Market Insight | 2026.5.11

그림: ChatGPT(Vision Banana: 데이터 학습에서 데이터 제공자로)

MIRAE ASSET 미래에셋증권

AI Bi-weekly

잘 그려야 잘 본다

한종목
chongmok.han@miraeasset.com

김은지
eunji.kim.a@miraeasset.com

CONTENTS

Highlight of the Week 3
이미지계의 범용 모델 = Vision Banana? 3

I. Physical AI 4
1. VLM과 VLA, 뇌의 뉴런과 말초신경계 4
2. VLA의 데이터 병목 8
3. Vision Banana, 이미지 생성 모델의 범용화 11

II. AI Peer Table 16

III. Appendix: Charts 18

Highlight of the Week

이미지계의 범용 모델 = Vision Banana?

피지컬 AI의 두 축인 VLM과 VLA는 학습 데이터의 본질이 완전히 다른 모델입니다. VLM은 인터넷 스케일의 이미지-텍스트 데이터를 활용할 수 있지만, VLA는 카메라 입력·자기 위치 감각·움직임 명령이 동기화된 로봇 시연 데이터를 필요로 하며 이는 인터넷에 존재하지 않습니다. VLM이 뇌의 뉴런이라면 VLA는 근육과 관절을 움직이는 말초신경계에 해당하며, NVIDIA GR00T, Physical Intelligence π, Figure AI Helix, Google Gemini Robotics 모두 이 비대칭을 듀얼 시스템 구조로 풀어내고 있습니다.

VLM이 매 세대마다 새 모달리티를 흡수하며 도약하는 반면 VLA의 개선이 점진적 수준에 머무는 이유는 데이터 공급 격차에 있습니다. 로봇 시연은 신체마다 별도 수집이 필요하고 한 작업 학습에 수백~수천 달러가 발생하며, 학계 최대 공개 데이터셋 Open X-Embodiment조차 휴머노이드 비중이 낮고 단순 동작에 편중되어 LLM 학습 데이터 대비 절대량이 부족합니다.

이러한 맥락에서 4월 22일 Google DeepMind가 공개한 Vision Banana는 컴퓨터 비전 산업의 전제를 흔든 발표였습니다. Nano Banana Pro에 가벼운 지시 조정만 거친 결과 의미 분할·거리 추정·표면 법선 추정 등에서 SAM 3, Depth Anything V3 같은 전문 모델을 능가했습니다. 모든 비전 작업의 출력을 RGB 이미지로 통일하는 단 하나의 트릭으로 달성한 결과이며, 이는 "이미지 생성 학습이 LLM의 다음 토큰 예측과 같은 역할을 한다"는 가설을 입증한 셈입니다. 자연어 처리가 GPT-3 이후 단일 범용 LLM으로 흡수된 것과 동일한 패러다임 전환이 컴퓨터 비전에서도 가능해졌습니다.

이 전환은 피지컬 AI와도 직접 맞닿아 있습니다. Vision Banana의 출력 형식인 분할·깊이·법선은 NVIDIA Cosmos Transfer 같은 월드 모델의 입력 형식과 정확히 일치하며, 비전 파운데이션과 합성 데이터 파이프라인이 같은 RGB 인터페이스로 통합될 가능성이 열렸습니다. 이는 VLA의 데이터 병목을 우회하는 또 하나의 경로가 될 수 있습니다.

표 1. AI 관련 주요 일정

일	월	화	수	목	금	토
3	4	5	6	7	8	9
	· PLTR 실적 · DOUL 실적	· SHOP 실적 · AMD 실적	· ARM 실적	· VST 실적 · NET 실적 · COIN 실적 · CRWV 실적	· SMIC 실적(예) · 위스트론 실적 · Alchip 실적 · HD현대 실적(예)	
10	11	12	13	14	15	16
	·		· BABA 실적 · 소프트뱅크 실적	· 폭스콘 실적	·	
17	18	19	20	21	22	23
	·	·	· NVDA 실적	·		

자료: Bloomberg, 미래에셋증권 리서치센터

I. Physical AI

1. VLM과 VLA, 뇌의 뉴런과 말초신경계

(1) 두 모델의 본질적 차이

VLM과 VLA는 피지컬 AI를 구성하는 모델로서 같은 계열로 묶이지만, 학습되는 데이터의 본질은 완전히 다르다. VLM은 이미지와 언어를 함께 이해하고 일종의 액션 플랜을 생성하는 모델로, 기존 LLM과 마찬가지로 인터넷에 존재하는 이미지-텍스트 쌍을 거의 무한에 가깝게 활용해 학습할 수 있다. CLIP, PaLI, Gemini의 vision encoder, GPT-4V 계열이 모두 여기에 속한다.

반면 VLA는 비전과 언어의 이해 위에 '움직임 명령'이라는 출력 모달리티를 더한 모델이다. 즉 로봇의 관절 각도, 손가락이나 집게의 여닫음, 팔의 궤적과 같은 물리적 액션을 직접 출력해야 한다.

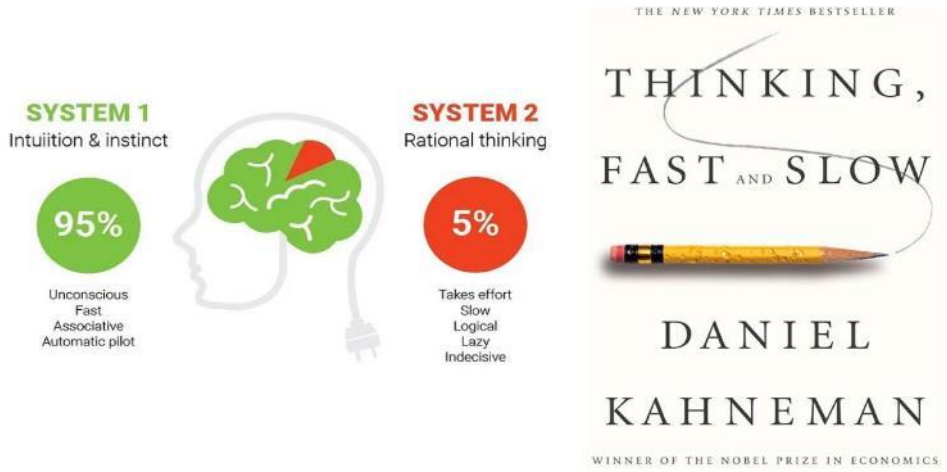
이 차이는 그대로 데이터의 차이로 이어진다. VLM은 인터넷 스케일의 이미지-텍스트 데이터로 학습할 수 있지만, VLA를 학습시킬 수 있는 데이터는 인터넷에 존재하지 않는다. 로봇이 어떤 작업을 수행할 때의 카메라 입력, 자기 위치 감각, 움직임 명령이 동기화된 데이터세트는 사람이 직접 로봇을 텔레오퍼레이션하거나, 시뮬레이션 환경에서 데이터를 합성하거나, 이미 배포되어 있는 로봇 군단(fleet)에서 수집해야만 모을 수 있다. Stanford, CMU 등 21개 기관이 협력해 만든 Open X-Embodiment 데이터세트는 22종의 로봇에서 500개 이상의 태스크, 1백만 에피소드를 모으며 자랑할 만한 규모를 달성했지만, LLM 학습에 쓰이는 수십 테라바이트의 인터넷 텍스트와는 여전히 비교할 수 없을 정도의 빈약한 양이다.

(2) 뇌의 뉴런과 말초신경계

이 본질적 차이를 가장 깔끔하게 설명하는 비유는 신경계의 분업 구조다. VLM이 인터넷 스케일 데이터로 학습된 뇌라면, VLA는 그 중추신경계에서 출발해 말초신경계(신경섬유·근육 신경분포)까지 이어지는 통합 운동 회로다. 인간도 글과 이미지로 다른 사람에게 텍스트로 가르칠 수 있는 것은 뇌에서 일어나는 인지와 추론 단계까지일 뿐이고, 근육과 관절을 직접 움직이는 신경 패턴은 책상 앞을 떠나 바깥으로 나가야만 학습할 수 있다. 인터넷에 사람의 모든 글과 사진을 다 모아도, 그것으로 누군가의 손 끝 움직임을 완벽히 재현할 수는 없다.

이러한 비유는 학계가 채택한 듀얼 시스템 구조와 맞아떨어진다. 엔비디아의 GR00T N1, Figure AI의 Helix, 구글의 Gemini Robotics 1.5 및 ER 1.6은 모두 명시적으로 System 2(느린 사고, VLM 기반 추론)와 System 1(빠른 사고, 움직임 수행)을 분리한다. 앞선 비유를 그대로 적용하면 System 2가 뇌의 뉴런, System 1이 척수와 말초신경계다. Daniel Kahneman의 〈Thinking, Fast and Slow〉 프레임이 robotics에 이식된 형태이기도 하다.

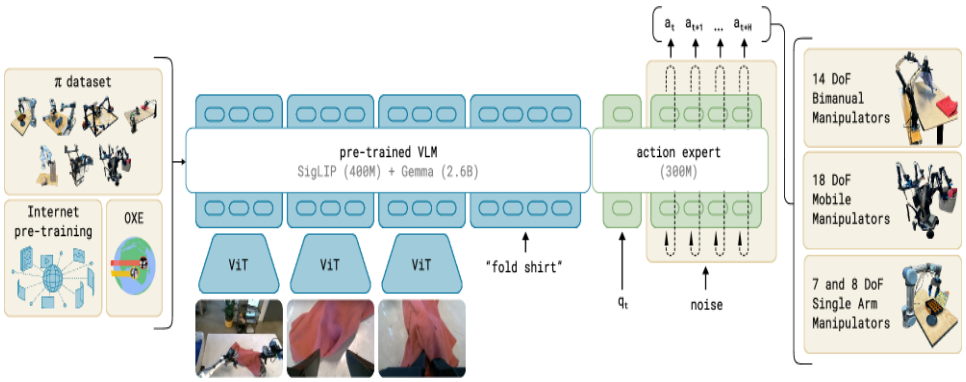
그림 1. System 1이 System 2보다 사고 속도는 느리지만, 전체적인 비중은 훨씬 높음



자료: 미래에셋증권 리서치센터

Physical Intelligence가 2024년 10월에 공개한 π_0 모델은 이 비대칭 구조를 모델 설계 단계에서 가장 명료하게 보여준다. π_0는 PaliGemma라는 Google의 사전학습 VLM(3B)을 그대로 백본으로 가져다 쓰고, 여기에 행동 전문가(action expert)라는 별도의 트랜스포머 가중치 0.3B 파라미터만 새로 추가했다. VLM과 행동 전문가가 하나의 신경망 안에 통합된 일체형 구조이지만, 학습 단계에서 백본과 액션 부분이 분리되어 있다는 점에서 dual-system VLA의 시조 격으로 평가된다. 결국 총 3.3B 모델 중 약 91%는 이미 만들어진 VLM이었고, 로봇 학습용으로 새로 만든 부분은 단 9%에 불과했다. 즉 Physical Intelligence는 "중요한 것은 액션 부분"이라는 인식을 모델 아키텍처에 그대로 반영한 셈이다.

그림 2. π_0 모델은 PaliGemma 모델을 VLM 부분으로 사용하고, VLA 부분만 자체 제작해 결합



자료: Physical Intelligence, 미래에셋증권 리서치센터

이 설계는 산업 전반에 걸친 합의에 가깝다. 엔비디아의 GR00T N1.5도 엔비디아 Eagle 2.5 VLM을 백본으로 쓰고 그 위에 Diffusion Transformer를 쌓았다. Figure AI Helix는 7B 규모의 VLM(System 2)을 7~9Hz로 돌리면서 비전-운동 정책(System 1)는 200Hz로 돌린다. Gemini Robotics 1.5 + ER 1.6 역시 같은 Gemini 3.x 백본에서 출발해 고레벨 추론과 움직임 명령을 분리하고 있다. 이는 모든 빅테크 진영이 같은 결론에 도달했음을 의미한다: VLM이 fancy해질수록 구동 부분은 분리되어야 한다.

(3) 발전 속도의 비대칭

VLM과 VLA의 본질적 차이는 출시 주기에도 그대로 반영된다. VLM 진영의 진화 속도는 VLA 진영을 명확히 앞선다.

Google Gemini 패밀리는 2023년 12월 1.0에서 출발해 약 분기 단위로 1.5, 2.0, 2.5, 3.0을 거쳐 3.1까지 진화했고, OpenAI의 GPT-4V부터 GPT-5.5에 이르는 멀티모달 라인, Anthropic Claude 3 Opus부터 4.7 Opus까지의 비전 능력 강화도 비슷한 속도로 진행되었다. 매 세대마다 이미지 이해 정밀도, 다국어 광학 문자 인식(OCR), 비디오 추론, 에이전틱 비전 같은 새로운 능력이 추가되었다.

VLA 진영의 출시 주기도 가속되고는 있다. 엔비디아 GR00T는 2025년 3월 N1에서 N1.5, N1.7과 지난 3월 공개된 N2 프리뷰까지 1년 사이 네 번 발표되었고, Physical Intelligence의 π 시리즈는 π_0, $\pi_{0.5}$로 약 6개월 단위로 진화했다. Google Gemini Robotics는 1.0과 ER 1.0, 1.5와 ER 1.5, ER 1.6으로 약 6~7개월 간격으로 출시됐다.

표 2. VLM 기반 주요 모델 출시 이력

모델 시리즈	버전	출시 시점	주요 능력
Google Gemini	1	23년 12월	텍스트·이미지·오디오·비디오·코드 통합 멀티모달
	1.5 Pro	24년 2월	1M 토큰 컨텍스트, 비디오 이해 강화
	2.0 Flash	24년 12월	에이전트 워크플로우, 도구 사용, 네이티브 이미지 생성
	2.5 Pro	25년 3월	thinking 모드 통합, 추론 능력 강화
	3.0 Pro	25년 11월	에이전틱 비전 도입, 코딩·수학 능력 도약
	3.1 시리즈	26년 2월	Flash Image가 Nano Banana 2, 이미지 편집 4배 가속
OpenAI GPT-4V 라인	GPT-4V	23년 9월	비전 입력 첫 도입
	GPT-4o	24년 5월	음성·이미지·텍스트 옴니모달, 실시간 응답
	GPT-5	25년 8월	통합 추론 모델, 컨텍스트 256K
	GPT-5.5 (Spud)	26년 4월	컴퓨터 사용·코딩·deep research 강화
Anthropic Claude	3 Opus	24년 3월	비전 능력 첫 도입
	3.5 Sonnet	24년 6월	컴퓨터 사용 베타
	4 Opus	25년 5월	장시간 에이전트 작업, 코딩 SOTA
	4.5 Opus	25년 1월	추론·도구 사용 강화
	4.7 Opus	26년 4월	현행 최신, 본 보고서 작성 시점 SOTA

자료: 각 사, 미래에셋증권 리서치센터

표 3. VLA 주요 모델 출시 이력

모델 시리즈	버전	출시 시점	주요 특징
엔비디아 GR00T	N1	25년 3월	첫 humanoid 오픈 파운데이션, dual-system 학술 발표
	N1.5	25년 6월	인간 비디오 학습 가능
	N1.7	26년 3월	휴머노이드 기능 점진 강화 (early access)
	N2	26년 3월	차세대 파운데이션 모델(preview)
Physical Intelligence π	π_0	24년 1월	플로우 매칭 행동 전문가, monolithic 구조
	π0-FAST	25년 2월	토큰화된 행동 표현으로 가속
	π0.5	25년 4월	오픈월드 일반화, 새 환경에서 동작
Google Gemini Robotics	1.0+ER 1.0	25년 3월	dual-model 구조 첫 발표, Apptronik 파트너십
	1.5+ER 1.5	25년 9월	thinking before action, 신체간 학습
	ER 1.6	26년 4월	에이전틱 비전 도입, 계기판 읽기 93~98%
Figure AI Helix	Helix	25년 2월	dual-frequency 구조, OpenAI 결별 후 자체 개발
Stanford·TRI OpenVLA	OpenVLA-7B	24년 6월	오픈소스 baseline, OXE 97만 데모로 사전학습

자료: 각 사, 미래에셋증권 리서치센터

그러나 출시 한 번당 능력 도약의 폭에서 비대칭이 명확히 드러난다. VLM은 한 세대 진화할 때마다 새로운 모달리티를 통째로 흡수한다. 이미지에서 비디오, 비디오에서 에이전틱 비전, 에이전틱 비전에서 도구 사용까지 본질적인 능력 추가가 연쇄적으로 일어났다.

반면 VLA의 개선은 매 세대 정교한 조작의 강건성이 조금씩 좋아지고 새로운 신체에 대한 적응력이 미세하게 늘어나는 점진적 수준에 머물렀다. GR00T N1에서 N1.7로의 진화, π_0에서 $\pi_{0.5}$로의 진화 모두 새로운 기능보다는 기존 동작의 신뢰도 향상에 가깝다.

이 비대칭의 원인은 단순하다. 앞서 언급했듯, 필요 데이터의 공급 차원이 문제다. VLM은 학습할 때마다 인터넷이라는 무한 자원에서 새로운 데이터 슬라이스를 흡수할 수 있지만, VLA는 모든 새로운 작업과 새로운 신체마다 사람이 직접 만든 로봇 시연 데이터를 필요로 한다. 즉 VLM과 VLA의 발전 속도의 격차는 학습 데이터의 본질적 형태 차이에서 온다. 이 데이터 병목이야말로 피지컬 AI의 근본적인 제약이 된다.

2. VLA의 데이터 병목

(1) 로봇 데이터 수집의 본질적 비용

로봇 시연 데이터를 모으는 비용은 이미지-텍스트 데이터를 모으는 비용과 단위가 다르다. 인터넷에서 이미지-텍스트 쌍을 수집하는 일은 크롤러를 돌리고 필터링하는 정도의 작업이지만, 로봇의 움직임 명령이 카메라 입력과 자기 위치 감각(proprioception)에 동기화된 데이터셋을 만들려면 사람이 로봇을 직접 텔레오퍼레이션하거나, 로봇이 실제로 시도하고 실패한 에피소드를 녹화하거나, 시뮬레이션 안에서 합성해야 한다. 이미지-텍스트 데이터가 단위 쌍당 비용이 사실상 0에 수렴하는 반면, 로봇 시연 데이터는 한 작업 학습에만 수백에서 수천 달러의 직접 비용이 발생한다.

데이터의 다양성 문제도 심각하다. LLM은 한 번 텍스트로 학습되면 그 능력을 어떤 작업에서든 어느 정도 활용 가능하지만, 로봇은 신체마다 별도의 학습 데이터가 필요하다. 7-DoF Franka 팔에서 학습한 정책은 Boston Dynamics의 Atlas 28-DoF 전신에 그대로 작동하지 않으며, ALOHA 2의 양팔 텔레오퍼레이션 데이터는 Apptronik의 Apollo 휴머노이드에 자동으로 전이되지 않는다. 이를 가능하게 하기 위한 신체간 학습이 별도의 연구 분야로 존재하는 이유다.

게다가 로봇이 실제 산업 환경에서 마주치는 수많은 엣지 케이스는 데이터로 모으기가 더더욱 어렵다. 정상적으로 작동하는 1만 에피소드를 모으는 것은 어떻게든 가능하지만, 100번에 1번 나오는 실패 케이스, 1,000번에 1번 나오는 안전 사고 직전의 상황, 10,000번에 1번 나오는 센서 노이즈 패턴 같은 것은 실제 세계에서 모으는 것이 불가능에 가깝다. 하지만 이 엣지 케이스 문제가 로봇이 실제 공장에서 무너지는 이유가 된다.

표 4. 이미지-텍스트 데이터 vs 로봇 시연 데이터 수집

비용 항목	이미지-텍스트 데이터	로봇 시연 데이터
데이터 자체 비용	0 (공개 웹 자원)	episode 수집마다 인건비·하드웨어 비용 발생
로봇 본체	-	플랫폼당 $40,000 +
인건비	거의 0에 수렴	숙련 운영자 $25~50/hr, 신규 운영자 4~8시간 학습 필요
수집 처리량	분 단위로 수십억 쌍 처리	운영자 1인당 시간당 5~50 episodes (작업 복잡도에 따라)
품질 필터링	CLIP 등 자동 필터링	20~30% episode가 품질 미달로 폐기
한 작업 학습 필요량	사전학습으로 보편 활용	작업당 50~5,000 episodes (알고리즘·복잡도에 따라)
신체 다양성 비용	-	새 신체마다 별도 데이터 수집 필요
단위 비용 추정	0에 수렴	$40/hour 기준 demo 1개당 약 $1 한 작업 학습에 직접 비용 수백~수천 달러

자료: Y Combinator, SVRC, 미래에셋증권 리서치센터

이 병목을 학계의 협력으로 풀어보려 한 시도가 Open X-Embodiment(OXE) 프로젝트다. Stanford, CMU, UC Berkeley, Peking University, Google, 엔비디아, Microsoft 등 21개 기관이 22종의 로봇에서 500개 이상의 작업, 1백만 에피소드 이상을 모아 공개 데이터셋으로 만들었다. RT-2, OpenVLA, π_0 등의 주요 VLA가 OXE 데이터를 학습에 사용하고 있다.

다만 OXE의 절대 규모는 LLM 학습 데이터 대비 여전히 미미하다. GPT 계열 학습에 쓰이는 텍스트 데이터가 수십 테라바이트 단위이고 토큰 수로는 수조 토큰에 이르는 반면, OXE는 수백 기가바이트 단위에 그친다. 또한 로봇 신체가 22종이나 포함되기는 했지만, 이중 휴머노이드는 거의 없고 산업용 매니퓰레이터 위주이며, 작업 분포 역시 픽 앤 플레이스, 밀기, 잡기 같은 단순 동작에 편중되어 있다. 즉, 학계가 모은 가장 큰 데이터셋조차 휴머노이드의 정교한 조작을 학습하기에는 부족하다는 것이다.

표 5. Open X-Embodiment 데이터세트의 구성 – 휴머노이드는 없음

로봇의 형태	데이터세트 수
Single Arm (단일 팔)	54
Mobile Manipulator (모바일 매니퓰레이터)	10
Wheeled Robot (바퀴형)	3
Bi-Manual (양팔)	2
Quadrupedal Robot (사족)	1
Human (사람)	1
혼합 (single-arm robot + single-arm human + tools)	1
합계	**72**

자료: 미래에셋증권 리서치센터
주: 양팔 로봇은 휴머노이드가 아니라, 로봇 팔 두 개를 책상 또는 프레임에 고정해 둔 텔레오퍼레이션 셋업을 의미

(2) 한계와 해결 시도

데이터 부족이 피지컬 AI의 본질적 제약이라는 인식은 학계와 산업계 모두에 공유되어 있다. 빅테크 진영과 학계가 이 병목을 우회하기 위해 추진하는 시도는 크게 여섯 갈래로 나뉜다.

첫째, 실세계 fleet 데이터 수집이다. Tesla가 자사 차량 fleet에서 모은 운전 데이터를 Optimus 학습에 활용하려 하는 것, Boston Dynamics가 현대차의 로봇 메타플랜트 애플리케이션 센터(RMAC)에서 산업 환경 데이터를 모으는 것, Figure AI가 BMW Spartanburg 공장에서 Figure 02를 운용하며 데이터를 수집하는 것이 모두 이 범주다.

둘째, 대규모 텔레오퍼레이션 인프라다. Boston Dynamics와 Apptronik는 자체 텔레오퍼레이션 시설을 두고 사람이 로봇을 직접 조작해 데이터를 모으고 있다. 사람의 시간과 임금이 들어가기 때문에 단위 비용이 가장 비싼 방식이지만, 그만큼 데이터 품질이 가장 높다.

셋째, 시뮬레이션과 월드 모델을 통한 합성 데이터 생성이다. 엔비디아의 Cosmos 월드 파운데이션 모델이 분할 맵, 깊이 맵, LIDAR 스캔, 자세 추정, 궤적 정보를 입력으로 받아 사실적인 비디오를 출력하는 것이 대표적이다. Google이 보유한 Genie 계열의 월드 모델도 같은 목적이다.

넷째, 합성 데이터 팩토리의 산업화다. 엔비디아가 GTC 2026에서 발표한 피지컬 AI 데이터 팩토리 블루프린트는 Cosmos Curator, Cosmos Transfer, Cosmos Evaluator를 묶어 실세계 데이터를 합성 데이터로 증폭시키는 참조 아키텍처다. Microsoft Azure와 Nebius가 첫 클라우드 플랫폼 파트너로 합류했고, FieldAI, Hexagon Robotics, Skild AI, Teradyne Robotics, Uber 등이 이를 채택했다. 데이터 생성 자체를 산업 표준 인프라로 만들려는 시도다.

다섯째, 인터넷 비디오 활용이다. Mimic Robotics가 인터넷 비디오에 플로우 매칭 행동 디코더를 결합한 mimic-video 모델로 기존 VLA 대비 10배 샘플 효율, 2배 빠른 수렴을 달성했다고 엔비디아 Robotics Week에서 발표했다. 정적 이미지-언어 백본을 비디오로 학습된 물리적 동역학으로 대체한 사례다.

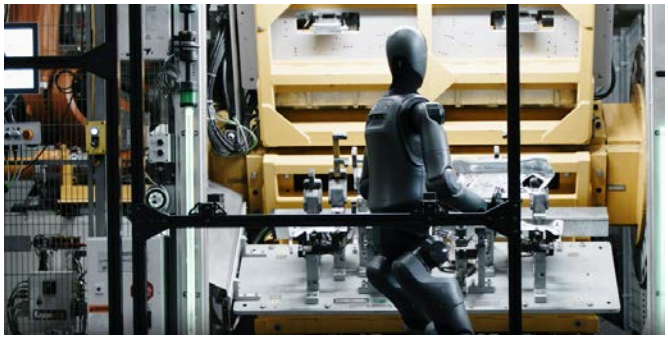
여섯째, 신체간 학습이다. Google Gemini Robotics 1.5가 ALOHA 2에서 학습한 작업이 Apptronik Apollo와 양팔 Franka에서도 작동한다고 보고한 것, 엔비디아 GR00T가 신체간 파운데이션 모델로 설계된 것이 여기 속한다. 서로 다른 구조와 자유도를 가지고 있는 기기의 데이터도 사용할 수 있도록 데이터 범용성을 확보하려는 움직임이다.

이 여섯 갈래는 서로 배타적이지 않으며, 보통 여러 갈래를 동시에 추구한다.

그림 3. Atlas 로봇

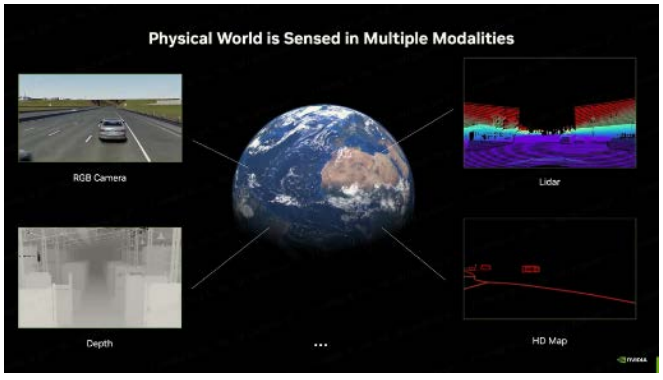
자료: Boston Dynamics, 미래에셋증권 리서치센터

그림 4. BMW 공장에 배치된 Figure 02



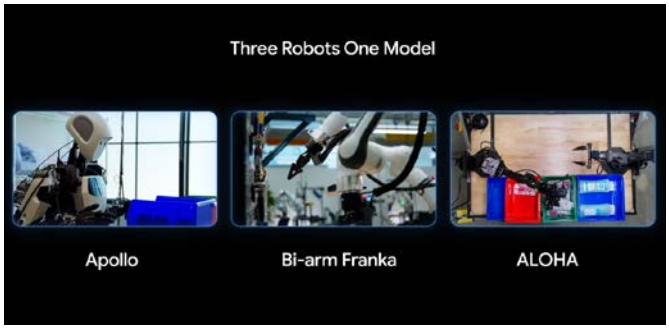
자료: Figure AI, 미래에셋증권 리서치센터

그림 5. 다양한 모달리티로 수집되는 물리적 세계 데이터



자료: 엔비디아, 미래에셋증권 리서치센터

그림 6. Gemini Robotics 1.5의 신체간 학습 능력 유지



자료: Google, 미래에셋증권 리서치센터

3. Vision Banana, 이미지 생성 모델의 범용화

(1) 바나나를 이해하고, 만들 줄도 아는 모델

4월 22일, Google DeepMind는 'Image Generators are Generalist Vision Learners'라는 제목의 논문을 공개하며 Vision Banana라는 통합 비전 모델을 발표했다. 같은 주 ICLR 2026 Google 부스에서 라이브 시연이 진행되었고, 발표 직후 올라온 리드 연구자들의 트윗이 1주일 안에 수만에서 수십만 뷰를 기록했으며, 컴퓨터 비전 학계가 오랫동안 기다렸던 순간이라는 평가도 나왔다.

핵심 발견은 단순하지만 강력하다. Nano Banana Pro에 가벼운 지시 조정만 거친 결과, 의미 분할, 인스턴스 분할, 단안 거리 추정, 표면 법선 추정 등 비전 이해의 핵심 작업에서 각 분야의 SOTA 전문 모델을 능가하거나 동등한 성능을 달성했다는 것이다. Meta의 SAM 3가 분할을, Depth Anything V3가 깊이 추정 성능을 전문가 모델을 만듦으로써 끌어올렸던 것에 반해, 구글은 범용 모델 하나로 한 번에 달성한 것이다. 게다가 Vision Banana는 이 과정에서 기반 모델인 Nano Banana Pro의 이미지 생성 능력도 그대로 유지했다. 이해와 생성을 하나의 모델로 통합한 셈이다.

이 발견의 본질적 의미는 베이스 모델이 가진 지식의 정체에 있다. 사실적인 이미지를 생성하려면 모델이 이미 객체 경계, 표면 기하, 깊이, 공간 관계, 재질 속성을 암묵적으로 알고 있는 상태여야 한다. Vision Banana는 단순히 그 잠재된 표현을 측정 가능한 형식으로 끌어올린 것으로, 새로 가르친 지식은 거의 없었다. 논문 저자들이 명시한 핵심 주장은 이렇다. "이미지 생성 학습이 LLM의 다음 토큰 예측과 같은 역할을 한다." 즉 LLM이 텍스트 생성 학습을 통해 일반화된 추론 능력을 우연한 부산물로 획득했듯이, 이미지 생성 모델도 이미지 이해 능력을 부산물로 획득했다는 것이다.

(2) 진리는 언제나 단순함에서 발견된다

Vision Banana의 기술적 우아함은 새로운 것을 거의 만들지 않았다는 데에 있다. 별도의 작업별 디코더 헤드도, 회귀 모듈도, 보조 손실 함수도, 새로운 아키텍처도 없다. 모든 비전 작업의 출력을 RGB 이미지로 표현하는 단 하나의 트릭으로 모든 것을 통일했다.

작동 방식은 다음과 같다. 사용자가 자연어 프롬프트로 작업을 지정하면(예: "이 이미지의 픽셀별 클래스 라벨링을 생성해라. 마카롱 케이크는 (255, 255, 0), 둥근 접시는 (255, 192, 128), 슬라이스 케이크는 (64, 192, 64)"), 모델은 그 지시에 따라 RGB 이미지를 생성한다. 그 다음 이 출력 이미지를 미리 정해둔 변환 규칙을 통해 다시 작업별 예측값으로 변환한다.

분할의 경우 픽셀을 가장 가까운 지정 색상으로 모아서 표현하면 되고, 깊이의 경우 RGB 채널이 인코딩한 거리 값을 역변환하면 된다. 물론 깊이는 0에서 무한대까지의 범위를 가지지만, 이를 가역적인 함수를 사용해 0~255 사이의 값을 갖는 RGB로 매핑했다. 따라서 어떤 깊이든 RGB 색조합으로 표현이 가능하며, RGB 색상을 보고 깊이 값으로 역연산하는 것도 가능하다. 표면 법선의 경우 더 단순하다. (x, y, z) 세 성분을 가진 단위 벡터를 그대로 (R, G, B) 세 채널에 매핑하면 끝이다.

이런 인코딩 트릭을 사용한 Vision Banana는 절대 거리(metric depth) 6개 벤치마크 평균에서 $\delta 1$(모델의 예측 거리가 일정 수준 이상으로 정확한 픽셀의 비율) 점수 0.882를 달성했고, 이는 Depth Anything V3의 0.823, Depth Pro의 0.715를 능가하는 수치다. 절대 거리는 단순한 원근 순서가 아니라 픽셀별 실제 거리 값을 미터 단위로 추정하는 작업으로, robotics나 자율주행에 직접 활용 가능한 형태라는 점에서 의미가 있다. 더 인상적인 것은 Vision Banana가 카메라 내재 매개변수를 학습이나 추론에 전혀 사용하지 않는다는 점이다. 대부분의 깊이 모델이 카메라 렌즈와 센서 정보에 의존하는 것과 달리, Vision Banana는 세계 지식만으로 거리를 추론한다.

또한 자연어 기반 모델이라는 특성상, 작업 어휘도 자연스럽게 열려 있다. 분할 작업에서 어떤 객체 카테고리를 분리할지는 프롬프트에 사용자가 지정한 색상-카테고리 매핑으로 결정된다. 즉 모델이 자체적으로 어휘에 제한을 두고 있지 않으며, 사용자가 프롬프트로 임의의 어휘를 정의할 수 있다. SAM 시리즈가 promptable concept segmentation을 통해 도달하려 했던 지점을 생성형 모델이 자연스럽게 달성한 결과다.

(3) 성능과 차별점

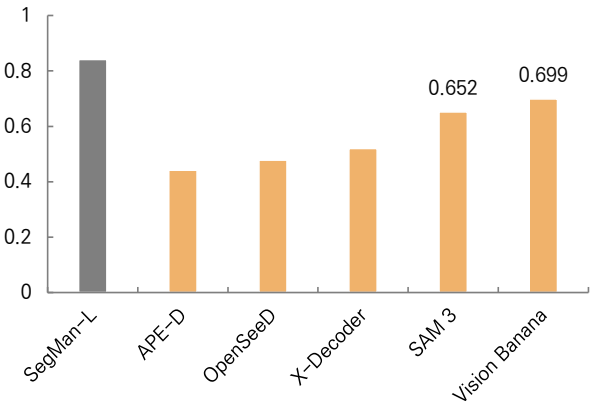
Vision Banana는 평가 데이터셋의 학습 세트가 튜닝 데이터에 일절 포함되지 않은 제로샷 전이(zero-shot transfer) 조건에서 평가되었다. 이는 모델이 진정한 일반화 능력을 보이는지를 검증할 수 있는 환경이다.

Vision Banana는 2D 이해 영역에서는 Meta의 SAM 3와 엎치락뒤치락하는 모습을 보인다. Cityscapes(의미 분할) 벤치마크에서는 SAM 3을 4.7%p 앞섰지만, SA-Co/Gold(인스턴스 분할)에서는 SAM 3에 비해 성능이 떨어졌다.

반면 3D 이해 영역에서는 더 결정적인 우위를 보였다. 거리 추정의 6개 벤치마크 평균 $\delta 1$ 점수가 0.882로 UniK3D의 0.823, Depth Pro의 0.715를 모두 앞섰다. 표면 법선 추정의 3개 벤치마크 평균 각도 오차도 Lotus-2, DSINE 등 기존 모델들을 능가하며 가장 낮은 수준이었다.

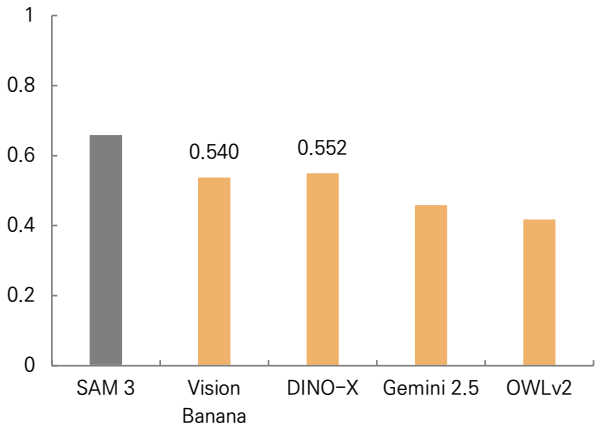
이러한 2D 및 3D 이해 능력을 추가하면서도 기존 Nano Banana 모델이 가지고 있던 생성 능력은 유지했다. 인간 평가에서 base 모델인 Nano Banana Pro 대비 53.5%의 승률을 보이며, instruction-tuning 과정에서 생성 품질이 거의 손상되지 않았음을 증명했다. 이는 Vision Banana의 가장 중요한 결과 중 하나다. 이해와 생성이 같은 모델 내부에서 양립 가능한 능력이라는 것을 처음으로 입증했기 때문이다.

그림 7. Cityscapes 벤치마크(의미론적 분할)



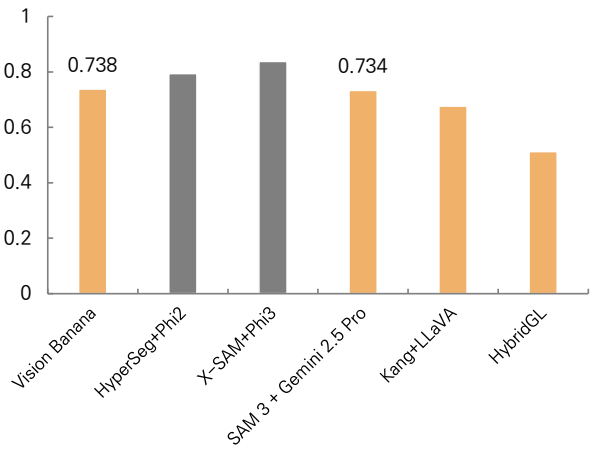
자료: Google, 미래에셋증권 리서치센터
주: 회색은 Non zero-shot

그림 8. SA-Co/Gold 벤치마크(인스턴스 분할)



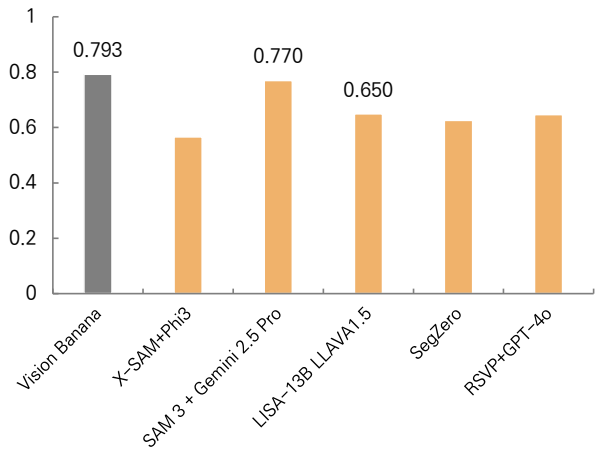
자료: Google, 미래에셋증권 리서치센터

그림 9. RefCOCOg 벤치마크(객체 관계 및 상황 묘사)



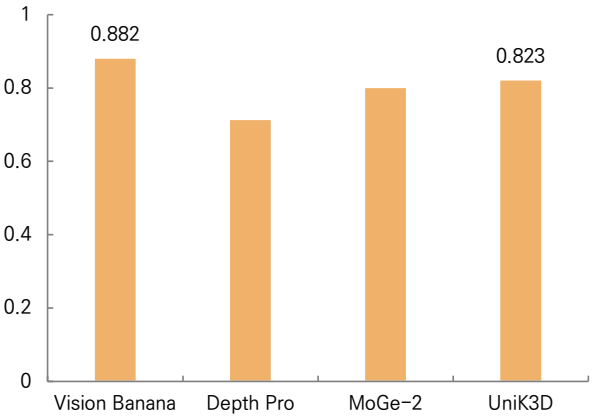
자료: Google, 미래에셋증권 리서치센터

그림 10. ReasonSeg 벤치마크(이미지+텍스트 분할 추론)



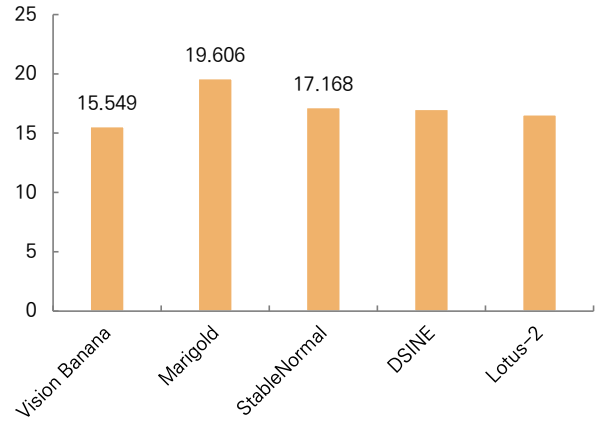
자료: Google, 미래에셋증권 리서치센터

그림 11. 절대 깊이 벤치마크 6개 평균



자료: Google, 미래에셋증권 리서치센터

그림 12. 표면 법선 벤치마크 3개 평균



자료: Google, 미래에셋증권 리서치센터

발표 직후의 반응은 '컴퓨터 비전의 GPT-3 순간'이라는 프레임으로 빠르게 수렴했다.

어드바이저로 참여한 Saining Xie는 X에 "조용히 큰 변화가 일어났다(Something big has quietly shifted)"고 표현하며 이 발표의 의미를 요약했다. 그는 컴퓨터 비전의 역사적 순환을 언급하며, 1990년대와 2000년대 초반의 비전 연구의 주류였던 생성형 접근이 2010년대에 들어 ImageNet과 ResNet의 등장과 함께 discriminative 접근으로 완전히 옮겨갔고, 다시 생성형 접근이 새로운 형태로 돌아왔다고 분석했다.

Xie는 또한 Vision Banana가 만든 분할 마스크가 인간 어노테이터의 마스크보다 픽셀 단위로 더 정확한 경우가 있다는 점, 정답(ground truth)보다 더 나은 예측이 가능할 수 있다는 점을 강조하며 같은 방식으로 비디오 생성 모델로 확장했을 때 월드 모델이 가질 가능성에 대한 기대감을 내비쳤다.

하지만 기술적 한계도 분명하다. 먼저, Vision Banana는 디퓨전 기반 추론을 사용하기 때문에 속도가 느리다. Saining Xie 본인도 "아직 실용적이지 않다"고 명시적으로 언급했을 정도로, 실시간 응답이 필요한 응용에는 쓸 수 없다. 또한, 부정 쿼리에 약하다. 이미지에 존재하지 않는 객체를 분할하라는 요청에서 모델이 잘못된 출력을 내는 경우가 있고, 이 점이 SA-Co/Gold 벤치마크에서 SAM 3에 뒤진 주된 원인이었다.

이러한 한계들은 Vision Banana가 여전히 개선할 수 있는 부분들이 있음을 보여준다. 다만 패러다임 전환의 가능성을 처음으로 정량 입증한 작업이라는 점은 비판자들 사이에서도 부정되지 않았다.

그림 13. Saining Xie, "단일 범용 모델이 처음으로 SAM 3, DepthAnything 3과 같은 도메인 특화 모델을 이겼다는 것이 중요"

자료: X(@Saining Xie)미래에셋증권 리서치센터

Vision Banana의 본질적 기여는 새로운 SOTA 모델을 추가했다는 것이 아니라 컴퓨터 비전 산업의 전제를 흔들었다는 데에 있다. 지난 20년 동안 산업용 컴퓨터 비전 파이프라인은 전문화 원칙 위에 세워져 왔다. 분할 모델, 거리 모델, 광학 흐름 모델, 결함 검출 모델이 각자 별도로 학습되고 별도의 통합 표면을 가지며, 별도의 재학습 주기와 별도의 지원 계약을 요구했다. 자율 주행, 산업 검사, 로봇 응용 분야의 컴퓨터 비전 시스템 대부분이 이런 전문가 모델 여러 개를 사슬로 엮은 형태로 구축되어 왔다.

하지만 Vision Banana는 하나의 통합 모델로 모든 작업을 한 번에 처리하는 구조다. 이는 시스템 복잡성을 줄이고, 개별 모델 유지보수 부담을 없애며, 학습 효율을 올린다. LLM의 발전 양상을 생각해 보면 이 패러다임 전환의 본질이 명확히 드러난다. 기존의 자연어 처리 분야는 감성 분석 모델, 개체명 인식 모델, 번역 모델, 요약 모델 등 작업별 전문가 모델로 분할되어 있었지만, GPT-3 등장 이후 이 모델들의 기능 대부분이 단일 범용 LLM 하나로 흡수되었다. 프롬프트 엔지니어링만으로 다양한 작업을 처리할 수 있었기 때문이다. 이미지 생성 사전학습이 LLM 사전학습과 같은 역할을 한다면 텍스트 생성이 언어 작업의 보편 인터페이스가 되었듯 이미지 생성도 비전 작업의 보편 인터페이스가 될 수 있을 것이라는 기대가 나타났다.

이 패러다임 전환은 피지컬 AI와도 맞닿아 있다. Vision Banana의 출력이 분할, 깊이, 법선을 나타내고 있고, 엔비디아 Cosmos Transfer 같은 월드 모델이 입력으로 받는 형식과 정확히 일치한다는 사실은 단순한 우연이 아니다. 비전 파운데이션과 합성 데이터 파이프라인이 같은 RGB 이미지 인터페이스로 통합될 가능성이 열린 것이다.

II. AI Peer Table

표 6. Peer Table (1)

종목명	티커	주가 (현지통화)	시가총액 (조원)	기간별 수익률(%)						실적발표 예정일	AI 밸류체인 관련 한줄평
				1W	1M	3M	6M	1Y	YTD		
최종 소비자(온디바이스, 앱)											
애플	AAPL US	287.44	6,144	5.9	13.4	3.4	7.3	47.1	5.8	07-31	엔드유저 데이터 보유 및 서비스 배포능력
테슬라	TSLA US	411.79	2,251	7.9	18.8	0.2	-4.1	49.1	-8.4	07-02	FSD v13 공공 출시 및 1Q25 중국 출시 가능성
Pony AI	PONY US	9.63	6	-2.4	9.7	-32.7	-31.4	-35.4	-33.6	05-26	중국의 자율주행 선도기업 중 하나
삼성전자	005930 KS	266,000	1,555	19.6	24.9	65.6	174.7	395.7	121.4	07-07	온디바이스 AI, HBM, 파운드리 영역에서의 가능성
듀오링고	DUOL US	113.61	8	3.2	18.0	-4.8	-43.5	-77.1	-35.3	08-06	AI 캐릭터와 상호작용하는 학습 플랫폼
크래프톤	259960 KS	271,000	13	0.6	14.8	13.7	6.0	-27.9	10.7	07-29	PUBG, inZOI 등 자사 게임에 AI CPC 기능 탑재
쇼피파이	SHOP US	111.74	211	-7.8	-4.5	-0.3	-26.7	18.2	-30.6	08-06	이커머스 콘텐츠 생성의 자동화(Shopify Magic AI)
코인베이스	COIN US	192.96	74	2.8	10.1	16.9	-37.6	-1.8	-14.7	05-07	AI 에이전트간 자율 암호화폐 거래 인프라 구축
비즈니스 효율화											
크라우드스트라이크	CRWD US	505.72	187	13.5	19.5	27.9	-6.3	19.7	7.9	06-03	기업 고객 대상 엔드포인트 보안(Charlotte AI)
IBM	IBM US	231.31	316	0.1	-5.6	-22.2	-23.7	-6.4	-21.5	07-22	기업 고객 대상 AI 모델 개발, 배포 플랫폼(watsonx)
액센츄어	ACN US	180.19	162	0.8	-7.9	-24.5	-25.6	-39.6	-31.9	06-18	각 산업별 맞춤형 AI 에이전트를 만들기 위한 첨병
SAP	SAP GY	151.38	318	5.8	6.0	-10.1	-28.6	-40.6	-26.1	07-23	기업용 SW 솔루션에 AI를 내장한 에이전트(Joule)
인포시스	INFO IN	1,162.70	73	-1.6	-13.2	-22.9	-21.3	-20.6	-28.0	07-23	각 산업별 맞춤형 AI 에이전트를 만들기 위한 첨병
일루미나	ILMN US	138.71	31	9.4	9.7	15.9	13.8	81.7	5.8	07-31	AI 기반 첨단 유전체 분석
HD현대	267250 KS	287,000	23	-9.3	13.7	19.9	38.9	268.2	53.0	05-13	AI, 로보틱스, 디지털 트랜스포메이션 관련 노력
AI Ops											
클라우드플레어	NET US	256.79	132	25.3	18.7	48.3	10.3	111.0	30.3	05-07	실시간 AI 추론에 CDN 수요증가 가능성(Workers AI)
팔란티어	PLTR US	137.05	478	-1.5	-8.7	0.8	-23.0	24.0	-22.9	08-04	비즈니스 현장의 의사결정을 돕는 AI 플랫폼(AIP)
파운데이션 모델											
알파벳	GOOGL US	397.99	6,995	3.4	30.3	23.4	42.9	163.8	27.2	07-23	OpenAI와의 AGI 경쟁이 가능한 업체(Gemini)
메타 플랫폼스	META US	616.81	2,279	0.8	7.3	-6.7	-0.6	3.7	-6.5	07-30	오픈소스 AI 개발의 선구자(Llama 등)
알리바바	9988 HK	140.90	503	11.8	11.4	-9.1	-12.0	17.9	-1.3	05-13	중국의 CSP이자 중국의 대표적인 언어모델(Qwen)
네이버	035420 KS	208,000	33	-2.1	2.2	-16.1	-18.9	11.3	-13.7	08-07	한국형 LLM 개발 및 AI 서비스(HyperCLOVA)

자료: Bloomberg, 미래에셋증권 리서치센터
주: 실적발표 일정은 변경될 수 있음.

표 7. Peer Table (2)

종목명	티커	주가 (현지통화)	시가총액 (조원)	기간별 수익률(%)						실적발표 예정일	AI 밸류체인 관련 한줄평
				1W	1M	3M	6M	1Y	YTD		
클라우드											
마이크로소프트	MSFT US	420.77	4,549	3.2	13.0	5.1	-15.0	-2.2	-12.8	07-30	OpenAI 모델 라이선스 독점권 보유한 세계 2위 CSP
아마존 닷컴	AMZN US	271.17	4,245	2.3	26.9	28.9	10.9	43.7	17.5	07-31	최고의 CSP로서 Anthropic에 수 조원 투자
오라클	ORCL US	194.59	815	20.6	36.4	36.7	-18.2	31.6	0.5	06-11	AI 데이터센터 capex 경쟁에 진입한 전통적 강자
소프트뱅크	9984 JP	6,424.00	341	18.4	68.1	60.9	18.6	256.0	46.2	05-13	'비전펀드'는 AI에 집중. 일본 최고의 AI 슈퍼컴퓨터.
하드웨어 인프라											
엔비디아	NVDA US	211.50	7,480	6.0	18.8	14.1	12.4	80.7	13.4	05-20	AI 모델 훈련 및 추론에 필수인 GPU계 현존 최강자
브로드컴	AVGO US	412.56	2,843	-1.2	23.5	24.2	18.5	103.1	19.4	06-03	이더넷 기반 네트워킹 반도체의 최강자
AMD	AMD US	408.46	969	15.2	84.4	96.0	74.9	307.0	90.7	08-05	서버용 CPU의 최강자. AI 가속기 분야 패스트팔로어
인텔	INTC US	109.62	802	16.0	107.2	116.7	187.5	439.7	197.1	07-24	파운드리로서 소버린 AI 미 지정학적 가치 부상
마이크론	MU US	646.63	1,061	25.0	71.3	63.9	172.0	684.6	126.7	06-25	SK하이닉스 추격 중인 "미국"의 메모리 반도체
SK하이닉스	000660 KS	1,619,000	1,154	25.1	55.2	89.0	183.7	755.5	148.7	07-23	HBM 부문 전세계 최강자
시놉시스	SNPS US	505.19	141	4.7	27.0	18.3	28.4	5.1	7.6	05-27	반도체 EDA 부문 리딩 업체, AI 에이전트 적극 활용
케이던스 디자인 시스템	CDNS US	356.90	143	8.3	27.7	25.9	9.8	15.7	14.2	07-28	반도체 EDA 부문 리딩 업체
암페놀	APH US	136.62	245	-7.2	6.4	0.5	-1.4	70.2	1.3	07-29	엔비디아향 AI용 서버 내 고속 I/O 및 커넥터 제조
버티브	VRT US	340.01	190	3.5	29.6	73.9	89.2	256.4	109.9	07-30	액체 냉각 방식에 있어 글로벌 선두업체
Arm	ARM US	213.31	330	1.4	48.3	72.4	40.0	71.8	95.1	07-30	AI 가속기 설계 위한 다수의 IP를 소유한 팹리스
마벨 테크놀로지스	MRVL US	160.01	204	-3.1	46.4	99.4	76.2	184.9	88.5	05-27	데이터 인프라용 네트워킹 및 스토리지 ASIC 설계
코어위브	CRWV US	128.84	102	15.4	51.1	43.2	23.9	140.4	79.9	05-07	AI용 클라우드 인프라 제공업체
중과서광	603019 CH	98.75	31	6.2	19.1	13.8	-11.1	52.3	15.4	08-28	중국과학원과 긴밀 관계를 유지 중인 HPC 제조업체
SMIC	981 HK	76.75	139	8.3	36.7	13.5	1.7	70.0	7.4	05-08	7나노 공정을 달성한 중국 유일의 미세공정 파운드리
TSMC	2330 TT	2,310.00	2,776	8.2	18.5	30.2	59.3	155.4	49.5	07-17	명실상부 반도체 파운드리 최강자
폭스콘	2317 TT	253.50	165	15.5	25.8	17.9	3.9	79.8	10.0	05-14	GPU 기반 보드 및 서버 시스템 조립 담당, 대만 업체
위스트론	3231 TT	146.00	22	6.6	11.5	13.2	6.2	41.2	-3.0	05-08	GPU 기반 보드 및 서버 시스템 조립 담당, 대만 업체
Alchip	3661 TT	4,795.00	18	16.0	67.7	53.9	34.5	106.4	36.6	05-08	빅테크들의 AI 가속기 설계를 돕는 대만의 팹리스
에너지 인프라											
GE 베르노바	GEV US	1,045.63	409	-3.5	14.8	34.2	82.0	157.6	60.2	07-23	AI 데이터센터용 가스터빈 인프라 솔루션
슈나이더 일렉트릭	SU FP	275.20	272	2.5	18.0	8.8	19.2	32.5	17.2	07-30	AI 데이터센터 전력관리 및 냉각시스템 통합 솔루션
비스트라	VST US	153.95	76	-2.5	0.2	3.0	-19.2	10.9	-4.4	05-07	미국 최대 독립 전력 생산 사업자
블룸 에너지	BE US	258.64	107	-8.7	90.3	80.8	91.3	1,464.7	197.7	07-31	AI 데이터센터용 연료전지 솔루션
넥스트에라 에너지	NEE US	93.32	283	-4.7	-0.4	5.0	12.7	42.7	17.0	07-23	AI 데이터센터용 엔드투엔드 전력 파트너

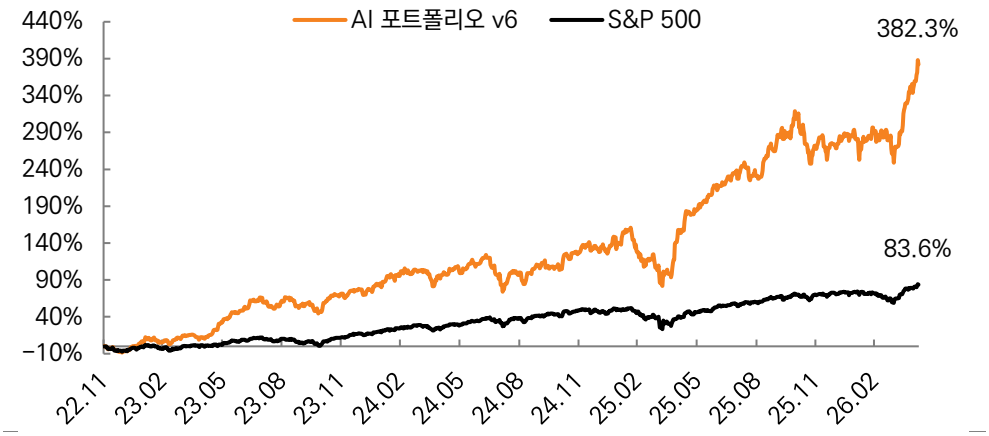
자료: Bloomberg, 미래에셋증권 리서치센터
주: 실적발표 일정은 변경될 수 있음.

III. Appendix: Charts

그림 14. ChatGPT 출시일 이후, S&P 500에서 Magnificent Seven이 차지하는 비중(시가총액 기준)

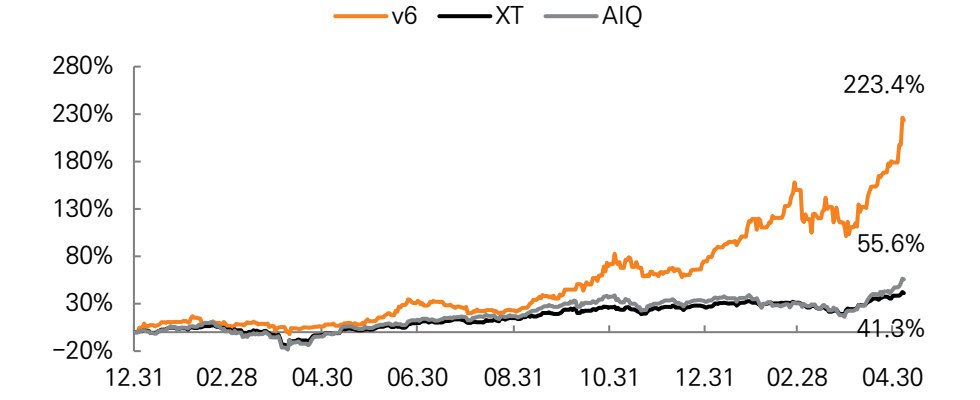
자료: Bloomberg, 미래에셋증권 리서치센터

그림 15. ChatGPT 출시일 이후, AI 밸류체인에 속한 유니버스 종목들의 평균 수익률(동일가중 방식)



자료: Bloomberg, 미래에셋증권 리서치센터

그림 16. 대표 AI ETF vs AI 유니버스(v.6) 수익률 추이(24.12.31 기준)
- Global X Artificial Intelligence & Technology ETF(AIQ), iShares Exponential Technologies ETF(XT)



자료: Bloomberg, 미래에셋증권 리서치센터

표 8. SimpleBench 기준 성능 상위 모델 10개(동일 모델 생략)

순위 (전 주 변동)	모델명	평균 점수(%)	기관명	라이선스	지식 컷오프
-	인간 최고 점수	95.4	-	-	-
-	인간 평균	83.7	-	-	-
1(-)	Gemini 3.1 Pro Preview	79.6	구글	독점 폐쇄소스	2025년 1월
2(New)	GPT-5.5 Pro	76.9	OpenAI	독점 폐쇄소스	2025년 12월
3(▼1)	Gemini 3 Pro Preview	76.4	구글	독점 폐쇄소스	2025년 1월
4(New)	GPT-5.5	69.0	OpenAI	독점 폐쇄소스	2025년 12월
5(▼2)	Claude Opus 4.6	67.6	앤트로픽	독점 폐쇄소스	2025년 8월
6(▼2)	Gemini 2.5 pro(06-05)	62.4	구글	독점 폐쇄소스	2025년 1월
7(▼2)	Claude Opus 4.5	62.0	앤트로픽	독점 폐쇄소스	2025년 8월
8(New)	Claude Opus 4.7	61.7	앤트로픽	독점 폐쇄소스	2026년 1월
9(▼3)	GPT-5 Pro	61.6	OpenAI	독점 폐쇄소스	2024년 9월
10(▼3)	Gemini 3 Flash Preview	61.1	구글	독점 폐쇄소스	2025년 1월

자료: simple-bench.com, 각 사, 미래에셋증권 리서치센터
주: SimpleBench는 인간에게는 쉽지만 기계에게는 어려운 문제로 이루어진 벤치마크

표 9. LLM 벤치마크 챗봇아레나의 Hard Prompts 기준 성능 상위 모델 10개(동일 모델 생략)

순위 (전 주 변동)	모델명	전체 평균 점수	기관명	라이선스	지식 컷오프
1(-)	claude-opus-4-6-thinking	1535	앤트로픽	독점 폐쇄소스	2025년 8월
2(-)	claude-opus-4-6	1531	앤트로픽	독점 폐쇄소스	2025년 8월
3(New)	claude-opus-4-7-thinking	1530	앤트로픽	독점 폐쇄소스	2026년 1월
4(New)	claude-opus-4-7		앤트로픽	독점 폐쇄소스	2026년 1월
5(▼2)	gemini-3.1-pro-preview	1513	구글	독점 폐쇄소스	2025년 1월
6(▼2)	muse-spark	1509	메타	독점 폐쇄소스	2026년 1월
7(▼1)	gemini-3-pro	1504	구글	독점 폐쇄소스	2025년 1월
8(New)	gpt-5.5-high	1504	OpenAI	독점 폐쇄소스	2025년 12월
9(▼4)	gpt-5.4-high	1502	OpenAI	독점 폐쇄소스	2025년 8월
10(-)	claude-sonnet-4-6	1500	앤트로픽	독점 폐쇄소스	2025년 8월

자료: lmarena.ai, 각 사, 미래에셋증권 리서치센터
주: Hard Prompts는 기존의 일반적인 요청보다 훨씬 더 까다롭고 복잡한 문제를 제시하여 LLM의 한계를 테스트하는 벤치마크

표 10. LLM 벤치마크 LiveBench의 성능 상위 모델 10개(동일 모델 생략)

순위 (전 주 변동)	모델명	전체 평균 점수	기관명	라이선스	지식 컷오프
1(New)	GPT-5.5 Thinking xHigh Effort	80.71	OpenAI	독점 폐쇄소스	2025년 12월
2(▼1)	GPT-5.4 Thinking xHigh Effort	80.28	OpenAI	독점 폐쇄소스	2025년 8월
3(▼1)	Gemini 3.1 Pro Preview High*	79.93	구글	독점 폐쇄소스	2025년 1월
4(▼1)	Claude 4.7 Opus Thinking xHigh Effort	76.91	앤트로픽	독점 폐쇄소스	2026년 1월
5(▼1)	Claude 4.6 Opus Thinking High Effort	76.33	앤트로픽	독점 폐쇄소스	2025년 8월
6(▼1)	Claude 4.5 Opus Thinking High Effort	75.96	앤트로픽	독점 폐쇄소스	2025년 7월
7(▼1)	Claude 4.6 Sonnet Thinking Medium Effort	75.47	앤트로픽	독점 폐쇄소스	2025년 8월
8(▼1)	GPT-5.2 High	74.84	OpenAI	독점 폐쇄소스	2025년 8월
9(▼1)	GPT-5.2 Codex	74.30	OpenAI	독점 폐쇄소스	2025년 8월
10(▼1)	GPT-5.1 Codex Max	73.98	OpenAI	독점 폐쇄소스	2024년 9월

자료: LiveBench, Huggingface, 미래에셋증권 리서치센터

표 11. 코딩 관련 벤치마크 SWE-Bench의 성능 상위 모델 10개

순위 (전주 변동)	모델명	점수(%)	기관명	등록일
1(-)	Claude 4.5 Opus (high reasoning)	76.80	앤트로픽	2026-02-17
2(-)	Gemini 3 Flash (high reasoning)	75.80	구글	2026-02-17
3(-)	MiniMax M2.5 (high reasoning)	75.80	미니맥스	2026-02-17
4(-)	Claude Opus 4.6	75.60	앤트로픽	2026-02-17
5(-)	GLM-5 (high reasoning)	72.80	Zhipu AI	2026-02-17
6(-)	GPT-5-2 (high reasoning)	72.80	OpenAI	2026-02-19
7(-)	GPT 5.2 Codex	72.80	OpenAI	2026-02-17
8(-)	Claude 4.5 Sonnet (high reasoning)	71.40	앤트로픽	2026-02-17
9(-)	Kimi K2.5 (high reasoning)	70.80	Moonshot AI	2026-02-17
10(-)	DeepSeek V3.2 (high reasoning)	70.00	DeepSeek	2026-02-17

자료: SWE-Bench, 미래에셋증권 리서치센터
주: 동일 모델이더라도, 추론 설정 등의 변경으로 점수가 재산정되는 경우 새로운 항목으로 판단. 등록일 업데이트를 참고

표 12. 지식 환각 벤치마크인 AA-Omniscience의 성능 상위 모델 10개

순위 (전주 변동)	모델	점수	환각률(%)	기관명	총 토큰 사용량 및 비용
1(-)	Gemini 3.1 Pro Preview	33	50%	구글	8.7M / $106
2(-)	Claude Opus 4.7(max)	26	36%	앤트로픽	9M / $212
3(New)	GPT-5.5(xhigh)	20	-	OpenAI	31M / $905
4(New)	GPT-5.5(high)	18	86%	OpenAI	-
5(New)	Grok 4.3	18	25%	xAI	18M / $45
6(New)	GPT-5.5(medium)	17	86%	OpenAI	-
7(New)	GPT-5.5(low)	15	86%	OpenAI	-
8(New)	Claude Opus 4.7(Non-reasoning, high)	14	36%	앤트로픽	-
9(▼4)	Claude Sonnet 4.6(max)	12	46%	앤트로픽	17M / $263
10(▼4)	Gemini 3 Flash	12	92%	구글	11M /$32

자료: Artificial Analysis, 미래에셋증권 리서치센터
주: AA-Omniscience는 모델의 환각을 평가하는 벤치마크. 정답을 맞추면 가점을, 환각에는 감점을 하며 답변 거부에는 페널티를 주지 않음. 점수는 -100점부터 100점 사이로, 0점은 정답과 오답 수가 같음을 의미. 음수는 오답이 정답보다 많았음을 의미.
*환각률 = 답변하지 않거나 모른다고 해야하는 문제에 대해, 답변을 (환각으로) 제공한 비율
= 부정확한 답변 / (부정확한 답변 + 부분 답변 + 시도하지 않음)

Compliance Notice

- 당사는 자료 작성일 현재 조사분석 대상법인과 관련하여 특별한 이해관계가 없음을 확인합니다.
- 당사는 본 자료를 제3자에게 사전 제공한 사실이 없습니다.
- 본 자료를 작성한 애널리스트 김은지은(는) 자료작성일 현재 알파벳 50주 보유하고 있습니다.
- 본 자료는 외부의 부당한 압력이나 간섭없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.

본 조사분석자료는 당사의 리서치센터가 신뢰할 수 있는 자료 및 정보로부터 얻은 것이나, 당사가 그 정확성이나 완전성을 보장할 수 없으므로 투자자 자신의 판단과 책임하에 종목 선택이나 투자시기에 대한 최종 결정을 하시기 바랍니다. 따라서 본 조사분석자료는 어떠한 경우에도 고객의 증권투자 결과에 대한 법적 책임소재의 증빙자료로 사용될 수 없습니다. 본 조사분석자료의 지적재산권은 당사에 있으므로 당사의 허락 없이 무단 복제 및 배포할 수 없습니다.