

최신 AI 10대 핵심 기술 가이드북

LLM부터 피지컬 AI를 넘어 AGI·ASI 까지

오늘날 산업과 학계를 재편하고 있는 10가지
AI 핵심 기술을 한 권의 작은 책으로 정리했습니다.



01	개요	1
02	기술 관계도	2
03	타임라인	3-6
04	10대 핵심 기술	7-16
05	주요 플레이어 비교	17-19
06	특허 동향	20



개요

LLM(대규모 언어 모델)이 일상의 도구가 되어 **생성형 AI 시대**가 도래했고, AI는 텍스트를 넘어 영상·음성·로봇·물리 세계로 영역을 빠르게 넓히며 산업 전반의 혁신을 다시 정의하고 있습니다.

본 가이드는 오늘날 AI 지형을 형성하는 **10가지 핵심 기술**에 대한 기술 관계도, 타임라인, 핵심 기술들의 특징 등을 정리하여, 특히 심사관과 출원인이 최신 AI 주요 기술과 빠른 변화의 흐름을 한눈에 파악할 수 있도록 돕는 것을 목적으로 합니다.



10가지 핵심 기술

- 1 **LLM(대규모 언어 모델)**
Large Language Model 대규모 언어 데이터를 학습해 인간 언어를 이해하고 생성하는 AI
- 2 **멀티모달 AI**
Multimodal AI 다양한 형태의 데이터를 연계해 의미를 통합적으로 이해하는 AI
(텍스트·이미지·음성·영상 등)
- 3 **에이전틱 AI**
Agentic AI 스스로 상황을 인식하고 계획·실행·평가를 반복하며 과업을 수행하는 AI
- 4 **피지컬 AI**
Physical AI 물리적 장치에 기반하여 현실 세계에서 행동을 수행하는 AI
- 5 **월드 모델**
World Models 현실 세계의 물리 법칙·시공간 특성을 이해하고 시뮬레이션하는 모델
- 6 **버티컬 AI**
Vertical AI 특정 산업이나 분야에 맞춰 설계된 전문형 AI
- 7 **모델 경량화**
Model Compression AI 모델의 크기와 연산량을 줄여 효율을 높이는 기법
- 8 **지식 증류**
Knowledge Distillation 큰 모델의 지식을 작은 모델로 전달해 성능을 유지하는 기법
- 9 **AGI(범용 인공지능)**
Artificial General Intelligence 인간이 할 수 있는 지적 과제를 이해·학습·추론·계획하여 수행하는 AI
- 10 **ASI(초인공지능)**
Artificial Super Intelligence 인간의 인지 능력과 판단력을 넘어서는 수준의 AI



인공지능은 LLM을 중심으로 **멀티모달·에이전틱·피지컬 AI**로 확장되고, **월드 모델**로 현실을 이해하며, **버티컬 AI·모델 경량화·지식 증류**를 통해 산업에 적용되고, 나아가 **AGI**와 **ASI**라는 미래 지능의 가능성으로 이어지고 있습니다.



AI 기술 관계도

오늘날 AI 분야의 기술들은 서로 독립된 영역이 아니라 유기적 관계를 가지며, 아래와 같이 **인간에 비유**할 수 있습니다.

아래 관계도는 기술들이 어떻게 유기적으로 연결·확장되는지 10가지 핵심 기술 관계 구조를 한눈에 보여줍니다.

[LLM이라는 ‘두뇌’를 중심으로, 다양한 능력이 더해지고 융합되어 AGI·ASI를 향해 나아가는 구조]



* 위 관계도는 직관적 이해를 돕기 위해 단순화된 형태로 재구성한 것이며, 실제 기술들은 더 복잡한 상호 관계를 가지며 영향을 주고받습니다.



AI 주요 타임라인 (1950 - 2022)

10대 핵심 기술을 포함한 전체 AI 발전 흐름에 대해 주요 이슈·모델 및 플레이어를 정리한 타임라인입니다.

	주요 이슈	관련 10대 핵심 기술	주요 이벤트·모델	연구자·플레이어	설명
1950 - 1979	AI 개념정립 태동 1차 겨울		튜링 테스트 (1950) 다트머스 회의 (1956) 퍼셉트론 (1958) 1차 겨울 (1969)	Alan Turing John McCarthy Frank Rosenblatt -	대화에서 인간과 구별되는지로 기계 지능을 판별하는 기준 제안 인공지능 용어의 첫 사용 인간의 뇌신경망 모방한 초기 인공 신경망 단층 퍼셉트론으로 XOR 논리 연산 해결할 수 없음 증명
1980 - 2005	전문가 시스템 흥망 2차 겨울		전문가 시스템 (1980) 역전파 알고리즘 (1986) 2차 겨울 (1987) Deep Blue (1997)	Edward Feigenbaum David Rumelhart - IBM	전문가 지식을 규칙 기반으로 표현, 특정 분야 의사결정 지원 시스템 다층 신경망의 오류 수정 학습 - 신경망 부활 예외 처리 불가능 등 전문가 시스템 한계·몰락 탐색·평가함수 기반 알고리즘, 세계 체스 챔피언 격파
2006 - 2016	딥러닝 혁명 생성형 AI 기반 확립		Deep Belief Network (2006) AlexNet (2012) GAN (2014) AlphaGO (2016)	Geoffrey Hinton Alex Krizhevsky Ian Goodfellow Google	제한된 볼츠만 머신 기반의 심층 신경망 구성 → 딥러닝 체제 GPU 도입, CNN 기반, 이미지 인식 대회 압도적 우승 → 딥러닝 혁명 진짜 같은 가짜 만드는 적대적 생성 신경망 → 이미지 생성 AI 기초 vs 이세돌 바둑 대국 승리 → 딥러닝·강화학습 대중화
2017 - 2019	LLM의 시작과 가능성	①	Transformer (2017) BERT (2018) GPT-2 (2019)	Google Google OpenAI	Attention 메커니즘, 현대 LLM의 기반 구조 양방향 Transformer 인코더 기반 언어 모델 대규모 Transformer 기반 언어 모델, 15억 개 파라미터
	지식 증류의 실용화	⑦ ⑧	DistilBERT (2019)	Hugging Face	지식 증류 기법으로 BERT를 더 작고 빠르게
2020	초대형 LLM 스케일링 충격	①	GPT-3	OpenAI	1,750억 개 파라미터로 GPT-2 대비 파라미터 100배 이상 ↑
	언어 모델 압축 경쟁 심화	⑦ ⑧	TinyBERT MiniLM	Huawei Microsoft	BERT 계열 모델의 증류·경량화 활발
	AI for Science 상징 등장	⑥	AlphaFold2	Google	단백질 구조 예측에서 혁신적 성능
	모델 기반 강화학습	⑤	DreamerV2	Google	환경 예측하는 잠재 모델 기반의 강화 학습
2021	이미지·텍스트 멀티모달 시작	②	CLIP DALL-E	OpenAI OpenAI	이미지·텍스트를 같은 표현공간 연결, 멀티모달 AI의 기반 텍스트→이미지 생성형 AI 대표 초기 모델
	LLM 파인튜닝 경량화의 전환점	① ⑦	LoRA	Microsoft	대형 모델 전체 학습 → 일부 저랭크 어댑터만 학습
2022	이미지 생성 AI 대중화	②	DALL-E 2 Stable Diffusion	OpenAI Stability AI	텍스트→고품질 이미지 생성 화제 이미지 생성 AI 오픈 생태계 확산
	대화형 LLM의 대중 충격	①	GPT-3.5(ChatGPT)	OpenAI	일반 사용자가 LLM 직접 경험, 생성형 AI 대중화 촉발
	도구 사용형 에이전트 시작	① ② ③	ReAct	Google	LLM이 생각·행동하며 도구 호출하는 에이전트 패턴 제시
	로봇 제어용 파운데이션 초기 모델	② ③ ④	RT-1	Google	텍스트·이미지 → 로봇 행동 토큰, Transformer 기반 모델

① LLM ② 멀티모달 AI ③ 에이전트 AI ④ 피지컬 AI ⑤ 월드 모델 ⑥ 버티컬 AI ⑦ 모델 경량화 ⑧ 지식 증류 ⑨ AG(근접) ⑩ ASI(미실현)



AI 주요 타임라인 (2023 - 2026)

10대 핵심 기술을 포함한 전체 AI 발전 흐름에 대해 주요 이슈·모델 및 플레이어를 정리한 타임라인입니다.

	주요 이슈	관련 10대 핵심 기술	주요 모델	플레이어	설명
2023	프론티어 LLM 성능 경쟁	① ②	GPT-4 Claude 1 · 2 Gemini 1.0	OpenAI Anthropic Google	고성능 프론티어 LLM 개발 경쟁 (LLM 3강의 첫 격돌 시기)
	오픈 LLM 생태계	① ② ⑦ ⑧	Llama 2 Mistral 7B Qwen	Meta Mistral AI Alibaba	폐쇄형 모델에 맞서 공개 가능 경량화 오픈 모델 생태계 확산
	< 동시 확장 > 에이전틱 AI	③	AutoGPT LangChain	오픈소스 LangChain	작업 분해 · 도구 호출하는 자율형 LLM 에이전트 화제
	피지컬 AI	④	PaLM-E, RT-2	Google	VLM, VLA(Vision Language Model · Action) 기반 로봇 제어
	월드 모델	⑤	Dreamer V3	Google	모델 기반 강화학습 (게임 마인크래프트 다이아몬드 채굴)
	버티컬 AI	⑥	BloombergGPT Med-PaLM 2	Bloomberg Google	방대한 금융 데이터 학습, 금융 특화 AI 의료 및 바이오헬스케어 특화 AI
2024	실시간 멀티모달 인터페이스 경쟁	① ② ③ ⑨	GPT-4o Gemini 1.5	OpenAI Google	텍스트 · 이미지 · 음성 통합 실시간 처리 긴 컨텍스트(200만 토큰 - 책 20권 분량) 한번에 처리 가능
	텍스트→영상 고품질 생성	②	Sora Veo	OpenAI Google	객체 · 환경 일관성 유지하는 고품질 영상 생성 AI
	소형 LLM 재부상	① ⑦ ⑧	Phi-3 Gemma Llama 3	Microsoft Google Meta	온디바이스 목표 소형 언어 모델(SLM) 경쟁 치열
2025	추론형 LLM 모델 경쟁 폭발	① ② ③ ⑨	o3, Gemini 2.5 Claude 4, DeepSeek-R1	OpenAI, Google Anthropic, DeepSeek	수학 · 코딩 · 복잡 문제 등을 해결 가능한 추론 모델 경쟁
	에이전트가 '업무 실행'으로 이동	② ③ ⑥	Claude Computer Use Operator Devin	Anthropic OpenAI Cognition	컴퓨터 조작 · 웹 작업 · 코딩 · 문서 검토 등 수행 에이전트
	월드 모델 경쟁	② ④ ⑤	Cosmos Genie	NVIDIA Google	피지컬 AI용 월드 모델 ↔ 실시간 상호작용 월드 모델
	증류 · 오픈 LLM 모델 재부상	① ② ⑦ ⑧	DeepSeek-R1-Distill Llama-3.3, Phi-4-Mini	DeepSeek Microsoft, Meta	강한 추론 모델 → 작은 모델 증류(DeepSeek 무단 증류 의혹/논란)
2026	추론 · 에이전트 고도화 모델 경쟁	① ② ③ ⑨	GPT-5.x Claude 4.x Gemini 3.x	OpenAI Anthropic Google	추론 고도화 · 자율 행동 지능 · 실시간 사고력 조정 개발자용 클로드 코드, 병렬 분산 에이전틱 AI 멀티모달 영상 생성 Omni, 자율 AI 비서 Spark
	피지컬 AI 플랫폼 경쟁	② ③ ④ ⑤ ⑨	GR00T Gemini Robotics Figure Helix	NVIDIA Google Figure AI	가상 환경 선행 학습 → 로봇 주입 'Sim-to-Real' 강자 VLA 기반, 처음보는 환경 · 도구에 임기응변 · 추론 강점 초고속 모방 학습 - 인간 행동 비디오 시청만으로 동작 학습
	초압축 경량화	① ⑦	TurboQuant	Google	LLM의 KV 캐시 정확도 손실 없이 압축 (한국 연구자 참여)
	미토스(Mythos) 충격	① ② ③ ⑥ ⑨	Claude Mythos	Anthropic	사이버보안 특화 - 제한적 파트너 접근 (프로젝트 글래스워)

① LLM ② 멀티모달 AI ③ 에이전틱 AI ④ 피지컬 AI ⑤ 월드 모델 ⑥ 버티컬 AI ⑦ 모델 경량화 ⑧ 지식 증류 ⑨ AGI(근접) ⑩ ASI(미실현)



CHAPTER 03. 타임라인

최신 AI 10대 핵심 기술 가이드북

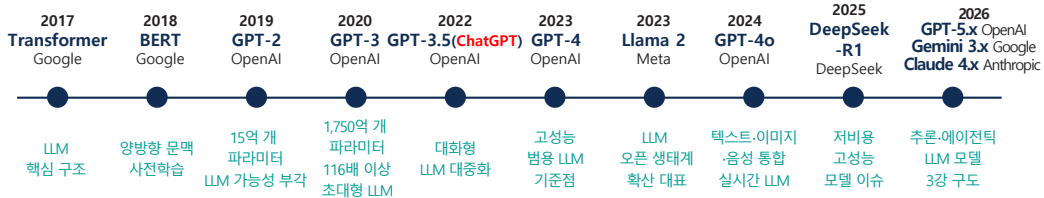


10대 핵심 기술별 주요 타임라인

10대 핵심기술별 주요 모델·이벤트 및 특징을 정리한 타임라인입니다.

1 LLM

Transformer → 스케일 확장 → 추론·에이전틱 모델



2 멀티모달 AI

이미지·텍스트 결합 → 이미지 생성 → 시각언어 추론 → 실시간 멀티모달



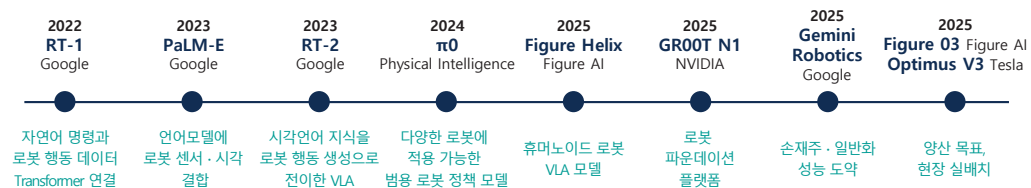
3 에이전틱 AI

도구 호출형 → 자율 작업 → 컴퓨터 사용·전문 업무 에이전트



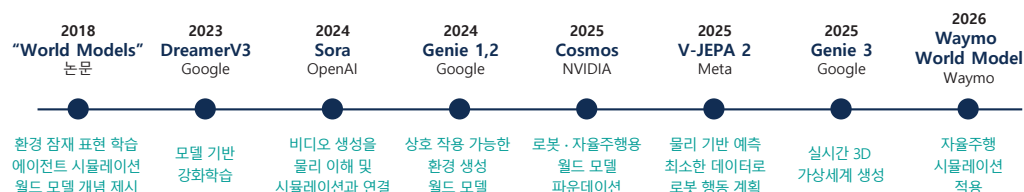
4 피지컬 AI

로봇 행동 Transformer → VLA 모델 → 로봇 파운데이션 모델



5 월드 모델

환경 상상형 강화 학습 → 영상 기반 세계 시뮬레이션 → 로봇·자율주행용 모델





CHAPTER 03. 타임라인

최신 AI 10대 핵심 기술 가이드북



10대 핵심 기술별 주요 타임라인

10대 핵심기술별 주요 모델·이벤트 및 특징을 정리한 타임라인입니다.

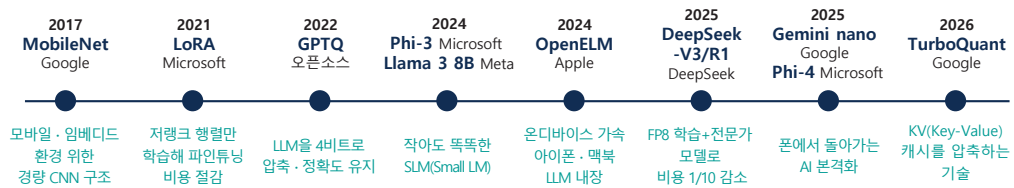
6 버티컬 AI

과학을 시작으로 코딩 · 금융 · 의료 · 법률 · 사이버보안 등 산업 전반으로 확장



7 모델 경량화

모바일 CNN → 어댑터 튜닝 → 양자화 → 소형 LLM → KV 캐시 압축



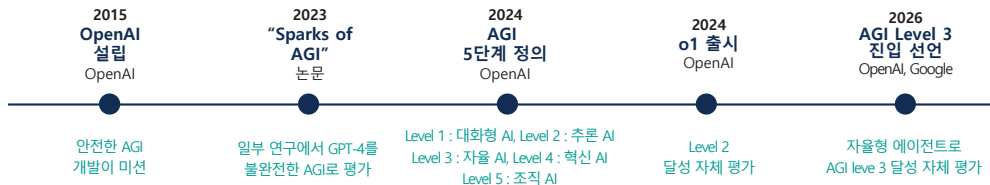
8 지식 증류

BERT 압축 → 모델 소형화 증류 → 추론 모델 증류



9 AGI

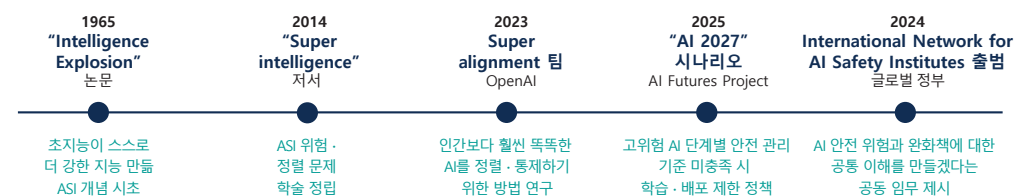
지능 판별 개념 → 게임 AI → 범용 LLM → 추론 · 행동형 AI *AGI를 향한 프론티어 모델은 등장했으나, 완성형 AGI는 아직 미흡의



10 ASI

초지능 가설 → 정렬 연구 → 프론티어 모델 안전 → 거버넌스 이슈

* ASI는 아직 가설적 · 이론적 개념 수준





1

LLM

Large Language Model

대규모 언어 데이터를 학습해

인간 언어를 이해하고 생성하는 AI

방대한 텍스트로 사전 학습된 거대 신경망으로, 다음에 올 단어를 확률적으로 예측하는 방식을 통해 번역·요약·대화·코드 작성 등 광범위한 언어 과제를 수행하는 범용 모델입니다.



CONCEPT DIAGRAM



• 주요 특징 · KEY FEATURES

1. 자가지도 학습

정답을 알려주지 않아도, 대규모 텍스트·코드 등에서 일부 단어를 예측하는 자가지도 학습으로 언어 패턴을 학습

2. Transformer 구조

긴 글에서도 단어들이 서로 어떻게 연결되었는지 효율적으로 파악·문맥 기억 구조

3. 거대 파라미터

수십억~조 개의 연결을 가진 거대한 신경망-신경망이 거대할수록 복잡한 추론까지 가능

4. 인컨텍스트 학습

추가 학습 없이 프롬프트(입력)에 제시된 문맥과 예시만을 참고해 즉각적으로 새로운 작업을 수행



이슈·논란

- 오픈소스 vs 폐쇄형 노선 갈등
- 학습 데이터 저작권 소송(NYT vs OpenAI 등)

• 대표 모델 / 플레이어



GPT

· OpenAI



Claude

· Anthropic



Gemini

· Google



Llama

· Meta



DeepSeek

· DeepSeek



Qwen

· Alibaba



HyperCLOVA X

· Naver



EXAONE

· LG AI





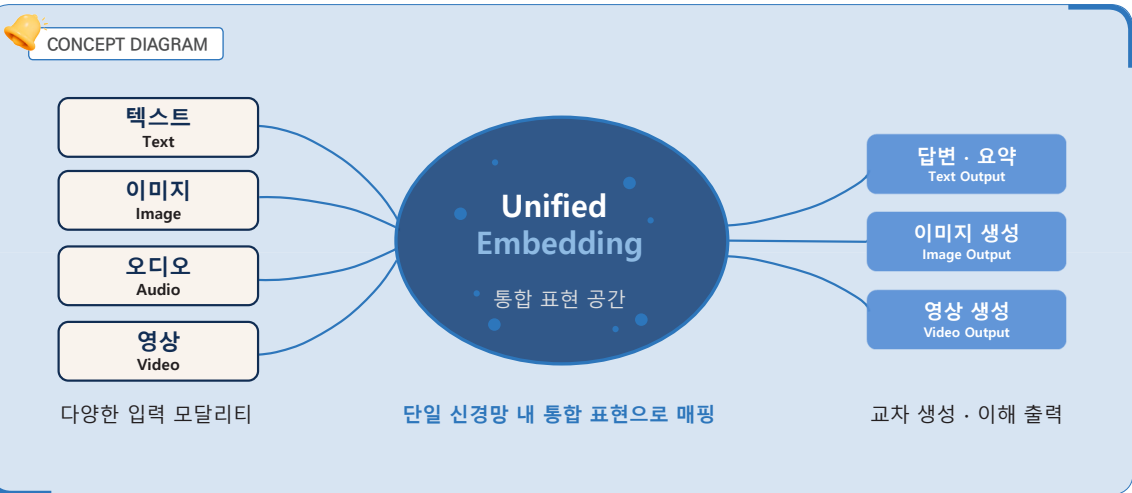
2

멀티모달 AI

Multimodal AI

다양한 형태의 데이터를
연계해 의미를 통합적으로 이해하는 AI

텍스트·이미지·음성·영상 등 서로 다른 형태의 데이터를 하나의 모델 안에서 통합 처리하는 AI입니다.
인간이 오감을 통합하듯 다양한 데이터를 함께 이해하고 생성할 수 있습니다.



• 주요 특징 · KEY FEATURES

1. 멀티 모달리티(다중 양식) 통합

인간이 보고·듣고·읽는 것을 동시에 하듯, 글·그림·소리·영상을 한 번에 이해

2. 모달리티 교차 생성

글로 그림을 그려내고, 그림을 보고 글로 설명하는 등 형태를 자유롭게 변환·다양한 형태로의 변환

3. 맥락적 추론

사진 속 표·차트·글씨를 함께 읽고 전체 맥락을 파악해 종합적으로 답변 생성

4. 응용 확장성

음성 비서·의료 영상 진단·자율주행 등 산업 전반에 폭넓게 적용 가능



이슈·논란

- 가짜 뉴스·성 착취물 등 딥페이크 논란
- AI 생성 콘텐츠 표시 의무화 논의

• 대표 모델 / 플레이어



GPT-4o
(텍스트·이미지·음성)

· OpenAI



Claude 3
(시각 입력 멀티모달)

· Anthropic



Gemini 1.0
(네이티브 멀티모달)

· Google



Sora-Veo
(텍스트 → 영상)

· OpenAI
· Google

Qwen2.5-VL
(다국어 멀티모달)

· Alibaba



Kling AI
(텍스트 → 영상)

· Kuaishou



HyperCLOVA X Omni
(국내 대표 옴니모달)

· Naver



VARCO Vision
(한국어·영어 VLM)

· NC AI





3

에이전틱 AI

Agentic AI

스스로 상황을 인식하고

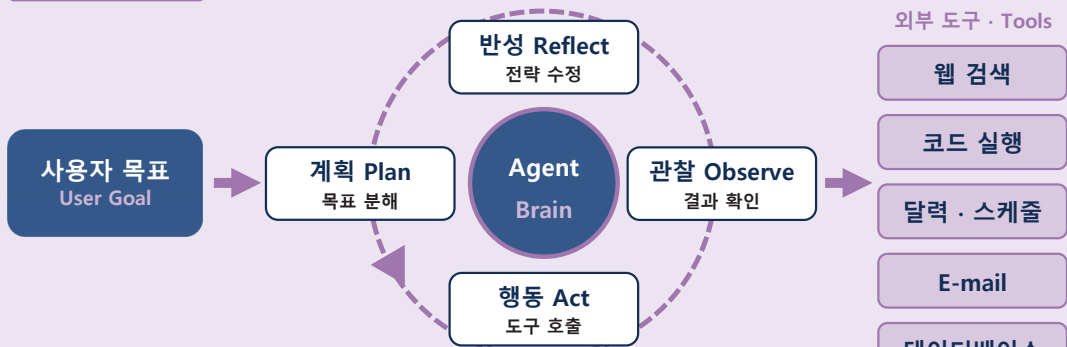
계획·실행·평가를 반복하며 과업을 수행하는 AI

스스로 과업(목표)를 분해하고 도구를 사용해 일을 처리하는 자율 AI입니다.

한 번의 질문·답변에 그치지 않고, 계획·실행·관찰·수정을 반복하며 실제 작업을 완수합니다.



CONCEPT DIAGRAM



에이전트는 목표 달성까지 루프를 반복 - 인간의 개입 없이 단단계 작업 수행

• 주요 특징 · KEY FEATURES

1. 자율적 작업 수행

큰 목표만 던져주면 세부 일정과 단계를 스스로 설계하고 실행

2. 도구 사용 능력

인간처럼 웹을 검색하고, 프로그램을 직접 실행하고, 외부 애플리케이션과 연결해 실제 작업을 처리

3. 장기 메모리

이전 대화와 작업을 기억해두고, 일관성 있게 후속 작업 진행

4. 자기 반성

수행 결과를 평가하고 다시 계획하여 행동을 수정



이슈·논란

- 에이전트 안전성 논쟁
- 에이전틱 AI 도입에 따른 일자리 대체 시점 논의

• 대표 모델 / 플레이어

Operator
(웹자동화)

· OpenAI

Claude Computer Use
(PC 조작)

· Anthropic

LangChain
(외부 도구 호출)

· LangChain

Devin
(코딩 에이전트)

· Cognition

Manus
(범용 에이전트)

· Butterfly Effect

Kimi K2.5
(병렬 에이전트)

· Moonshot

에이닷(A.)
(통화/비서 에이전트)

· SK Telecom

Kanana
(대화방 무인비서)

· Kakao





4

피지컬 AI

Physical AI

물리적 장치에 기반하여
현실세계에서 행동을 수행하는 AI

로봇·자율주행 등 물리적 신체를 통해 현실세계와 상호작용하는 AI입니다.

디지털 공간을 넘어 센서로 보고 모터로 움직이며, 물리 법칙 안에서 학습하고 판단합니다.



CONCEPT DIAGRAM

감각(Sensors)

카메라 · LiDAR

마이크 · IMU

촉각 센서

환경 센서

Embodied Brain
VLA Model
Vision-Language-Action

행위(Actuators)

관절 모터

로봇 손

이동 · 주행

음성 출력

물리 세계에서 행동 → 결과 생성 → 다음 행동 결정(Closed Loop)

• 주요 특징 · KEY FEATURES

1. 체화된 지능

센서와 로봇 몸을 갖춰 실제 공간에서 보고, 만지고, 움직일 수 있으며, 환경과 상호작용을 통해 학습

2. VLA(Vision, Language, Action) 모델

시각을 인지하고, 말을 이해하고, 손으로 행동하는 세 가지를 한 모델 안에서 처리

3. 시뮬레이션 학습 (Sim-to-Real / Real-to-Sim)

가상 환경에서 안전한 시뮬레이션을 먼저 충분히 수행한 후 그 능력을 현실세계(로봇)에 적용

4. 신뢰성·안정성 확보

실패 비용이 매우 큰 영역이며, 인간 곁에서 일하는 만큼 부딪힘이나 실수가 없도록 항상 주의를 기울임



이슈·논란

- AI 로봇의 가정 · 산업 현장 진입 기대감 확대
- 노동력 대체·안전 규제 지속 논의

• 대표 모델 / 플레이어



RT-2 · Gemini Robotics · Google
(VLA 모델)



GR00T N~
(로봇 파운데이션) · NVIDIA



Figure ~
(산업·휴머노이드 로봇) · Figure AI



π0 (Pi-Zero)
(범용 로봇 정책) · Physical Intelligence



H1 / G1
(저가 휴머노이드) · Unitree



Neo
(가정·물류용 로봇) · 1X Tech.



Atlas
(휴머노이드 로봇) · Boston Dynamics · Hyundai



RB-Y1
(양팔형 매니퓰레이터) · Rainbow Robotics · Samsung





5

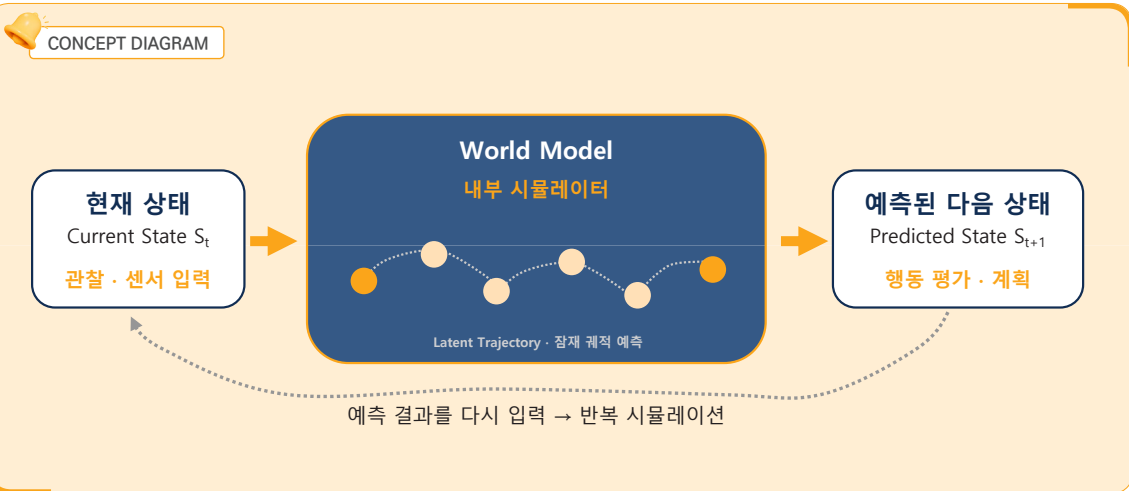
월드 모델

World Models

현실 세계의 물리 법칙·시공간 특성을
이해하고 시뮬레이션하는 모델

AI가 현실 세계의 물리 법칙·시공간·인과관계를 이해하고, 시뮬레이션하는 모델입니다.

현재 상태에서 다음 상태를 예측하고 “만약 ~ 한다면 ” 을 가상으로 설정하며 행동을 계획합니다.



주요 특징 · KEY FEATURES

1. 물리 동역학 학습

중력·관성·충돌 등 현실 세계의 물리 법칙을 데이터를
보며 자연스럽게 학습

2. 시뮬레이션 기반 계획

실제 행동 전에 가상 시뮬레이션(이렇게 하면 어떻게
될까?)를 미리 검토

3. 샘플 효율성

적은 수의 실제 데이터로도 가상으로 다양한 상황(경험)을
자체적으로 만들어 학습

4. 피지컬 AI 연결

자율주행·로봇이 ‘다음에 무엇이 일어날지’ 예측하는
로봇의 “두뇌”로 작동하는 차세대 핵심 기술



이슈·논란

- 진짜 세계의 이해인가, 그럴듯한 예측인가?
- 현실과 시뮬레이션의 차이로 인한 안정성 문제

대표 모델 / 플레이어



Dreamer V3

(모델 기반 강화학습)

· Google



Genie 2-3

(대화형 환경 생성)

· Google



V-JEPA 2

(잠재 예측 모델)

· Meta



Cosmos

(로봇·자율주행용)

· NVIDIA



Waymo World Model

(자율주행용)

· Waymo



Marble

(대형 월드 모델)

· World Labs



GigaWorld-1

· GigaAI



K-World Model

(월드 파운데이션 모델)

· NC AI





6

버티컬 AI

Vertical AI

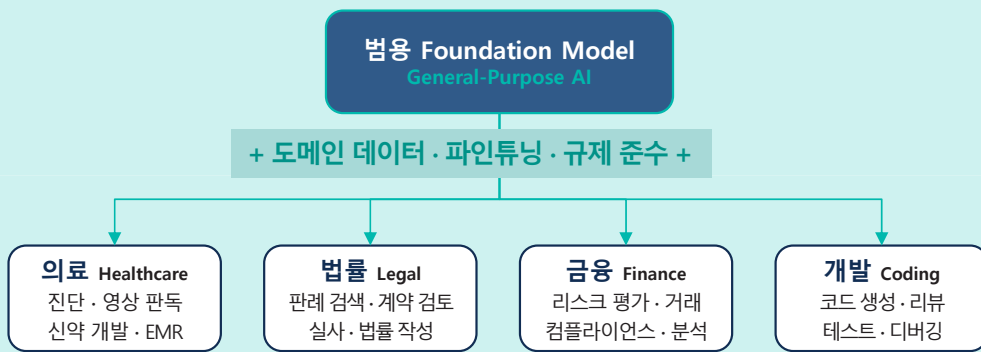
특정 산업이나 분야에 맞춰 설계된 전문형 AI

의료·법률·금융 등 특정 산업 도메인에 특화된 AI입니다.

범용 모델로는 다루기 어려운 전문 용어·규제·워크플로우를 산업 맥락에서 정확히 처리합니다.



CONCEPT DIAGRAM



• 주요 특징 · KEY FEATURES

1. 도메인(전문 분야) 특화

범용 AI를 그대로 사용하는 것이 아니며, 특정 전문 분야의 데이터로 추가 학습시켜 깊은 지식을 갖추

2. 규제·컴플라이언스

의료·금융·법률 등 분야·산업별로 다른 법규와 안전 기준을 모델 설계에 처음부터 반영

3. 워크플로우 통합

기존에 쓰던 의료 차트·법무 DB·금융 시스템 등 기존 업무 시스템과 자연스럽게 연결되어 작동

4. 전문가 협업

최종 결정은 전문가가 하고, AI는 일을 빠르게 처리하고 정확도를 높이는 보조 역할



이슈·논란

- 의료 AI 책임 소재 논쟁
- 법률 AI의 환각 사례(미국 판례 등)

• 대표 모델 / 플레이어



Med-PaLM

(의료 진단)

· Google



Harvey · CoCounsel

(법률 AI)

· Harvey AI

· Casetext



BloombergGPT

(금융 도메인)

· Bloomberg



GitHub Copilot · Cursor

(코딩 에이전트)

· Microsoft

· Anysphere



Mythos

(사이버보안)

· Anthropic



XuanYuan 2.0 · Spark LLM

(금융·의료)

· Du Xiaoman

· iFLYTEK



P-GPT · H-ACS

(제조 버티컬 AI)

· Posco

· Hyundai



Lunit ~ · VUNO

(국내 의료영상 AI)

· Lunit

· VUNO





7

모델 경량화

Model Compression

AI 모델의 크기와 연산량을 줄여
효율을 높이는 기법

거대 모델의 크기·연산량을 줄여 엣지·온디바이스에서 구동 가능하게 만드는 기술입니다.
성능 손실을 최소화하면서 메모리 사용량·전력 소모·추론 지연을 개선합니다.



CONCEPT DIAGRAM



• 주요 특징 · KEY FEATURES

1. 비트 정밀도 감소

AI가 쓰는 숫자의 정밀도를 줄여 메모리를 4~8배 절약
- FP32 → INT4까지 양자화

2. 불필요 파라미터 제거

불필요한 연결을 제거해 모델을 더 가볍게 만들
- 가지치기·희소화로 연산 효율 극대화

3. 추론 가속

전용 칩에 맞춰 최적화해 응답 시간을 크게 줄임
- GPU·NPU 최적화 커널

4. 온디바이스 AI 실현

스마트폰·자동차·IoT 등의 저사양·휴대용
기기에서도 AI를 직접 사용할 수 있게 함



이슈·논란

- 환경(전력 소비) 문제로 중요성 부각
- 온디바이스 AI 시대 개막

• 대표 모델 / 플레이어

	LoRA (저랭크 행렬 학습)	· Microsoft	
	Phi ~ (가상 합성 데이터)	· Microsoft	
	OpenELM (비대칭 계층 구조)	· Apple	
	Gemini nano·Gemma (윈도우 어텐션)	· Google	
	TurboQuant (벡터 양자화)	· Google	
	Mistral 7B (연산 방식 개선)	· Mistral AI	
	Qwen2.5-7B (연산 방식 개선)	· Alibaba	
	Solar-10.7B (모델 구조 개선)	· Upstage	



8

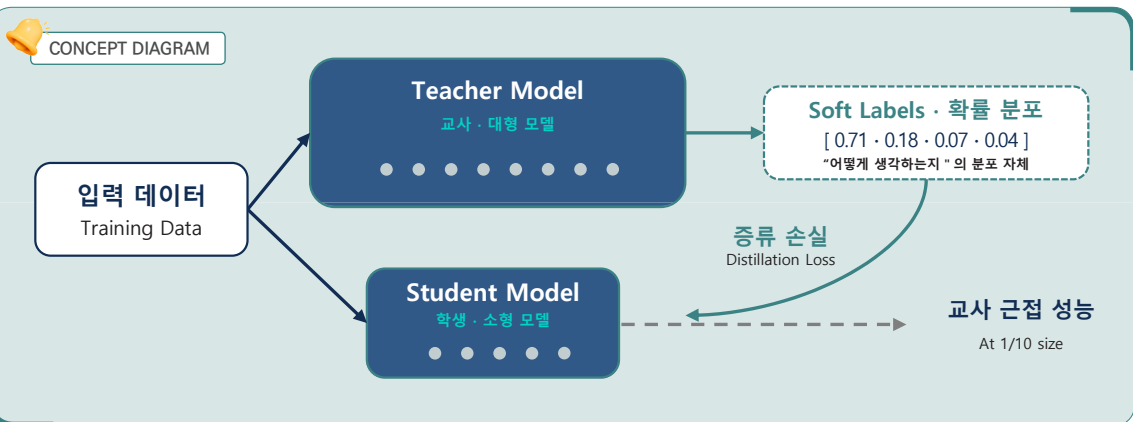
지식 종류

Knowledge Distillation

큰 모델의 지식을 작은 모델로 전달해
성능을 유지하는 기법

큰 모델(교사)의 지식을 작은 모델(학생)에게 이전하는 학습 기법입니다.

학생은 정답뿐 아니라 교사의 판단에 대한 “확률 분포” 또한 모방함으로써, 훨씬 작은 크기로도 교사에 근접한 성능을 달성합니다.



• 주요 특징 · KEY FEATURES

1. 교사-학생 학습 구조

성능이 좋은 큰 모델을 교사 모델, 작은 모델을 학생 모델로 구성
학생 모델은 교사 모델의 출력, 판단 경향, 중간 표현 등을 학습

2. Soft Label 학습

정답(Hard label)만 배우는게 아니라, 교사 모델이 답에 부여한 확률 분포를 함께 배워, 판단 경향까지 학습

3. 추론 비용 절감

학생 모델을 이용한 서비스를 통해 GPU 사용량·전력
소모·추론 시간 등이 대폭 축소
- 지식 증류는 경량화 방법 중 하나에 해당

4. 도메인·과업 전수

전문 교사 모델을 이용하여 특정 분야에 강한 소형 학생
모델을 양산 가능하고, 도메인 특화에 유리



이슈·논란

- DeepSeek 무단 증류 의혹/논란
- 모델 붕괴(품질 저하) 우려

• 대표 모델 / 플레이어



DistilBERT

(BERT 40% 경량화)

· Hugging Face



Llama ~ Dist.

(대형 → 소형 변환)

· Meta



Phi ~ Dist.

(SLM·합성 증류)

· Microsoft



Gemma ~ Dist.

(Gemini 기반 경량)

· Google



TinyBERT

(모바일·임베디드)

· Huawei



DeepSeek ~ Dist.

(추론 능력 증류)

· DeepSeek



Mistral 7B

(대형 → 소형 변환)

· Mistral AI



Mi:dm

(대형 → 소형 변환)

· KT





9

AGI

Artificial General Intelligence

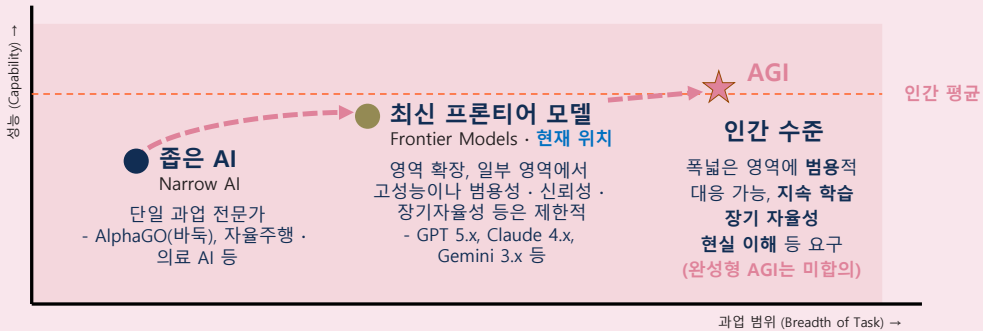
인간이 할 수 있는 지적 과제를
이해·학습·추론·계획하여 수행하는 AI

특정 과업의 초고성능 AI가 아닌, **폭넓은 분야의 지적 과제를 인간 수준 이상으로 안정적·자율적으로 수행할 수 있는 범용 지능**을 의미합니다.

현재 프론티어 모델들은 일부 영역에서 인간 수준 이상의 성능을 보이거나, 지속 학습성·신뢰성·장기 자율성·현실 이해의 측면에서 완성형 AGI로 합의된 사례는 아직 없습니다.



CONCEPT DIAGRAM



● 주요 특징 · KEY FEATURES

1. 구분 기준

폭넓은 지적 과제를 인간 수준 이상으로 수행하는 범용성
- 다양한 영역의 문제를 이해·추론·해결

2. 필요 조건

범용성, 견고한 추론·신뢰성, 지속 학습, 장기 기억, 장기 자율성, 현실·인간 관계 이해 등

3. 도달 경로

규모 확장, 추론 고도화, 에이전트·멀티모달·피지컬·월드 모델 등의 고도화 및 통합

4. 현재 수준

현재 프론티어 모델들이 추론 능력을 구비하여 특정 영역에서 인간 이상 성능을 보이거나, 범용성·지속학습·신뢰성·자율성 측면에서 완성형 AGI와는 차이가 있음

● 주요 모델 · 연구 / 플레이어



GPT
(AGI 미션 명시)

· OpenAI



Gemini
(“levels of AGI”)

· Google



Claude
(안전한 AGI 연구)

· Anthropic



JEPA
(월드 모델 기반)

· Meta



Safe Superintelligence · SSI
(안전 중심 연구소)



이슈·논란

- 시기 논쟁 : 낙관(‘26~27) / 중립(‘30 이후) / 회의(LLM으로는 불가)
- 경로 논쟁 : 규모 확장 / 추론 모델 기반 / 월드 모델 기반



10

ASI

Artificial Super Intelligence

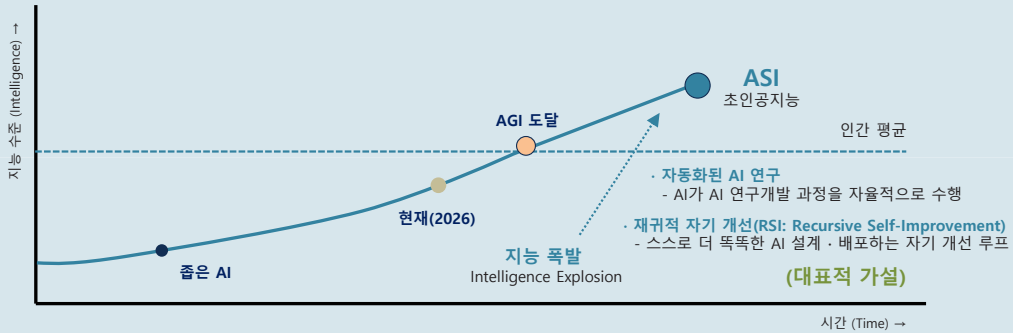
인간의 인지 능력과 판단력을 넘어서는 수준의 AI

AGI 이후 가능하다고 논의되는 이론적 초지능 개념으로, 인간 최고 전문가를 거의 모든 지적 영역에서 능가할 것으로 가정되는 초지능입니다.

재귀적 자기개선과 자동화된 AI 연구 등이 대표적 경로 가설이나, 검증된 실현 방법은 아직 없으며, 현재 논의의 중심은 정렬·통제·위험평가·거버넌스에 있습니다.



CONCEPT DIAGRAM



주요 특징 · KEY FEATURES

1. 구분 기준

인간 최고 수준을 넘어서는 초지능-과학 발견, 전략 수립, 기술 개발 등 거의 모든 영역에서 인간 최고 전문가를 크게 능가

2. 필요 조건

AGI의 범용성+자동화된 AI 연구, 자기개선 가능성, 정렬·감독·통제 메커니즘이 핵심 논의 대상

3. 도달 경로

재귀적 자기 개선과 자동화된 AI 연구가 대표적 가설
- 검증된 기술 경로는 아직 없으며, 안정성 연구 중심

4. 현재 수준

아직 미실현-이론적 개념-기술 개발보다 정렬·위험평가·거버넌스 논의가 선행되는 영역



이슈·논란

- 인간보다 똑똑한 AI를 어떻게 인간 가치에 맞출 것인가?(정렬)
- 공식적으로 ASI라 부를 시스템은 부재·도래 전 대비 논의

주요 연구 · 논의 / 플레이어

	Safe Superintelligence . SSI (안전 ASI 전용)	
	Superalignment . OpenAI (정렬 연구)	
	Responsible Scaling Policy . Anthropic (위험 평가 정책)	
	AGI Safety . Google (AGI/ASI 안전팀)	

학계 · 정부 차원 접근

MIRI·FHI (이론적 안전성 연구)	· 학계	
AI Safety Institute (AI 안전 연구소)	· 정부	
AI 관련법 (AI 규제 법안)	· 정부	



기술별 주요 플레이어 비교

1 LLM

대표 모델	플레이어	국가	주요 특징
GPT	OpenAI	미국	범용 LLM 대표 모델- 대화·코딩·추론·멀티모달·에이전트까지 폭넓게 확장
Claude	Anthropic	미국	안정성· 긴 문맥 처리- 긴 컨텍스트(200K+), 코딩· 분석· 에이전트 강점
Gemini	Google	미국	생태계 연동 강점- Google 검색· Workspace 생태계 통합
Llama	Meta	미국	오픈웨이트 생태계 대표- 공개 가중치 기반 자유롭게 미세조정· 오픈프레이미스 구축 가능
DeepSeek	DeepSeek	중국	저비용· 고효율 추론- 낮은 학습· 추론 비용, 고성능 수학· 코딩· 추론 성능
Qwen	Alibaba	중국	중국 대표 범용 모델- 다국어· 코딩· 수학 등, 자사 클라우드 기반 기업 적용 강점
HyperCLOVA X	Naver	한국	국내 한국어 특화- 한국어· 한국 문화 이해, 네이버 서비스 연계, 검색· 쇼핑 강점
EXAONE	LG AI	한국	산업· 전문지식 특화- 제조· 화학· 바이오 등 전문 도메인 적용, 기업용 AI 제공

2 멀티모달 AI

대표 모델	플레이어	국가	주요 특징
GPT-4o	OpenAI	미국	대표 옴니모달- 음성· 이미지· 텍스트 실시간 이해, 인간과 대화하듯 빠른 음성 응답
Claude 3	Anthropic	미국	문서 이해 강한 업무형- 긴 문서·차트·이미지 기반 질의 응답, 기업용 분석 강점
Gemini 1.0	Google	미국	네이티브 멀티모달- 처음부터 멀티모달로 설계, 자사 생태계 연결 강점
Sora· Veo	OpenAI· Google	미국	비디오 영상 생성 대표 모델- 물리 현상을 고려한 사실적인 영상 생성
Qwen 2.5-VL	Alibaba	중국	중국 대표 VLM- 이미지· 영상 이해, 표· 영수증· 문서 구조 분석에 강점
Kling AI	Kuaishou	중국	중국 대표 영상 생성 모델- 짧은 고품질 영상, 이미지 기반 영상화, 크리에이터 지원
HyperCLOVA X Omni	Naver	한국	국내 대표 옴니모달- 텍스트· 이미지· 오디오를 처음부터 함께 학습(네이티브)
VARCO Vision	NC AI	한국	한국어· 영어 이미지- OCR· 이미지 이해 강점, 자사 게임· 콘텐츠 제작 확장 가능성

3 에이전틱 AI

대표 모델	플레이어	국가	주요 특징
Operator	OpenAI	미국	브라우저 직접 조작- 웹페이지 보고 클릭· 입력· 스크롤하여 예약· 검색· 주문
Claude Computer Use	Anthropic	미국	화면 보고 마우스· 키보드 사용- 스크린샷 기반 PC 환경 이해, 단단계 업무 수행
LangChain	LangChain	미국	AI 에이전트 생성· 연결 도구- LLM· 도구· 워크플로우 등 연결하여 에이전트 구축
Devin	Cognition	미국	SW 개발 자동화- 코드 작성· 디버깅· 테스트 등 개발 업무 수행
Manus	Butterfly Effect	중국	범용 자율 에이전트- 리서치· 문서 작성· 웹사이트 제작· 데이터 분석 등 수행
Kimi K2.5	Moonshot	중국	중국 오픈 웨이트 모델- 화면· 문서 이해와 복잡한 작업 수행 능력 강조
에이닷(A.)	SK Telecom	한국	AI 개인비서 서비스- 검색· 기억· 일정· 통화/노트 요약 등 대화형 에이전트
Kanana	Kakao	한국	개인· 그룹 대화 맥락 기억- 개인용 Nana와 그룹용 Kana 통해 대화 맥락 기반 맞춤 지원



CHAPTER 05. 주요 플레이어 비교

최신 AI 10대 핵심 기술 가이드북



기술별 주요 플레이어 비교

4 피지컬 AI

대표 모델	플레이어	국가	주요 특징
RT-2 · Gemini Robotics	Google	미국	VLA 모델 선도 – 언어 · 시각 이해를 로봇 행동으로 연결
GR00T N~	NVIDIA	미국	로봇 파운데이션 모델– 휴머노이드 로봇의 언어 지시 이해 · 일반화 · 행동 생성
Figure ~	Figure AI	미국	상용화 추진, 범용 휴머노이드– 제조 현장 작업, 인간형 작업 수행, AI 자율 작업 강조
π 0(Pi-Zero)	Physical Intelligence	미국	범용 행동 정책 모델 – 다양한 로봇 적용 가능, generalist robot policy를 지향
H1 / G1	Unitree	중국	보급 · 연구용 휴머노이드– 상대적 낮은 가격과 빠른 상용화 속도
Neo	1X Tech.	미국 노르웨이	집안일 보조 휴머노이드 목표 – Redwood AI 기반 청소 · 정리 · 세탁 등 가정 작업 지원
Atlas	Boston Dynamics Hyundai	미국 한국	산업 현장 강점 휴머노이드– 뛰어난 운동성, 균형 제어, 공장 작업 적용 가능성
RB-Y1	Rainbow Robotics Samsung	한국	양팔형 모바일 피지컬AI– 바퀴형이동부와 양팔 조작 구조 결합, 연구 · 산업용

5 월드 모델

대표 모델	플레이어	국가	주요 특징
Dreamer V3	Google	미국	모델 기반 강화학습– 환경 모델을 학습한 뒤 가상의 미래 궤적을 상상해 정책 개선
Genie 2 · 3	Google	미국	텍스트로 플레이 가능한 3D 환경 생성– 조작 가능 · 실시간 상호작용 환경 생성
V-JEPA 2	Meta	미국	비디오 관찰해 세계 이해– 자기지도 학습, 시각 이해, 예측, 계획 → 로봇 제어
Cosmos	NVIDIA	미국	피지컬AI 학습용 – 로봇 · 자율주행차 실제 투입 전 가상 세계에서 학습 · 검증
Waymo World Model	Waymo	미국	회귀 · 위험 주행 상황 가상 생성– 실제 주행 영상, 언어 · 행동 제어 바탕
Marble	World Labs	미국	탐색 가능 3D 세계 생성 – 텍스트 · 이미지 · 비디오로 탐색 가능, 3D 월드 생성 · 편집 등
GigaWorld-1	GigaAI	중국	중국형 월드 모델 프레임워크– VLA 학습 데이터 생성, 픽셀-행동 동역학, 행동 디코딩
K-World Model	NC AI	한국	국내 월드 모델 – 잠재공간 기반 로봇 조작 행동 직접 생성, 효율성 · 정확도 향상

6 버티컬 AI

대표 모델	플레이어	국가	주요 특징
Med-PaLM	Google	미국	의료 특화– 의학 지식, 임상 질의응답, 의료 문답 안정성 평가 중심
Harvey · CoCounsel	Harvey AI · Casetext	미국	법률 특화– 계약 검토 · 감사 · 리서치 · 법률 문서 분석에 특화
BloombergGPT	Bloomberg	미국	금융 특화– 방대한 금융 데이터 학습, 금융 뉴스, 시장 분석, 기업 정보 등에 특화
GitHub Copilot · Cursor	Microsoft · Anysphere	미국	코딩 개발자 특화– 코드 자동 완성 · 코드 생성 · 디버깅 등 개발 업무 전반 지원
Mythos	Anthropic	미국	고위험 사이버보안 특화 – 강력한 기능으로 '프로젝트 글래스wing' 통해 일부 제한 공개
XuanYuan 2.0 · Spark LLM	Du XiaomanFinan. iFLYTEK Health.	중국	금융 · 의료 특화 – 중국의 대표적인 금융 챗 모델, 의료 LLM
P-GPT · H-ACS	Posco · Hyundai	한국	제조 특화– 철강 · 자동차 스마트 팩토리 등 제조 산업 특화형 AI 구축
Lunit ~ · VUNO ~	Lunit · VUNO	한국	의료 특화– 암 진단 · 판독 보조, 의료 데이터 분석 등을 지원하는 의료 AI



CHAPTER 05. 주요 플레이어 비교

최신 AI 10대 핵심 기술 가이드북



기술별 주요 플레이어 비교

7 모델 경량화

대표 모델	플레이어	국가	주요 특징
LoRA	Microsoft	미국	저랭크 행렬만 학습 - 대형 모델의 가중치 고정, 작은 어댑터만 추가 · 학습
Phi ~	Microsoft	미국	학습 데이터의 질 향상- 최상급 가상 합성 데이터만 사용하여 적은 모델로 고성능
OpenELM	Apple	미국	계층적 스케일링 - 뒤쪽 레이어 크기와 파라미터 조절하는 비대칭 계층 구조 사용
Gemini nano · Gemma	Google	미국	슬라이딩 윈도우 어텐션- 일정 범위(윈도우)만 집중 기억하도록 구조 설계
TurboQuant	Google	미국	고압축 양자화- KV 캐시 압축, 벡터 검색 지원(메모리 6x ↓, 추론 8x ↑)
Mistral 7B	Mistral AI	프랑스	그룹 쿼리 어텐션 - 데이터 처리를 합쳐주는 그룹 쿼리와 슬라이딩 윈도우 어텐션 도입
Qwen2.5-7B	Alibaba	중국	로터리 위치 임베딩(RoPE)- 문맥 위치 정밀하게 기억하도록 하는 RoPE기술 도입
Solar-10.7B	Upstage	한국	Depth Up-Scaling- 복잡한 MoE구조 없이 Depth Up-Scaling, 지속 학습으로 성능 향상

8 지식 종류

대표 모델	플레이어	국가	주요 특징
DistilBERT	Hugging Face	미국	BERT 압축 대표 종류 모델- BERT 40% 경량화로 성능 97% 유지
Llama ~ Dist.	Meta	미국	종류 · 파인튜닝 오픈 모델군- 자사 대형 모델 선행 학습 후 지식 종류
Phi~ Dist.	Microsoft	미국	합성 · 교과서형 데이터 기반 종류- 고품질 · 생성 데이터 활용 작은 모델 성능 향상
Gemma ~ Dist.	Google	미국	Gemini 기반 - Gemini 계열 연구 · 기술 바탕으로 작고 효율적인 모델 제공
TinyBERT	Huawei	중국	BERT의 여러 층 지식 전달- 임베딩·은닉상태 ·어텐션·출력 로직까지 단계별 모방
DeepSeek ~ Dist.	DeepSeek	중국	추론 모델 사고 과정 종류 - DeepSeek-R1의 추론 능력을 Qwen· Llama 기반 모델에 종류
Mistral 7B	Mistral AI	프랑스	유럽 대표 LLM- 자체구조 최적화와 고품질 학습으로 작은 모델 효율성 입증
Mi:dm	KT	한국	한국 특화 종류 - 한국어 문맥 이해 능력 · 기업 규정 · 도메인 지식 압축 종류

9 AGI

주요 모델 · 연구	플레이어	국가	주요 방향성
GPT	OpenAI	미국	"안전한 AGI 달성"이 목표- 추론 능력 강화모델 ·스케일링 확대 중심의 접근
Gemini	Google	미국	"Levels of AGI" 프레임워크 발표- AGI 단계적 정의 · 평가, Gemini 기반 연구
Claude	Anthropic	미국	안전한 AGI 연구- Constitutional AI, 해석 가능성 · 정렬 연구 선도
JEPA	Meta	미국	월드 모델 기반 접근 - LLM 스케일링만으로 부족 문제의식, 월드모델 · JEPA 접근
Safe Superintelligence	SSI	미국	안전 중심 단일 미션 - 제품보다 '안전 우선'을 내세우는 AGI · ASI 플레이어

10 ASI

주요 연구 · 논의	플레이어	국가	주요 방향성
Safe Superintelligence	SSI	미국	안전 중심 단일 미션 - 제품보다 '안전 우선'을 내세우는 AGI · ASI 플레이어
Superalignment	OpenAI	미국	초정렬연구- 인간보다 똑똑한 AI를 인간 의도에 맞추기 위한 정렬 연구
Responsible Scaling Policy	Anthropic	미국	안전 조치 정책- 성능 강해질수록 안전조치도 함께 높이는 단계별 정책
AGI Safety	Google	미국	위험 평가 · 안전 조치 - 위험 능력 수준을 평가하고, 심각도에 비례해 안전조치 적용
MIRI · FHI 후속	학계	글로벌	이론적 안정성 연구 - AI 실존위험 · 정렬 문제를 오래 전부터 제기한 학계 · 비영리 흐름
AI Safety Institute	정부	글로벌	정부 주도 안전 인프라 - 프론티어AI 모델을 시험 · 평가, 국제 AISI 네트워크 협력
AI 관련법	정부	글로벌	초지능 이전 단계부터 고위험 AI 관리 법적 안전망- EU AI Act 등 관련 규제 · 법안



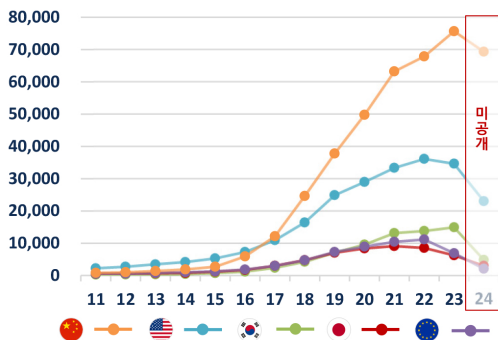
특허 동향

10대 핵심 기술을 포함하는 인공지능 기술에 대한 특허 동향*

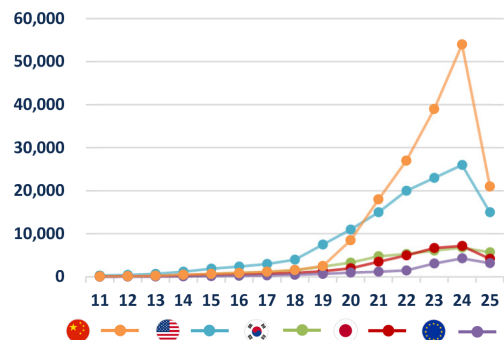
- 14년간('11~'24년) 출원 공개된 IP5 특허문헌 837,437건 대상

(출처 : 지식재산처 2025년 산업 맞춤형 심사혁신사업 "AI 기술이 접목된 발명에 대한 전략적 심사방안")

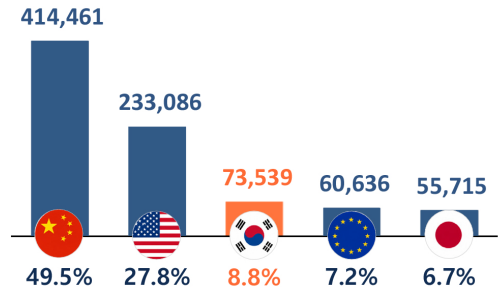
발행국별 출원 동향



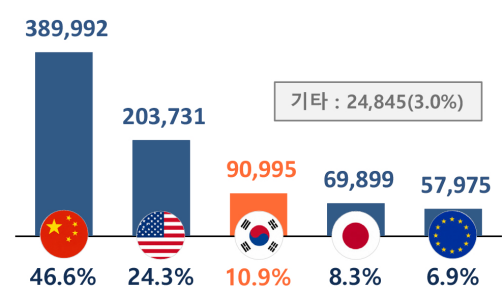
발행국별 등록 동향



발행국별 출원량 및 점유율



출원인 국적별 출원량 및 점유율



발행국별 주요 출원인

순위	MOIP	USPTO	JPO	EPO	CNIPA
1	Samsung Elec. (11,621)	IBM (10,220)	NTT (1,949)	Samsung Elec. (4,336)	Tencent (11,288)
2	LG Elec. (2,333)	Samsung Elec. (8,844)	NEC (1,801)	Google (1,942)	CAS(중국과학원) (6,178)
3	ETRI (1,576)	Google (5,351)	Fujitsu (1,757)	Microsoft (1,792)	Baidu (5,181)
4	KAIST (904)	Microsoft (5,113)	Canon (1,495)	Philips Elec. (1,600)	Huawei (4,230)
5	Google (886)	Intel (3,561)	Hitachi (1,217)	Huawei (1,310)	Ping An Tech. (3,544)