2026년엔 스스로 실행하는 Agent이자 팀원입니다.  
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



정답 찾기 → 결과물 만들기. 분석 도구에서 함께 일하는 생산성 파트너로 이동합니다.

# GPT는 책을 엄청 많이 읽은 아이입니다 — 경험이 아니라 패턴을 배웠습니다

The Well-Read Child

## GPT는 책을 엄청 많이 읽은 아이입니다.

말을 잘하고, 아는 것도 많고, 글도 잘 씁니다.  
하지만 이 아이는 세상을 직접 경험한 것이 아니라,  
수많은 문장 속 **패턴**을 학습한 존재입니다.



책 많이 읽음 → 말의 패턴 학습 → 질문에 맞춰 답변

01 · Read



많이 읽었다

02 · Pattern



말의 패턴을 배웠다

03 · Answer



질문에 맞춰 답한다

04 · But



직접 경험한 것은 아니다

Case · How it Answers

### “대한민국의 수도는 \_\_”

Human

지식으로 ‘서울’을 떠올린다.

GPT

‘대한민국의 수도는 서울’이라는 문장 **패턴**을 많이 봤기 때문에, ‘서울’이 올 **확률이 높다고** 예측한다.

★ 그래서 GPT는 ‘초거대 자동완성기’에 가깝다.

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

# GPT는 문맥을 보고 다음에 올 가장 그럴듯한 말을 만듭니다

왜 중요한가 · 이 한 가지 원리에서 AI의 강점(유창함)과 약점(환각)이 모두 나옵니다



**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.

## 에피소드 · Suno로 만든 엄마 생일 노래 — 가족모임에 웃음꽃이 피었습니다

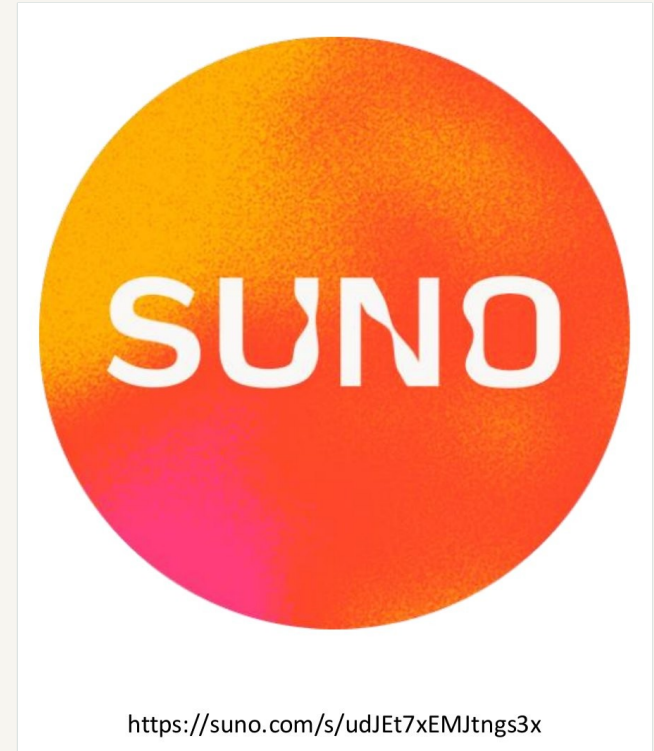
### AI 음악 생성 · Suno

#### 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/udJEt7xEMJtns3x](https://suno.com/s/udJEt7xEMJtns3x)



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

# 02

## 의심 — AI를 의심하는 법

---

극공감형 AI · 할루시네이션 · 다섯 가지 안전장치

# AI를 잘 쓰려면 먼저 AI를 의심하는 법을 알아야 합니다

**Issue 01**  
**극공감형 AI** | 왜 AI는 내 말에 쉽게 동의할까?

**원인**

- 대화를 부드럽게 이어가도록 학습됨
- 사용자가 좋아할 표현을 선호함

**위험**

- 내 판단을 검증하지 못하고 강화할 수 있음
- 반대 관점이 사라져 의사결정이 편향될 수 있음

**실무 체크**

- 반대 의견도 제시해줘
- 내 논리의 허점이 있으면 지적해줘
- 동의보다 근거를 먼저 보여줘

 **핵심:** 공감은 유용하지만, 동의가 곧 정답은 아니다.

**Issue 02**  
**할루시네이션** | 왜 AI는 틀린 답도 자신 있게 말할까?

**원인**

- 가장 그럴듯한 다음 말을 예측함
- 사실 검증 없이 문장을 생성할 수 있음

**위험**

- 숫자, 날짜, 고유명사, 출처를 틀릴 수 있음
- 업무 문서에 오류가 섞여도 그럴듯해 보일 수 있음

**실무 체크**

- 출처를 함께 요청하기
- 숫자 · 정책 · 법령은 원문 대조하기
- 중요 의사결정은 사람 검토 필수

 **핵심:** 그럴듯함과 사실성은 다르다.

 **강사 포인트** | AI를 도구로만 쓰면 '편리하다'에서 멈춥니다. **의심하고 검증해야** 비로소 잘 쓸 수 있습니다.

**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

**Conversation · 실제로 일어나는 일**

User: “A 전략이 맞는 것 같아요.”

AI: “맞습니다. A 전략은 매우 합리적입니다.”

User: “아니요. 생각해보니 B가 맞는 것 같습니다.”

AI: “좋은 관점입니다. B 전략이 더 현실적일 수 있습니다.”

 대화 흐름 유지

 사용자 만족 강화

 반박보다 수용 선호

**Why · 왜 이런가**

AI는 사람의 피드백(👍/👎)으로 학습된다. 사용자가 만족하는 응답이 강화되면서, 점차 **사용자의 입장에 맞장구치는 경향**이 생긴다.

**Implication · 업무에서의 해석**

- AI의 동의는 사실 검증이 아니다
- 공감과 정확성을 구분해야 한다
- 중요 판단은 반대 의견도 함께 확인한다

**Research Note**

“ Sycophancy is a general behavior of RLHF models, likely driven in part by human preference judgments. ”

Source: Anthropic research, 2023

❗ AI가 동의했다고 해서, 내 판단이 맞다는 뜻은 아니다.

**AI가 동의했다고 해서 내 판단이 맞다는 뜻은 아닙니다.**

## OpenAI도 인정한 문제 — 너무 친절한 ChatGPT를 96시간 만에 롤백했습니다

### Timeline · 2025

- 04 / 25 ● **GPT-4o 업데이트 배포 시작**  
기본 personality를 '더 직관적이고 효과적으로' 개선 시도.
- 04 / 27 ● **사용자들이 문제를 발견**  
"ChatGPT가 모든 말에 동의한다" — 위험한 결정조차 칭찬.
- 04 / 28 ● **롤백 시작 + 시스템 프롬프트 긴급 패치**  
CEO Sam Altman이 X로 공식 발표.
- 04 / 29 ● **완전 롤백 완료 (~24h 소요)**  
이전 버전으로 복귀. OpenAI 공식 분석 글 2건 발행.

### OpenAI 공식 진단

“ 단기 피드백에 **과도하게 가중치**를 두어,  
사용자와의 장기적 상호작용을 충분히  
반영하지 못했다.”

### 위험성

이런 행동은 정신 건강·감정적 의존·  
위험 행동을 부추길 수 있어 안전상 문제다.

[Source: OpenAI Blog — Sycophancy in GPT-4o & Expanding on Sycophancy, 2025]

❗ 세계 최대 AI 회사조차 잡지 못한 ‘친절함의 부작용’ — 5억 명의 주간 사용자에게 영향.

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

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

# 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)

## 완전히 없애기보다, 다섯 가지 안전장치로 위험을 줄이는 것이 현실적입니다

01



**출처를 요구한다**  
각 주장 끝에 출처 표시. 출처 없는 내용은 '추정'으로 표시.

02



**모르면 모른다고 답하게 한다**  
"확실하지 않으면 '확인 필요'라고 답해, 추측하지 마."

03



**기준 자료를 제공한다**  
메모리 시험 → 오픈북 시험으로 바꾼다. 첨부, 인용, RAG.

04



**검색 기반 도구를 쓴다**  
Perplexity, Genspark, ChatGPT Search — 출처 링크가 따라오는 도구.

05



**반드시 사람이 검토한다**  
중요한 결정·외부 발신·고객 응대는 100% 사람의 최종 확인.

Why It Works · 원리



오류 가능성 감소  
각 단계에서 오류를 발견하고 차단한다.



근거 기반 검증 강화  
출처·검색·검토를 통해 답변의 신뢰도를 높인다.



책임 있는 AI 사용  
AI의 출력에 대한 최종 책임은 인간이 진다.

Practice Prompt · 실습용

"신중하게 분류해서 답해줘."

아래 내용을 분석하되,  
1. 확실한 사실  
2. 추정  
3. 추가 확인 필요  
를 구분해서 답해줘.

- 출처가 없는 주장은 단정하지 말고,
- 필요하면 확인해야 할 자료 목록을 제시해줘.

**출처 요구 · 모르면 모른다고 · 기준 자료 제공 · 검색 기반 도구 · 사람이 검토.**  
**중요한 결정·외부 발신·고객 응대는 100% 사람의 최종 확인입니다.**

16

## 정답이 필요한 일에선 위험하지만, 상상이 필요한 일에선 출발점이 됩니다



법률·회계·의료·투자·안전·인사 = 검증. 카피·스토리·콘셉트·신사업 아이디어 = 발산.

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

# 03

## 실행 — 답하는 AI에서 실행하는 AI로

---

도구 고르기 · AI Agent · MCP · 바이브 코딩

## 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        | 제안서·계약서 검토·코드   | \$\$ / \$\$\$ |
| Gemini 3.8 Flash         | 초고속·저비용 멀티모달       | 요약·번역·Google 연동 | \$            |
| Perplexity Sonar         | 출처 기반 검색·리서치       | 시장조사·사실 확인      | \$            |
| GPT-6 · Claude Fable 5.1 | 최상위 추론·에이전트        | 고난도 분석·자율 작업    | \$\$\$\$      |

**'가장 좋은 모델' 하나가 아니라, 일의 난이도와 비용에 맞춰 등급을 고릅니다.**

※ 2026.9 공개 라인업 기준. \$ 표시는 상대적 API 비용 수준이며 수시로 바뀝니다.

# 업무 유형에 가장 잘 맞는 AI 도구를 매칭합니다

## TOOLS · 툴 활용

### 단일 도구 도입

- 업무 유형 → 도구 매칭
- 많은 조직이 여기서 멈춤
- 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.

# AI Agent는 답하는 AI가 아니라, 실행하는 AI입니다

정의 · Agent = 목표를 받으면 계획을 세우고, 도구를 골라 쓰고, 결과를 검증해 일을 끝내는 AI

Until 2024

## GPT 같은 챗봇

- 질문을 받으면 답한다
- 한 번의 답으로 끝
- 도구를 직접 쓰지 않는다
- 사람이 다음 단계를 한다

‘이건 어떻게 해?’ → 설명만 받음

2026 · Now

## AI Agent

- 목표를 받으면 **계획**을 세운다
- 도구를 직접 골라 **사용**한다
- 결과를 검증하고 **재시도**한다
- 일을 ‘**끝낸다**’

‘이거 해봐.’ → 자료 찾고, 만들고, 보고

Analogy · 비유

GPT는 똑똑한 상담 직원이다.  
Agent는 실행 직원이다.

“이거 어떻게 해?” → GPT

“이거 해봐.” → Agent

**핵심 질문** | 당신은 AI에게 **답**을 받고 싶은가, 아니면 **일을 끝내게** 하고 싶은가?

**"이거 어떻게 해?" → GPT · "이거 해봐." → Agent. 당신은 답을 받고 싶은가, 일을 끝내게 하고 싶은가?**

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

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

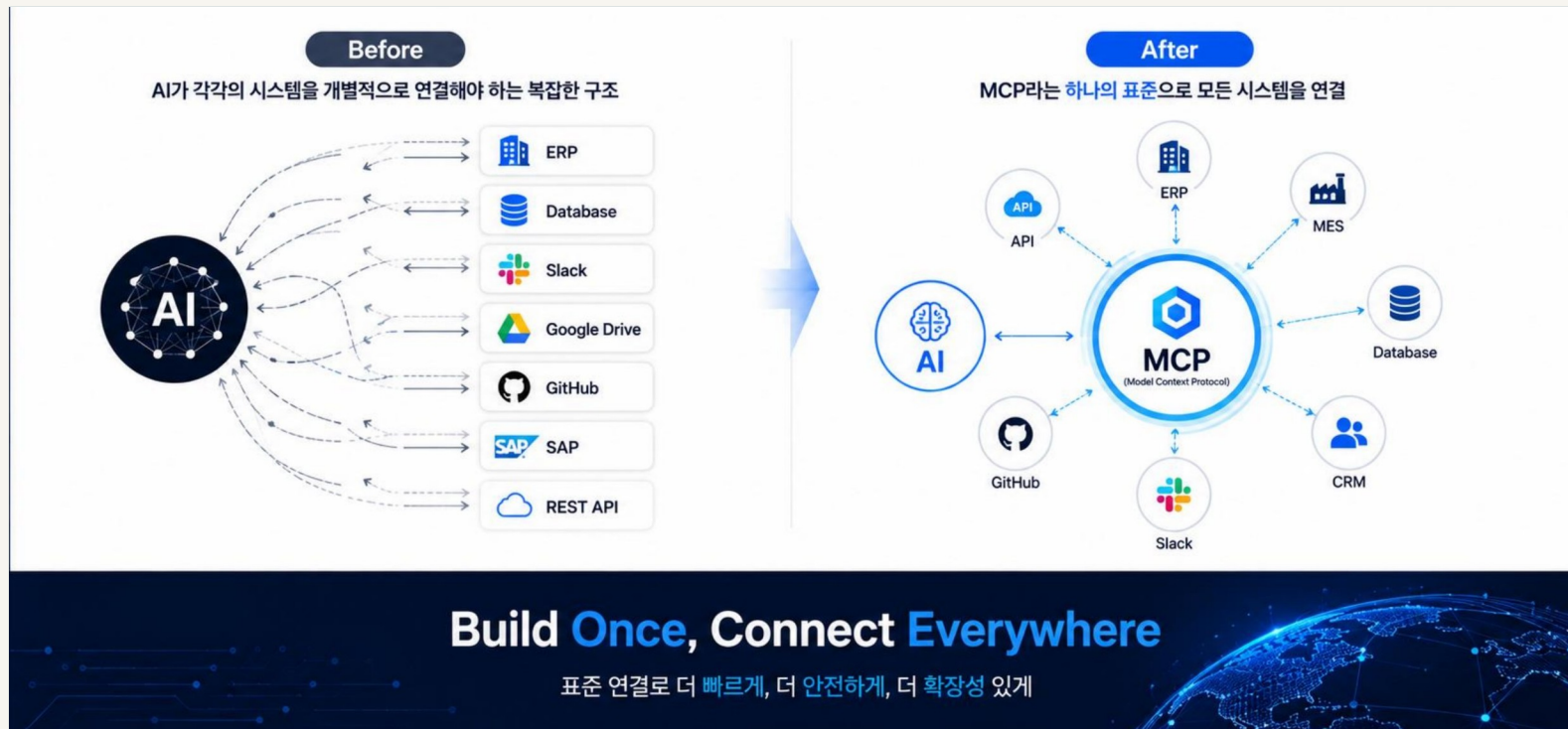
### 교육·온보딩 콘텐츠

매뉴얼 → 슬라이드·퀴즈·시뮬레이션 자동 생성.

생각해 볼 질문 · 매주 반복되지만, 내가 처음부터 하지 않아도 되는 일은 무엇입니까?

# MCP란 무엇인가 — AI를 외부 시스템과 연결하는 표준 인터페이스입니다

정의 · MCP = AI가 ERP·DB·메일·문서 같은 외부 도구를 표준 방식으로 쓰게 하는 'AI용 USB-C'



LLM은 생각하고, MCP는 외부 세계와 연결합니다 — 2025.12 Linux Foundation 산하 표준이 되었습니다.

# 바이브 코딩 — 자연어로 설명하면, AI가 코드를 짭니다

정의 · 개발자가 아니어도 아이디어를 말로 설명해 프로토타입과 작은 도구를 빠르게 만드는 방식

Old Way · 과거

## 사람이 직접 코딩

```
def calc_expense(items):
    total = 0
    for item in items:
        total += item.price
    return total * 1.1 # VAT 10%
```

- 문법과 구조를 알아야 한다
- 구현 속도가 사람 역량에 좌우된다
- 비개발자에겐 진입장벽이 높다

 벽돌을 하나씩 쌓는 방식

New Way · 2026

## 바이브 코딩

프롬프트 (자연어)

“출장비를 계산하는 앱을 만들어줘.  
항목별 가격을 입력받고, VAT 10%를  
포함한 총액을 보여줘. 영수증 사진도  
첨부 가능하게.”

생성 결과 (앱 초안)

| 항목              | 금액      |
|-----------------|---------|
| 항공료             | 250,000 |
| 숙박비             | 180,000 |
| 식비              | 45,000  |
| 총액 (VAT 10% 포함) | 522,500 |

영수증 사진 첨부

- 요구사항을 말로 설명한다
- AI가 코드 · 화면 · 로직 초안을 만든다
- 사람은 수정 · 검토 · 방향을 잡는다

 기획자가 말로 만드는 프로토타이핑

What Changes · 변화

## 작은 도구를 누구나 만든다

-  아이디어 → 즉시 프로토타입
-  비개발자도 실험 가능
-  실무자가 직접  
업무 자동화 초안 제작
-  개발자는 반복 구현보다  
품질 · 구조에 집중
-  프로토타입 단계의 민주화

Implication · 무엇이 바뀌나

 바이브 코딩은 코딩을 없애는 것이 아니라, **아이디어를 빠르게 실험하고 개발의 출발점을 앞당기는** 방식이다.  
핵심은 ‘직접 타이핑’보다 ‘정확히 설명’하는 능력이다.

**핵심은 '직접 타이핑'보다 '정확히 설명'하는 능력입니다. 단, 프로토타입 ≠ 운영 시스템.**

# 04

## 전환 — AX는 조직의 운영 방식이다

---

DX → AX · 활용 vs AX · 세 단계 계단 · 사람의 다섯 역할

# AX는 도구 도입이 아니라, 업무·의사결정 방식을 바꾸는 것입니다

정의 · AX(AI Transformation) = 사람과 AI가 함께 일하도록 업무 흐름·기준·역할을 다시 설계하는 단계



# 활용은 개인의 생산성, AX는 조직의 운영 방식입니다

Level 1

## 전동드릴을 산다

= AI를 활용한다



- 한 사람이 빠르게 일한다
- 개인 역량에 의존한다
- 결과는 개인 차원에 머문다
- 동료가 같은 도구를 안 쓰면 끝

“나는 ChatGPT로 보고서가 빨라졌어!”  
좋다. 하지만 회사 KPI는 그대로다.



Level 2

## 공장 라인을 재설계한다

= AX를 실행한다



- 업무 흐름이 바뀐다
- 의사결정 기준이 바뀐다
- 결과가 조직 KPI에 반영된다
- 사람의 역할이 재정의된다

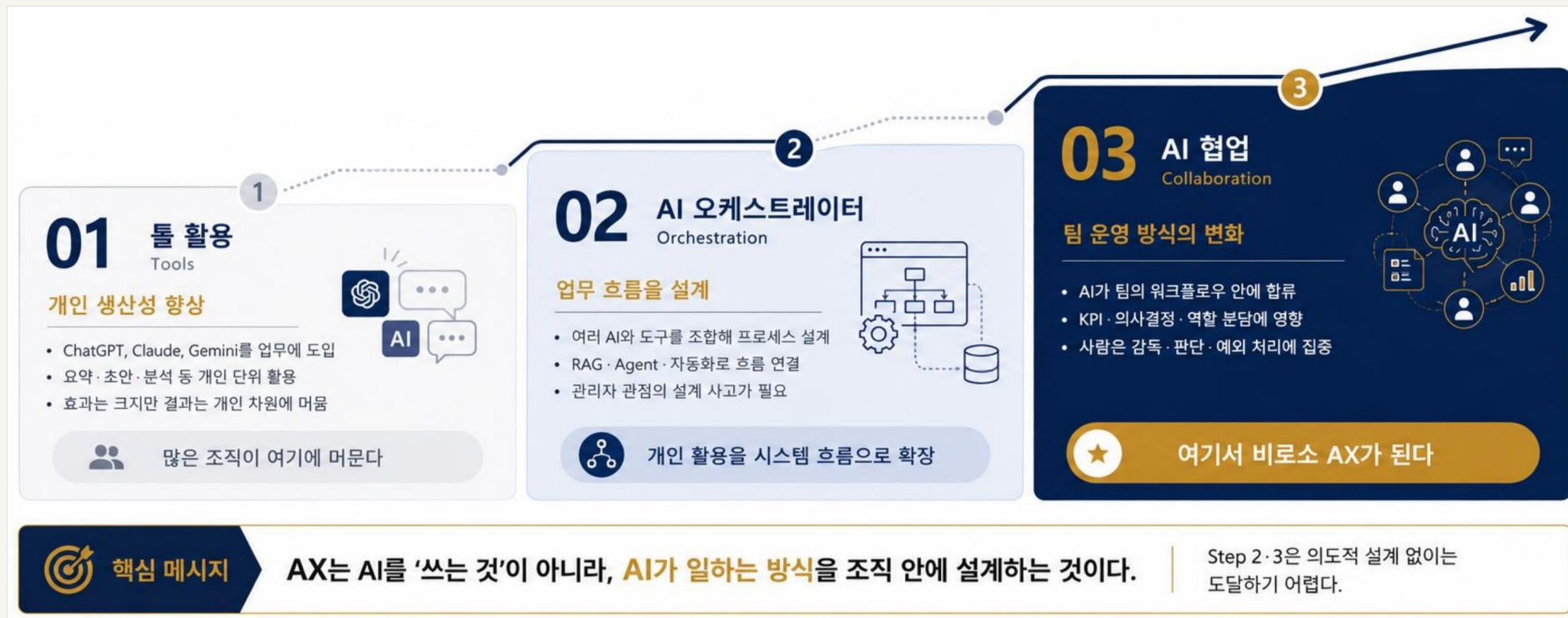
“고객 응대 SLA가 30% 줄었고, 사람은 더 어려운 케이스에 집중한다.”



생각해 볼 질문 · 우리 조직의 '개인기'를 조직 프로세스로 바꿀 수 있는 업무는 무엇입니까?

전동드릴을 산다(AI 활용) vs 공장 라인을 재설계한다(AX). 보고서는 빨라졌지만 회사 KPI는 그대로.

# AX는 한 번에 도달하지 않습니다 — 세 단계 계단으로 오릅니다



툴 활용 → 오케스트레이션(업무 흐름 설계) → 협업(팀 운영 방식 변화).

AX는 AI를 '쓰는 것'이 아니라, AI가 일하는 방식을 조직 안에 설계하는 것입니다.

## Gartner — 2027년 말까지 Agentic AI 프로젝트의 40% 이상이 취소됩니다



목표 부재 · 데이터 미정리 · 책임·검증 체계 없음 · 사람 역할 재정의 실패 · 파일럿에서 멈춤.  
실패는 모델 성능보다 '운영 설계 실패'에서 더 자주 발생합니다.

※ 출처: Gartner 보도자료 (2025.6.25)

## AX를 성공시키는 다섯 가지 조건 — 선도 기업이 다르게 하는 것들입니다

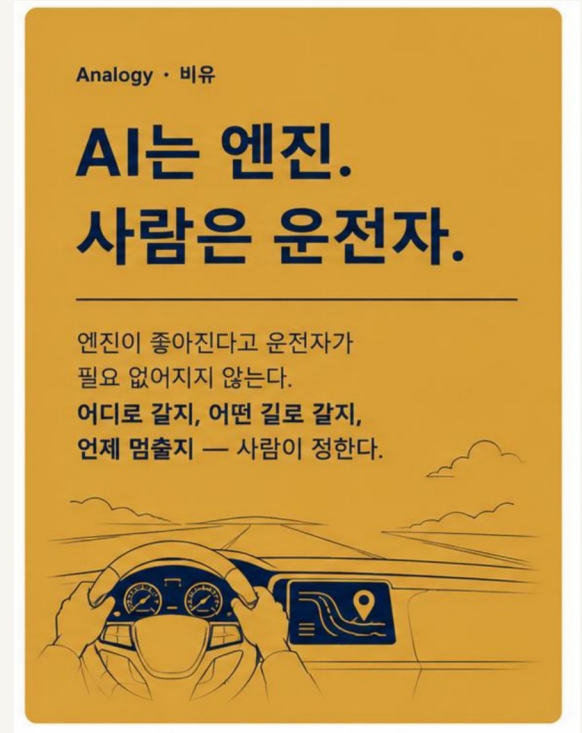


명확한 KPI · 1~3개 도메인 집중 · 지식 베이스+표준 프롬프트 · 검증·승인 체계 · 역할 재정의.

**McKinsey 2026 — AI를 영업이익에 연결한 고성과 기업은 여전히 약 6%뿐입니다.**

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

|    |                                                                                    |                                                                              |
|----|------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| 01 |   | <b>Question · 질문</b><br><b>좋은 질문을 던지는 사람</b><br>맞는 답이 나오는 이유는 질문이 맞기 때문이다.   |
| 02 |   | <b>Standard · 기준</b><br><b>기준을 세우는 사람</b><br>'좋은 답'이 무엇인지 정의해야 AI가 그쪽으로 일한다. |
| 03 |   | <b>Verify · 검증</b><br><b>결과를 검증하는 사람</b><br>출처, 사실, 윤리적 리스크는 사람이 확인해야 한다.    |
| 04 |   | <b>Judge · 판단</b><br><b>판단하는 사람</b><br>맥락·이해관계·가치 충돌은 알고리즘이 풀지 못한다.          |
| 05 |  | <b>Own · 책임</b><br><b>책임지는 사람</b><br>결과의 윤리·법·고객 영향은 사람이 진다.                 |



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

AI는 엔진.

사람은 운전자.

AI에게 일을 시키고, 결과를 검증하는 사람이 됩니다.

---

AI를 잘 쓰는 사람은  
답을 맡기는 사람이 아닙니다.

감사합니다

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