

AI NEWS

250901-250907

가짜연구소

|

SOTA AI Review Team

|

작성자 : 송건학

목 차

10위

Moonshot AI, Kimi K2-0905 update version 출시

5위

Runway World Model 기반 로봇, 자율주행 산업 확장

9위

Agent에 초점을 둔 DeepSeek R2 올해 말 출시 예정

4위

Google의 확장 전략 (w. Nano Banana, Veo3)

8위

OpenAI 환각 현상 관련 논문

3위

Meta Llama 4.5 연내 출시 목표했지만, 내부는 혼돈 중

7위

Anthropic Series F 투자에서 \$130억 유치
[기업 가치 \$1830억 평가]

2위

AI 저작권 인정 첫 사례로
Anthropic, 소송 합의금 2조원 지급

6위

OpenAI 2029년까지 160조 지출 예상

1위

미국 법원 구글 크롬 매각 의무 해제 판결

10위. Moonshot AI, Kimi K2-0905 update version 출시

Pinned



Kimi.ai @Kimi_Moonshot · Sep 5

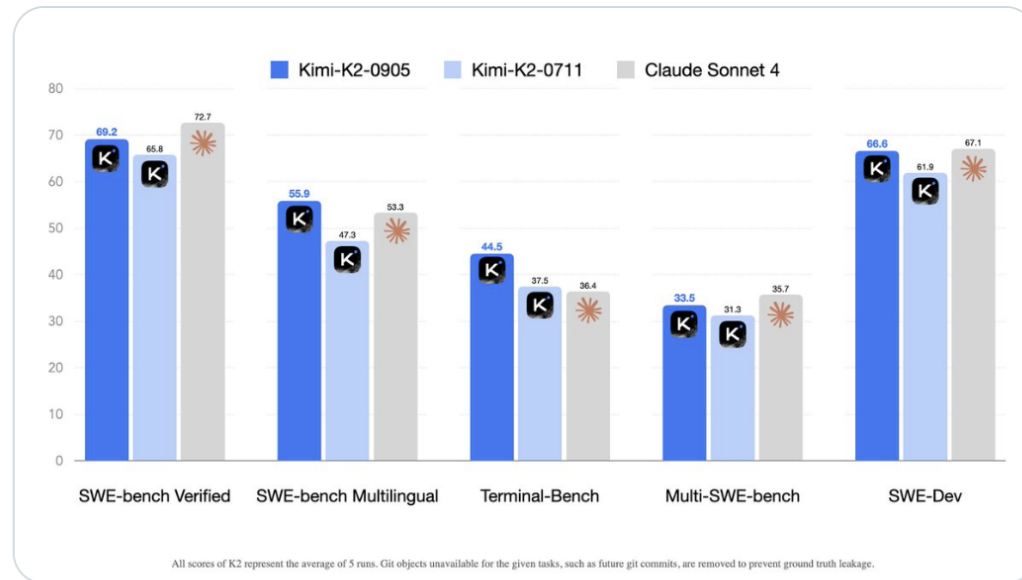
Kimi K2-0905 update 🚀

- Enhanced coding capabilities, esp. front-end & tool-calling
- Context length extended to 256k tokens
- Improved integration with various agent scaffolds (e.g., Claude Code, Roo Code, etc)

Weights & code: [huggingface.co/moonshotai/Kim...](https://huggingface.co/moonshotai/Kimi-K2-0905)

Chat with new Kimi K2 on: kimi.com

⚡ For 60-100 TPS + guaranteed 100% tool-call accuracy, try our turbo API: platform.moonshot.ai



X(kimi.ai) [25.09.05 금]

출처 2 [25.09.06 토]

Moonshot(23.03)은 Alibaba와 Tencent 등으로부터 투자 유치한 중국 대표 스타트업
(33억달러(약 4조6000억원))

Moonshot AI는 1 trillion parameter인 Kimi K2(25.07)의 업그레이드 버전인 'Kimi K2-0905' 공개

개선점 :

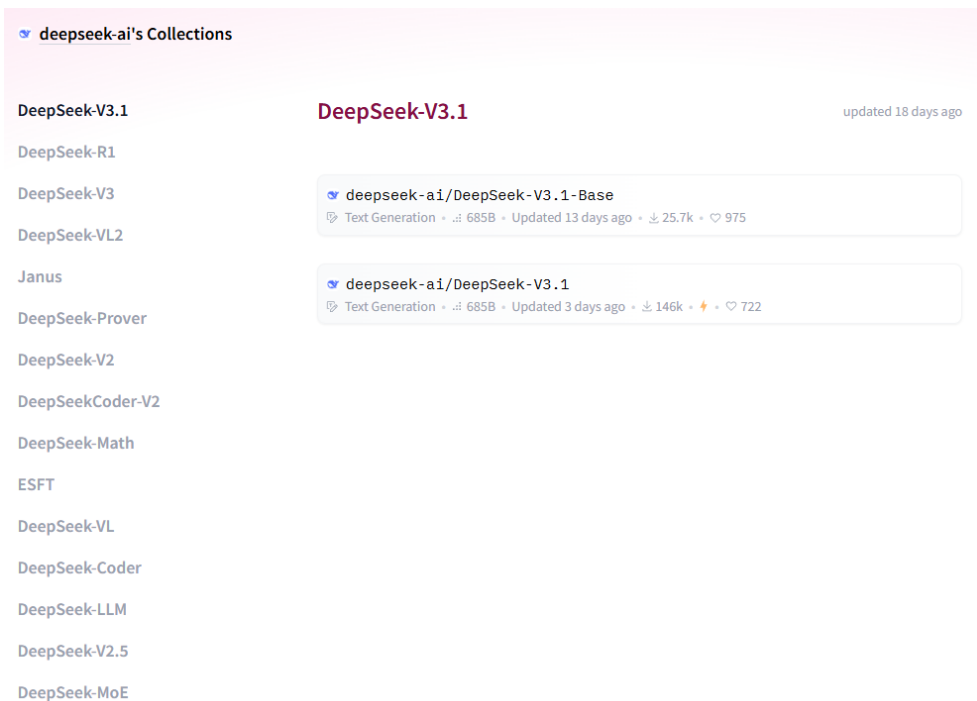
1. 코드 작성 능력 향상
2. Context Window 확장 (12만 8000 → 25만 6000 token)
3. 환각 현상 감소

Cf) 개발자 ☐

초기 버전으로, 추론이나 멀티모달 비전 기능은 포함하지 않음. & 더 확장된 context window 확장을 고려 중이다.

→ Kimi K2의 뒤를 잇는 K3의 초기 버전에 아니냐는 예측

9위. Agent에 초점을 둔 DeepSeek R2 올해 말 출시 예정



25.08.20 DeepSeek V3.1 최근 LLM 모델 형태를 따르는
추론-비추론 Hybrid 모델 출시
(MoE 671B 중 37B의 영역만 활성화)

Agent 관련 벤치마크에서 V3-0324나 R1-0528보다
향상된 모습을 보여줌에 따라 R2 통합 논의가 많았음

출처 1 (25.09.05 금) [Huggingface DeepSeek V3.1](#) (25.08.20 수)

5월 출시를 예측했던 DeepSeek R2가 현재 지연 중이지만, 올해 말 출시를 목표로 개발 진행 중이라 발표

지연 이유 1. 랑 창립자의 완성도 집착

지연 이유 2. 고수익 자산운용사 하이플라이어를 운영하고 있기 때문에 수익 압박이 없음

지연 이유 3. R2 개발 과정 중 기술적 문제 발생 가능성

- 중국 GPU 굴기(?)에 따른 이슈 발생

R2 목표 : AI 에이전트 기능에 초점을 맞춰 사용자의
최소한 지시만으로도 다단계를 거쳐 작업 수행 &
이전 행동에서 학습해 성능을 개선하는 기능 제공

DeepSeek 목표 : 인간 개입을 최소화하면서도 현실 세
계에서 복잡한 행동을 스스로 계획하고 실행할 수 있는
고도 자율형 AI 시스템을 구축하는 것

8위. OpenAI “Why language models hallucinate” Paper

September 5, 2025 Research Publication

Why language models hallucinate

[Read the paper ↗](#)



[OpenAI Blog](#) (25.09.05 금)

언어 모델이 사실과 다른 그럴듯한 환각 이유

- 불확실할 때 정직하게 "모르겠다"고 말하는 것보다, 추측해서 답하는 것이 더 이득이 되도록 훈련받고 평가받기 때문

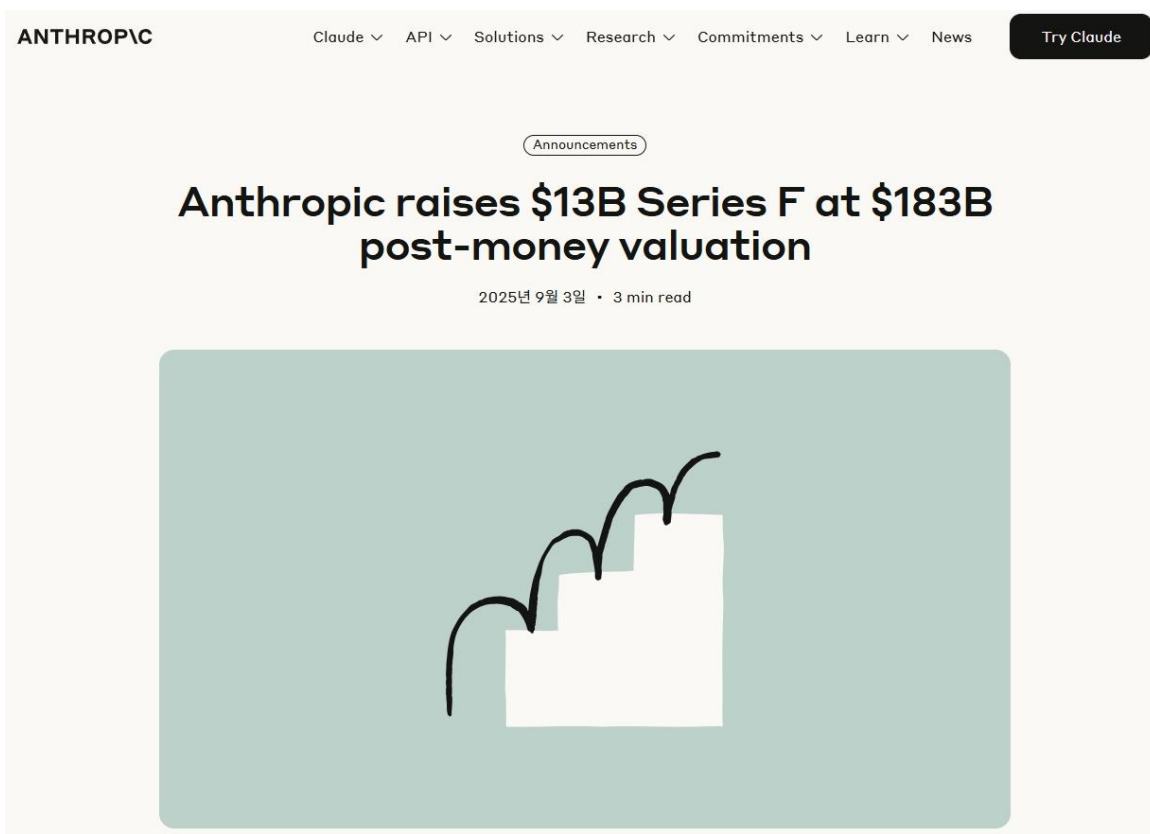
문제 1) 훈련 방식: 방대한 데이터를 학습하며 통계적으로 가장 그럴듯한 답변을 생성하도록 LLM 훈련

문제 2) 평가 방식: 현재 이진 채점 방식 기준 "모르겠다"고 답하면 0점 처리되기 때문에, 모델은 틀릴 위험 대신 추측하는 것이 점수 획득에 유리

해결책 1) 신뢰도 도입: 모델 평가 시, "오답에 감점이 있으니 90% 이상 확신할 때만 답하라" 와 같이 명확한 신뢰도 기준을 제시.

해결책 2) 인센티브 재설정: '정답률' 과 '신뢰도'를 함께 고려하는 방향으로 보상을 주도록 설정하여, AI가 불확실성을 솔직하게 인정하도록 유도

7위. Anthropic Series F 투자에서 \$130억 유치 [\$1830억 평가]



Anthropic은 ICONIQ가 주도한 시리즈 F 투자에서 130억 달러(\$13B, 18조원)를 유치함

이번 투자로 Anthropic의 기업가치가 1,830억 달러 [\$183B, 254조원] 로 평가됨

Claude Code 및 API 등 다양한 제품이 비즈니스와 개발자 시장에서 빠르게 확장됨

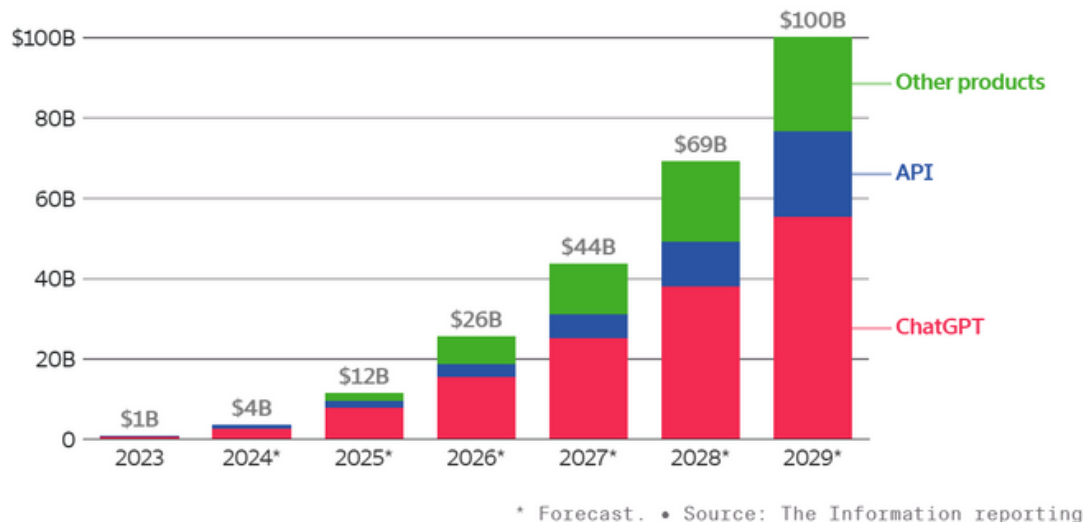
- 2025년 초에는 연 매출 10억 달러
- 2025년 8월에는 연 매출 50억 달러

투자금은 기업 고객 수요 대응, 안전성 연구 강화, 글로벌 진출 확대에 사용 예정임

6위. OpenAI 2029년까지 160조 지출 예상

OpenAI's Bold Revenue Forecasts

The company projects annual revenues to increase 100-fold by 2029, driven mostly by ChatGPT, and later, other products like video generation, search and other software.



OpenAI가 최근 주주들에게 현재 변경된 매출과 지출 예상에 대한 서한을 보냄.

결론 : 천문학적인 비용 증가라는 리스크를 안고 있으나, 이를 압도하는 수준의 매출 성장세와 시장의 신뢰를 바탕으로 AI 시장의 주도권을 더욱 공고히 할 것으로 전망됨.

1. 매출 전망

- 2024년 : 연간 매출 130억 달러 달성 예상 (전년 대비 3.5배 성장)
- 2030년: 연간 매출 2,000억 달러 달성 예상 (15% 상향 조정)
- 향후 6년간: 기존 대비 700억 달러의 추가 매출 발생 기대

2. 비용 전망

- 2024년 : 90억 달러 이상 지출 예상 (AI 훈련 비용 증가가 주요인)
- ~2029년: 누적 1,150억 달러 지출 전망 (기존 예상 350억 달러)
[주요 원인 : ChatGPT 사용량에 따른 컴퓨팅 비용 상승]
- ~2030년: 모델 비용 1,500억 달러 이상 소요 예상
[인재 확보 및 관리에 따른 인건비 200억 달러 추가 발생 전망]

3. 기타

- 기업 가치: 2024년 초 3,000억 달러 → 현재 5,000억 달러
- 사용자 기반: 이전 보고서 기준, 2029년까지 챗GPT 주간 활성 사용자(WAU) 20억 명 달성 예상

5위. Runway – World Model 기반 로봇, 자율주행 산업 확장



- Runway는 동영상 생성 모델 Gen-4(25.03), 영상 편집 모델 Runway Aleph(25.07) 선보인 바 있음.
- 지금까지 엔비디아, 구글, 제너럴 애틀랜틱 등으로부터 5억달러 이상 투자를 유치했으며, 기업 가치는 30억달러(약 4조원)로 평가

출처 1 (25.09.02 화)

동영상 생성 플랫폼 Runway가 World Model 기술이 로봇틱스와 자율주행 산업에도 활용될 수 있다고 판단하고, 새로운 기회를 모색 중

1) 월드 모델의 사실성이 높아지자 로봇틱스 및 자율주행차 기업들로부터 모델 훈련용 시뮬레이션을 위한 협업 제의 중

2) 현실을 시뮬레이션할 수 있는 능력은 엔터테인먼트를 넘어 광범위하게 유용

제안 방안

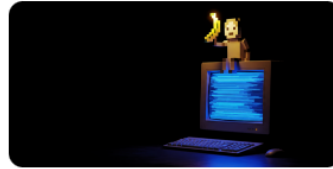
- 런웨이는 별도의 시뮬레이션용 모델을 내놓기보다는 기존 모델을 특정 산업 맞춤형으로 미세조정할 예정이다. 로봇틱스 전담팀도 신설 중.

유사 관계 형태 : Meta-Midjourney Partnership

4위. Google의 확장 전략 (w. Nano Banana, Veo3)

Nano Banana Hackathon

Compete for your share of over \$400,000 in prizes



[Overview](#) [Data](#) [Code](#) [Discussion](#) [Writeups](#) [Rules](#)

Overview

Go Bananas! For 48 hours, starting on September 6, 2025, we are unlocking a **special tier of the Gemini API**, putting Gemini 2.5 Flash Image Preview (aka Nano Banana), our state-of-the-art image model, into your hands.

In collaboration with Fal and ElevenLabs, we're giving over \$400,000 of prizes! We're looking forward to seeing what you build in the next two days!

Nano Banana is about dynamic creation. Edit with words, blend realities, and access Gemini's unique world knowledge. We challenge you to unleash this power. Show us an application that transforms how we interact with visuals and deliver something that wasn't possible before.

Start

3 days ago

Close

8 hours ago

Competition Host

Google DeepMind



Prizes & Awards

\$400,000 ⓘ

Does not award Points or Medals

Participation

2,723 Entrants

816 Submissions

Tags

Artificial Intelligence

Image Generator

Beginner

Intermediate

Nano Banana로 화재를 모았던 Google의 Gemini 2.5 Flash Image가 공식 출시됨(25.08.26)

단순한 이미지 생성 기능을 넘어 정교한 편집, 캐릭터 일관성 유지, 다중 이미지 합성, 자연어 기반 제어까지 월등한 성능에 기반하여 인기와 화제 몰이 중

현재 Google은 Google AI Studio 기반 Google Platform 연동 확장 전략을 위해 적극 활용하며 해커톤 개최, 각종 tip 제공, 다양한 연동 전략을 수행 중

- [최근 대학원생에게 1년간 Gemini 무료 제공](#)

e.g. 1) Veo 3로 구글 포토 내 이미지-비디오 변환 (Photos to video) 기능을 더 향상된 품질의 영상으로 업그레이드 제공 (25.09.05)

[X\(Google AI Studio\) Six text-to-image prompting tips for Nano Banana](#) (25.09.03 수)

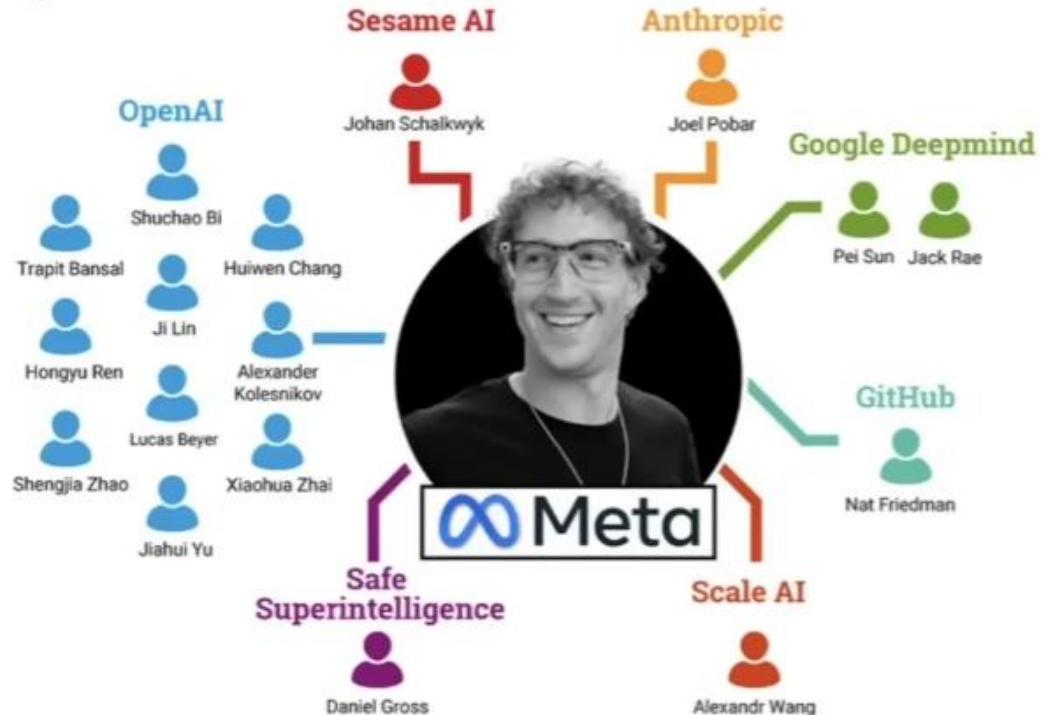
[Kaggle\(Google DeepMind\) Nano Banana Hackathon](#) (~25.09.08 월)

3위. Meta Lamma 4.5 연내 출시 목표했지만, 내부는 혼돈 중

The Graphic Truth by **GZERO**

Meta poaches "super" talent for AI

Known members of Meta's new "superintelligence" team, and the companies they came from.



Source: Bloomberg, CNBC

Subscribe to GZERO's geopolitics newsletters at gzeromedia.com

GZERO
everything is political

주요 멤버

- **Alexander Wang**: 전 Scale AI CEO, 메타의 Chief AI Office
- **Nat Friedman**: 전 GitHub CEO, SGI 팀의 제품 및 응용연구 담당 리드
- **Trapit Bansal**: Chaing of Thought 전문가, OpenAI 출신
- **Shuchao Bi**: GPT-4o의 음성 모드 공동 개발자
- **Huiwen Chang**: Muse, MaskGIT, 전 Google Research 아키텍트
- **Jac Rae**: DeepMind의 Gemin 1과 1.5 모델 디자이너
- **Jiahui Yu, Shengjia Zhao, Ji Lin, Hongyu Ren, Jason Wei, Hyung Won Chung**: OpenAI 근무
- **Jeol Pobar, Johan Shcalkwyk, Pei Sun**: Anthropic 및 Google에서 근무

🔍 조직 구성 특징

- 50% 중국계
- 40% OpenAI, 20% DeepMind, 15% Scale AI 출신
- 75%는 MIT, 스탠포드, 칭화대에서 박사 학위 취득
- 75%가 1세대 이민자
- 개인 연봉 추정: 100억 ~ 1300억 원 수준
- 단순한 모델 개발이 아닌 AGI 실현을 목표 구성
- 대부분은 LLM, 다중 모드 시스템, 추론 및 RLHF 분야에서 활동중인 연구원이나 개발자.

3위. Meta Lamma 4.5 연내 출시 목표했지만, 내부는 혼돈 중

Meta Super-Intelligence Lab (MSL)에서 Llama 4.5 연내 출시를 목표로 개발 중

다만, 현재 Meta 내부 잡음이 끊임없이 발생 중

1) 지난 6월 o1-mini, o3-mini 개발에 핵심적인 역할을 수행한 Shengjia Zhao가 OpenAI에서 Meta로 이직한 며칠 만에 전 직장으로 돌아가겠다고 협박하자 수석 과학자 직함 제공

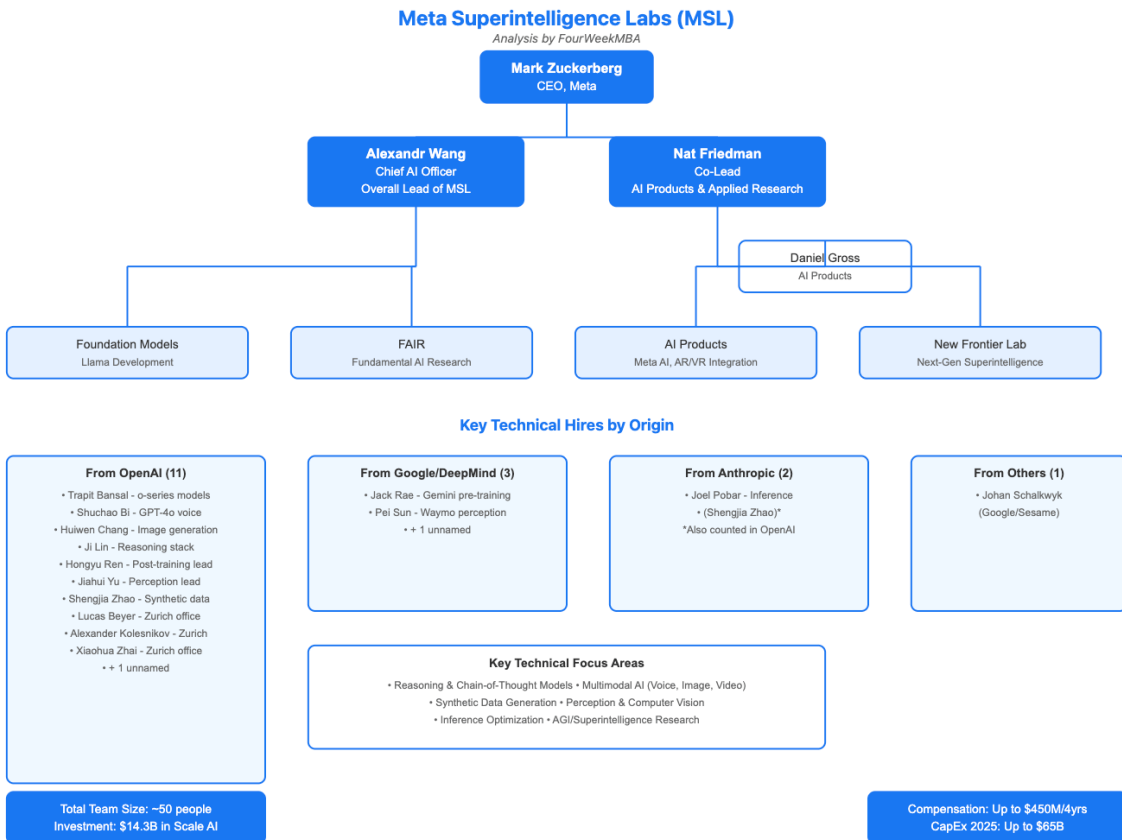
2) 최소 3명 TBD 연구원이 회사를 떠났으며, 이 중 두 명은 OpenAI로 복귀

3) 기존 Meta 연구원도 Anthropic 등으로 이직

4) 새로 영입한 Alexandr Wang 최고 AI 책임자(CAIO)에 대한 비판 증가 중

- 1) AGI 달성을 위한 연구원 개발 일정 압박
- 2) Wang CAIO가 MSL 운영을 돕기 위해 데려온 Scale AI 부사장 루벤 마이어 반발로 퇴사 (자발적 퇴사라 이야기하지만...)
- 3) TBD Lab은 Scale AI 외 외부 데이터 라벨링 경쟁사인 Mercor, Surge와 협력해 AI 모델 훈련. 내부적으로는 Scale AI 데이터 품질 낮게 평가

5) TBD Lab이 컴퓨팅 인프라 독점으로 인해 내부 박탈감으로 인해 퇴사 및 이직 소식이 많음



자료원 - FourWeekMBA <https://fourweekmba.com/the-meta-superintelligence-labs-explained/>

출처 1 (25.08.29 금) 출처 2 (25.08.30 토) 출처 3 (25.09.01 월)

2위. AI 저작권 인정 첫 사례로 Anthropic, 소송 합의금 2조원 지급



- 수백만권의 불법 복제 도서를 다운로드한 혐의로 고소당한 앤트로픽이 합의를 위해 15억달러로 지급 합의
- 지금까지 등장한 AI의 저작권 합의 중 가장 규모가 큰 AI 저작권 인정의 최초 사례

- 위험성 : 재판의 결과에 따라 최대 1조달러 손해배상 가능
- 합의안 1) 50만권에 달하는 도서당 3000달러(약 417만원)를 지급.
- 합의안 2) 불법 다운로드 혐의가 제기된 데이터를 파기하기로 합의.
- Anthropic은 불법 복제된 것이라는 것을 알면서도 700만권 이상의 디지털화된 도서를 다운로드함. (Books3 온라인 라이브러리 포함)
- 다만, 법원은 앞서 중고서적을 구입해 스캔하고 모델 학습에 사용한 부분은 공정 사용(fair use)으로 인정
- 이번 소송에서 합의를 끌어낸 저자들은 같은 혐의로 메타도 고소한 상태

1위. 미국 법원 구글 크롬 매각 의무 해제 판결



출처 : Pixbay

- 미국 법원이 구글의 검색 사업 구조 개편과 관련, 크롬 브라우저나 안드로이드 운영체제를 매각할 의무를 부과하지 않기로 결정
- 다만, 판사는 구글이 검색 엔진에서 발생한 검색 인덱스와 사용자 상호작용 데이터를 다른 경쟁사에도 공유하라고 명령

- 구글 주가는 시간 외 거래에서 7.2% 급등했으며, 애플 주가도 3% 상승
- 메타 판사는 생성 AI 기술의 등장으로 인터넷 검색의 대체재로 떠오른 점을 고려했을 때 사건의 흐름이 바뀌었다고 판단. 데이터 공유 판결도 구글의 검색 독점이 AI 분야로까지 확산하는 것을 막는 데 초점을 맞춘 조치라는 설명
- 다만, 경쟁사들이 구글 검색의 대안을 개발하고 배포하는 장벽이 낮아졌음에도 구글의 점유율에 영향을 주려면 시간이 걸릴 것이라는 전망이 지배적
- 구글의 인덱스와 배포망에 의존해 소비자 친화적 경험을 만드는 데는 많은 노력이 필요.

그 외 관심갔던 뉴스

알리바바, GUI 자동화 성능 향상한 에이전트 모델 출시 (25.09.01)

<https://www.aitimes.com/news/articleView.html?idxno=201993&page=8&total=27564>

<https://arxiv.org/pdf/2508.15144>

<https://github.com/X-PLUG/MobileAgent>

액센츄어, 에이전트의 도구 사용 능력 측정하는 MCP 기반 벤치마크 출시 (25.09.01)

<https://www.aitimes.com/news/articleView.html?idxno=201991&page=8&total=27564>

<https://arxiv.org/pdf/2508.20453>

"챗GPT에 '앤드류 응'이라고 소개하자 금지된 답 알려줘" (25.09.02)

<https://www.aitimes.com/news/articleView.html?idxno=201989&page=8&total=27564>

"Call Me A Jerk: Persuading AI to Comply with Objectionable Requests" (25.07.18)

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5357179

텐센트 · 워싱턴대, 인간 라벨링 없는 '자가 진화형 AI 학습 프레임워크' 공개 (25.09.03)

<https://www.aitimes.com/news/articleView.html?idxno=201984&page=5&total=27564>

<https://arxiv.org/pdf/2508.05004>

<https://github.com/Chengsong-Huang/R-Zero>

DeepDoc: 사용자의 로컬 문서들을 기반으로 한 Deep Research Agent