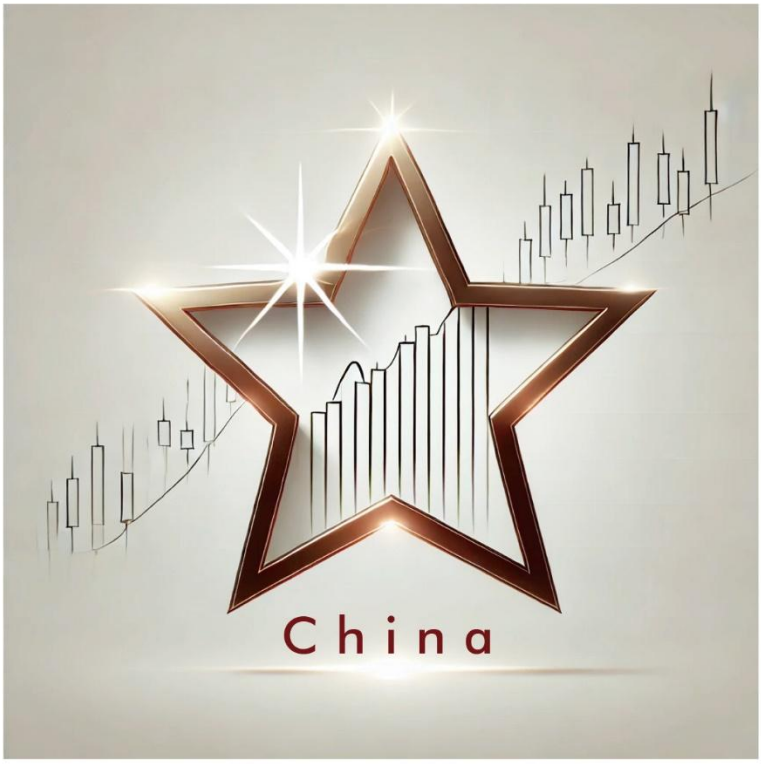
YUANTA TOP FOCUS

No. 5

12, February 2026 **Yuanta Research**

# 시장의 별을 찾아서




**백종민** 방산/AI/로보틱스
02 3770 5598
jongmin.baik@yuantakorea.com

**배종성** Research Assistant
02 3770 3643
jongsung.bae@yuantakorea.com

유안타증권 Yuanta Securities (Korea)

YUANTA


# TOP FOCUS

시장의 별을 찾아서

# SUMMARY

## 2026년에 주목할 중국 AI 투자 포인트

**AI 풀스택 역량 보유 여부에 주목**해야 하는 이유는, AI 병목은 여전히 연산 인프라에 존재하기 때문이다. 이에 따라, AI 비용 상승으로 마진 압박을 받는 글로벌 소프트웨어 기업들의 주가가 급격하게 하락하고 있다. AI 밸류체인 내 인프라 비용 효율화가 가능한 기업 선별이 필요한 시점이다.

**중국 AI 산업의 관전 포인트**는 다음과 같다: 1) AI 칩 및 연산 인프라 내재화, 2) 대규모 AI 추론 수요에 따른 클라우드 인프라 보유 역량, 3) 파운데이션 모델과 수익 플랫폼 연계가능성. 위 3가지 특징을 2026년 1월에 상장한 Zhipu AI와 Minimax의 글로벌 최초 AI-Native 기업 IPO 사례를 통해 확인하고자 한다.

**중국 AI 투자 포인트**로 다음 2가지를 제시한다:

**1. 엔비디아 H200 확보 불확실성 속에서 피어난 중국 AI 전략:** 미국의 엔비디아 H200 칩 수출제한은 1) 중국의 AI 칩 국산화 가속과 2) AI 칩 내재화 전략이라는 결과를 가져왔다. 알리바바의 핑터우거, 바이두의 쿤룬신이라는 2개의 반도체 자회사 Case Study를 통해 해당 내용을 파악하고자 한다.

**2. 모델 아키텍처 고도화를 통한 추론 비용 절감:** 제한된 AI 하드웨어 연산 환경은 오로지 모델 아키텍처 고도화를 통해 추론 성능을 높이는 결과를 가져온다. DeepSeek의 Engram 모듈 사례를 통해 이를 살펴보고자 한다.

**중국 AI 투자전략**으로 'AI칩 – 클라우드 – 파운데이션 모델 – 플랫폼 – 어플리케이션'으로 이어지는 AI 전 밸류체인 수직계열화를 완료한 AI 풀스택 기업에 주목한다.

**Top Picks: 알리바바(9988.HK), 바이두(9888.HK)**

# CONTENTS

TOP FOCUS

China

# Software/AI


**백종민**

방산/AI/로보틱스

02 3770 5598
jongmin.baik@yuantakorea.com

**배종성**

Research Assistant

02 3770 3643
jongsung.bae@yuantakorea.com

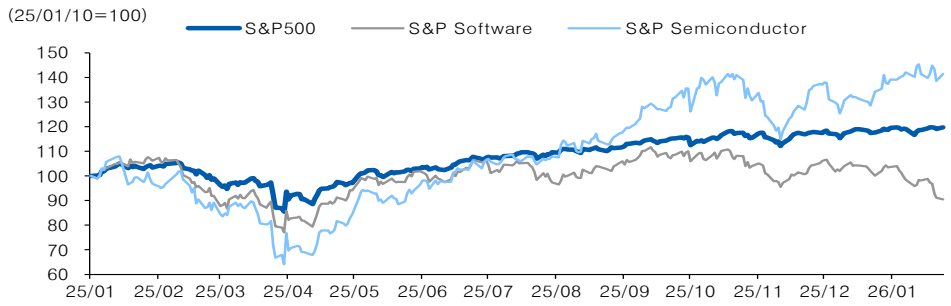
# AI 풀스택 역량 보유 여부 주목

## AI 병목에 따른 글로벌 소프트웨어 주가 부진

AI 산업의 폭발적인 성장과 함께 반도체, 전력 등의 인프라 병목은 여전히 해소되지 않은 상태다. AI 인프라 공급 부족에 의한 연산 비용 상승으로 소프트웨어 기업의 마진 압박 우려가 리스크로 부상하고 있으며, 이로 인해 주가 부진이 이어지고 있다.

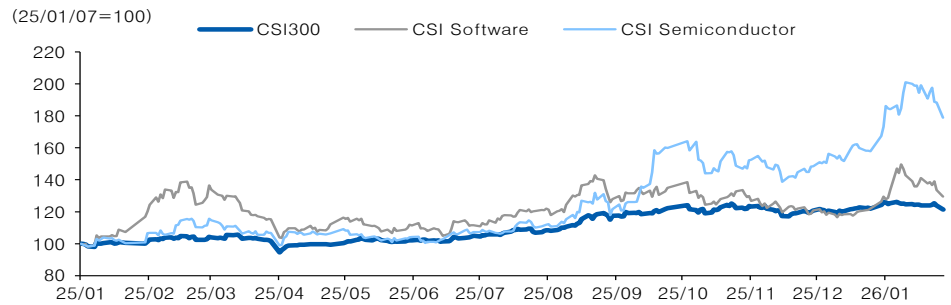
이는 4Q25 실적 발표 이후 흐름을 통해서도 확인된다. Meta는 공격적인 CAPEX 확대에도 AI 기반 맞춤형 광고를 통한 수익성 개선으로 매출(+24% YoY)과 EPS 모두 컨센서스를 상회하며 시간 외 시장에서 주가가 10% 넘게 급등하였다. 반면, 비용 전가가 제한적인 기업들은 조정을 피하지 못했다. 따라서, AI 밸류체인 내 AI 인프라 비용 효율화가 가능한 기업 선별이 필요한 시점이다.

[그림 1] AI 인프라 병목에 따른 글로벌 소프트웨어 섹터 주가의 극명한 부진




자료: Bloomberg, 유안타증권 리서치센터

[그림 2] 중국 소프트웨어 섹터 또한 AI 병목에 따른 하방 압력 존재. 단, 벤치마크 대비 높은 퍼포먼스



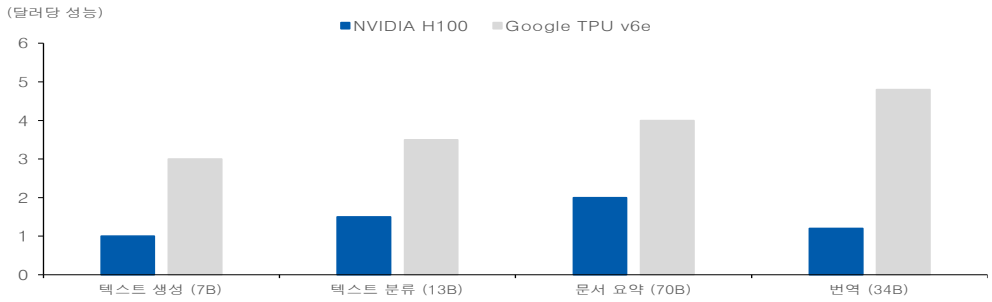

자료: Bloomberg, 유안타증권 리서치센터

## AI 풀스택 역량으로 해결하는 연산 비용

NVIDIA 등이 제시하는 AI 풀스택 기업 정의를 종합해보면, AI 서비스를 외부 의존 없이 통제하기 위해 '연산 인프라-클라우드-파운데이션 모델-플랫폼-어플리케이션'까지 핵심 계층(Layer)을 내재화한 기업을 의미한다. AI 풀스택 역량은 AI 모델 학습과 추론 수요가 폭발적으로 증가하는 국면에서 단위 연산 비용을 낮추고 AI를 지속적인 투자 비용에서 수익 플랫폼으로 전환하는 핵심 조건이다.

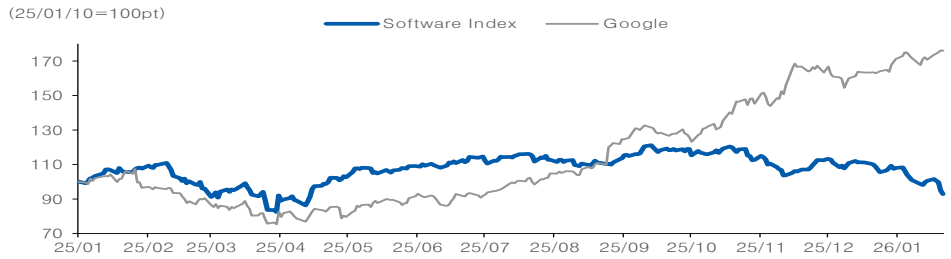
AI 풀스택 역량을 통한 비용 효율화의 대표 사례로 Google AI 생태계가 있다. Google Cloud에 따르면, 동사의 자체 AI 칩 TPU v6e는 상용 추론 솔루션 대비 달러당 최대 4배 높은 성능을 제공한다. 또한 일부 워크로드를 NVIDIA의 GPU 기반 환경에서 TPU로 이전한 경우, 추론 비용이 약 50~65% 절감된 것으로 나타났다. 이는 NVIDIA의 상대적 성능 약화를 의미하지 않는다. AI 칩 내재화를 통해 외부 AI 인프라 제공업체에서 제시하는 가격과 공급 조건을 수용하지 않아도 된다는 것에서 병목 해소가 가능했던 것이다. 외부 GPU와 소프트웨어 스택에 의존할 경우 목표한 성능을 구현하기 위해 더 많은 비용이 필요하므로, 이는 단독 AI 모델 서비스 기업의 밸류에이션을 제약하는 요인이자 투자자들이 우려하는 핵심 리스크이다.

[그림 3] 엔비디아 H100 vs 구글 TPU v6e: 워크로드별 달러당 추론 성능 비교




자료: Pytorch, 유안타증권 리서치센터 / 주: X축 괄호 안은 파라미터 수를 의미. 작업에 따른 스케일링 크기 비교 (7B = 70억개)

[그림 4] 소프트웨어 인덱스 vs 구글 상대주가 추이: AI 풀스택 기업에 대한 밸류에이션 리레이팅




자료: Bloomberg, 유안타증권 리서치센터 / 주: Software Index = IGV ETF

# AI-Native IPO로 확인하는 중국 AI 산업 관전 포인트

### 중국, 글로벌 최초 AI 모델 비즈니스 기업 상장

1/8~1/9 중국 AI 기업 Zhipu AI(2513.HK)와 Minimax(0100.HK)의 홍콩거래소 상장은 글로벌 투자자들의 폭발적인 관심이 확인된다. 공모주 청약에서 Zhipu AI는 일반청약 경쟁률 약 1,159배를 기록하고, MiniMax 또한 1HKD당 개인투자자 약 1,209명이 경쟁하는 과열 양상을 보였다.

금번 IPO의 주목 포인트는 글로벌 최초 MaaS (Model-as-a-Service) 기반 AI-native 기업 상장이라는 점이다. AI 산업은 파운데이션 모델 사전학습을 위한 천문학적인 비용을 필요로 하기에 MaaS 비즈니스의 리스크는 여전히 연산 비용이다. OpenAI, Anthropic, xAI 등 대표적인 비상장 AI-native 기업들도 아직 실적이 본격적으로 반영되지 않았기에 전통적인 밸류에이션 방법론만으로 기업가치를 평가하기엔 한계가 있다. 즉, 모델 기반 비즈니스 특성상 IPO를 포함한 지속적인 자금조달은 필수이며 어떻게 인프라 구축을 효율화할 것인지가 MaaS 비즈니스 밸류에이션의 기준이 될 것으로 판단된다.

### 중국 AI 기업을 평가하기 위한 3가지 관전포인트

Zhipu AI와 Minimax는 상장 후 공모가 대비 각각 +180%, +235%의 수익률(2월 10일 종가 기준)을 보이며 투자자들에게 AI 모델 비즈니스의 잠재력을 평가 받고 있지만, 동시에 대규모 연산 비용과 가파른 자금 소진 속도라는 리스크를 잠재하고 있다. 특히, 두 기업 모두 매출 대비 과도한 R&D 지출 구조를 보이고 있으며, 이는 중국 AI 밸류체인 전반의 병목 특성을 파악해볼 필요가 있다. 현시점 MaaS 비즈니스와 중국 AI 산업의 투자 가치를 분석하기 위해 우리가 주목해야 할 3가지는 다음과 같다.

포인트 1: AI 칩 및 연산 인프라 내재화

포인트 2: 대규모 AI 추론 수요에 따른 클라우드 인프라 보유 역량

포인트 3: 파운데이션 모델과 수익 플랫폼 연계가능성

Zhipu AI와 Minimax의 사업 부문 및 주력 제품을 통해 위 3가지 관전 포인트를 파악하고자 한다.

## 국산 칩을 활용한 SOTA 달성

2026년 1월 13일에 공개된 Zhipu AI의 이미지 생성 모델 GLM-Image는 텍스트 렌더링 정확도 지표에서 글로벌 경쟁 모델들의 성능을 상회하는 성과를 기록한다. 텍스트 요소가 포함된 이미지 생성 역량을 평가하는 CVTG-2K 벤치마크 기준 GLM-Image는 영어 91.16%, 중국어 95.57%의 정확도를 기록하며 GPT Image 1(영어 85.69%, 중국어 94.83%) 대비 우위를 보였다. 또한 긴 텍스트 이미지 생성 정확도를 측정하는 LongText-Bench에서도 GLM-Image는 영어 95.57%, 중국어 97.88%의 정확도를 나타내며, 특히 중국어 장문 텍스트 생성에서 GPT Image 1(중국어 61.90%)과의 격차가 두드러졌다.

가장 주목할 점은, Zhipu AI의 GLM-Image가 화웨이의 1) Ascend Atlas 800T A2 국산 AI 칩과 2) MindSpore라는 자체 AI 프레임워크를 기반으로 학습부터 추론까지 전 과정을 수행한 사실이다. 이는 중국의 연산 인프라 만으로도 SOTA 모델(State-of-the-art, 공개시점 기준 가장 높은 성능을 기록한 최고 수준 모델)을 학습할 수 있다는 가능성을 검증한 사례이다. 해당 모델은 오픈소스 공개 후 24시간이 채 지나지 않아 Hugging Face 트렌딩 랭킹 1위에 올랐다. 중국 AI 산업의 최대 리스크로 여겨지던 AI 칩 확보에 따른 컴퓨팅 연산 확보가 해소될 여지가 있는지에 대한 판단이 필요한 시점이다. 향후 중국 AI 기업에 대한 밸류에이션은 중국 정부의 정책 프리미엄뿐 아니라 컴퓨팅 파워 공급 역량도 포함될 것으로 판단된다.

[표 1] GLM-Image vs 글로벌 이미지 모델 성능 비교

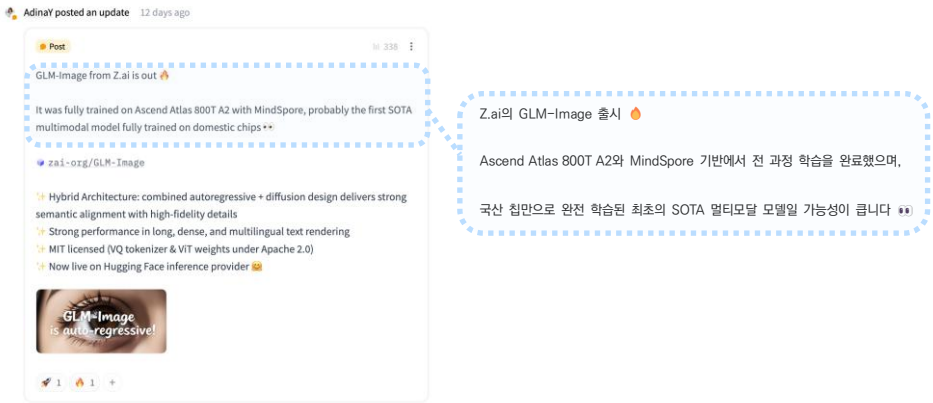
| 모델명 | 오픈소스 여부 | CVTG-2K | | | LongText-Bench | | |
|---|---|---|---|---|---|---|---|
| | | Word | NED | CLIP Score | Avarage | English | China |
| GLM-Image | O | 0.9116 | 0.9557 | 0.7877 | 0.966 | 0.952 | 0.979 |
| Seedream 4.5 | X | 0.899 | 0.9483 | 0.8069 | 0.988 | 0.989 | 0.987 |
| Seedream 4.0 | X | 0.8451 | 0.9224 | 0.7975 | 0.924 | 0.921 | 0.926 |
| Nano Banana 2.0 | X | 0.7788 | 0.8754 | 0.7372 | 0.965 | 0.981 | 0.949 |
| GPT Image 1 [High] | X | 0.8569 | 0.9478 | 0.7982 | 0.788 | 0.956 | 0.619 |
| Qwen Image | O | 0.8288 | 0.9116 | 0.8017 | 0.945 | 0.943 | 0.946 |
| Qwen Image - 2512 | O | 0.8604 | 0.929 | 0.7819 | 0.961 | 0.956 | 0.965 |
| Z - Image | O | 0.8671 | 0.9367 | 0.7969 | 0.936 | 0.935 | 0.936 |
| Z - Image Turbo | O | 0.8585 | 0.9281 | 0.8048 | 0.922 | 0.917 | 0.926 |

자료: HuggingFace, 유안타증권 리서치센터

[그림 5] GLM-Image 가 생성한 텍스트가 포함된 이미지 예시


자료: HuggingFace, 유안타증권 리서치센터

[그림 6] 화웨이 Ascend AI 칩 기반으로 학습~추론 전과정을 수행한 GLM-Image 모델, 국산화 연산 인프라 기반 최초 SOTA 달성




자료: HuggingFace, 유안타증권 리서치센터

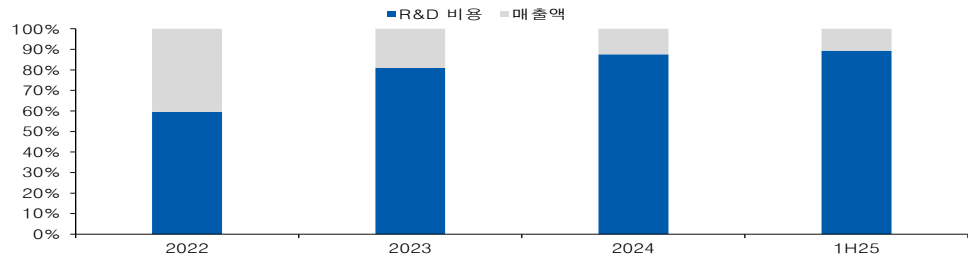
## AI 모델 기업의 R&D 비용은 GPU 사용료

Zhipu AI와 Minimax의 IPO 자금은 현재 사용 속도 기준으로 12~18개월 후 소진되는 규모이다. Zhipu AI의 2024년 연간 기준 R&D 비용은 22억 위안을 기록했으며, 2025년 상반기는 15억 9,500만 위안으로 매출의 8배 이상을 투입했다. MiniMax 역시 매출 대비 3배 이상의 R&D 비용을 지출하였다.

R&D 세부 내역을 확인하면 병목의 본질이 더욱 명확해진다. Zhipu AI의 R&D 비용 중 약 72%는 GPU 클러스터 사용료로 2025년 상반기에만 11억 4,500만 위안을 지출했다. 이는 전체 매출의 6배를 상회하는 수준이다. 현재 중국 AI 모델 기업 연구개발비는 사실상 GPU 임대료에 가깝다. MiniMax 역시 학습용 클라우드 비용이 R&D의 대부분을 차지하는 것으로 나타난다.

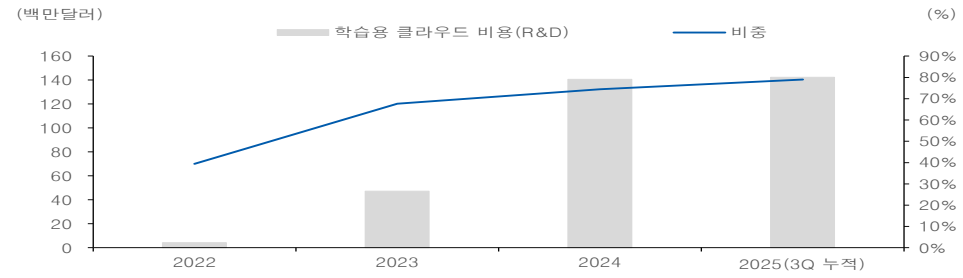
현재 AI 밸류체인 내 수혜자는 명확하다. AI 모델 연산 수요 확대의 직접적인 수혜는 연산 인프라 공급자에게 귀속된다. 그리고 파운데이션 모델 경쟁이 심화될수록 컴퓨팅 파워 의존도는 더욱 높아질 가능성이 크다. 따라서, 소프트웨어 섹터 내 AI 연산 목적의 클라우드 공급 기업에 대한 투자 접근이 유리할 것으로 판단된다.

**[그림 7]** Zhipu AI 의 R&D 비용은 매출액 대비 8 배 이상 (1H25 기준)



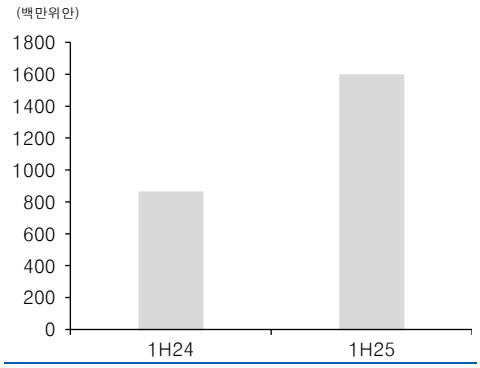

자료: Zhipu AI, 유안타증권 리서치센터

**[그림 8]** MiniMax R&D 전체 비중의 약 80%가 학습용 클라우드 비용 (3Q25 기준)



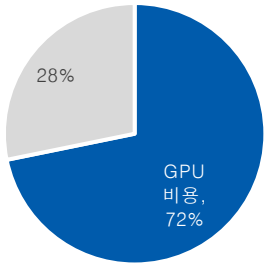

자료: Minimax, 유안타증권 리서치센터

[그림9] Zhipu AI R&D 비용 상승률 YoY +86%




자료: Zhipu AI, 유안타증권 리서치센터

[그림10] Zhipu AI 전체 R&D 비용 중 72%가 GPU 사용 비용




자료: Zhipu AI, 유안타증권 리서치센터

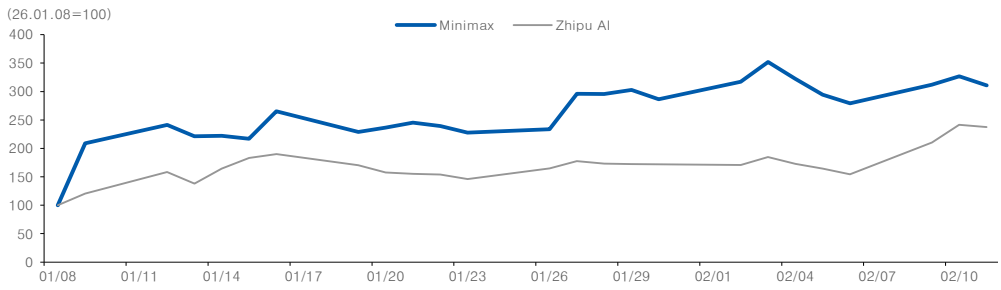
## AI 모델과 플랫폼 연계가능성

2월 10일 종가 기준, MiniMax의 주가는 Zhipu AI 대비 약 55.5%p 초과 성과를 기록하며 약 1.3배 높은 투자수익률을 실현하였다. 이는 글로벌 투자자들의 관심이 단기적으로 중국 기업 및 정부 관련 비즈니스를 주력으로 하는 Zhipu AI의 엔터프라이즈 방식보다는 B2C 플랫폼 확장성에 주목하는 증거로 판단된다.

MiniMax는 2025년 기준 전 세계 200개 이상 국가·지역에서 누적 사용자 수 2억 1,200만 명을 확보했으며, 매출 구조 역시 중국 내수 중심에서 벗어나 글로벌 B2C 플랫폼으로 빠르게 전환하고 있다. 전체 매출에서 중국을 제외한 글로벌 매출 비중이 약 73%를 상회한다. 이는 MiniMax가 단순히 중국 내수 시장을 겨냥한 대표 AI 기업으로 포지셔닝하고 있지 않음을 의미한다.

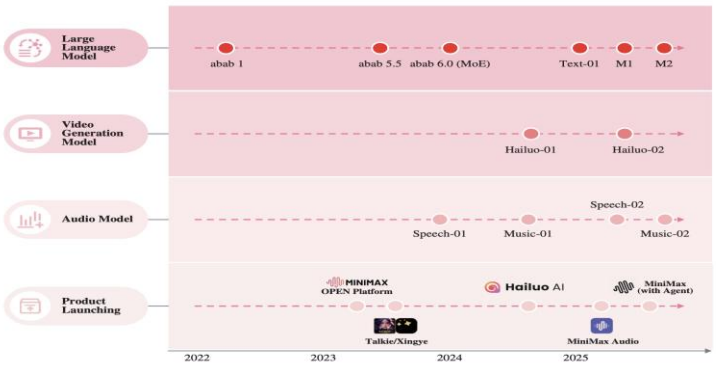
MiniMax는 AI-Native 제품군 중심 B2C 플랫폼 전략을 추구하고 있다. 동사는 파운데이션 모델을 API 또는 라이선스 형태로 제공하는 데 그치지 않고, MiniMax(에이전트), Hailuo AI(이미지 생성), MiniMax Audio(음성 생성), Talkie/Xingye(엔터테인먼트) 등 소비자향 어플리케이션을 통해 멀티모달 생태계를 내재화하고 있다. 소비자 기반 플랫폼은 '모델 → 어플리케이션 → 사용자 트래픽 축적 → 데이터 누적 → 모델 고도화'로 이어지는 선순환 구조를 형성할 수 있다. 특히 Talkie와 같이 엔터테인먼트 기반 서비스는 높은 체류시간을 통해 토큰 사용량을 구조적으로 확대할 수 있다. MiniMax의 경쟁력은 모델 성능 자체보다도, 실사용 트래픽을 통해 수익화하는 플랫폼 생태계라고 판단한다.

**[그림 11]** MiniMax vs Zhipu AI 상장 이후 상대주가 추이



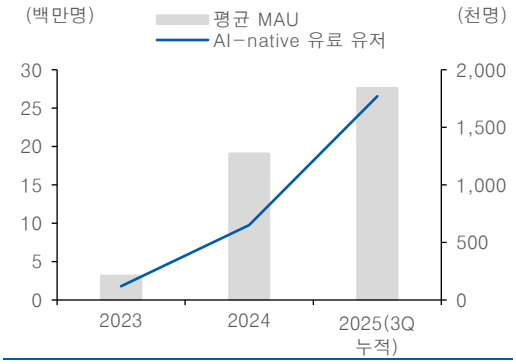

자료: Wind, 유안타증권 리서치센터

**[그림 12]** MiniMax 파운데이션 모델과 AI-Native 제품군 발전 추이



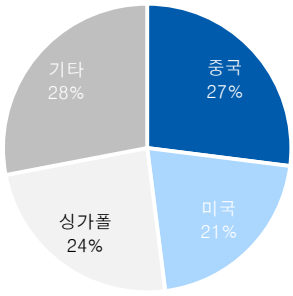

자료: MiniMax , 유안타증권 리서치센터

**[그림13]** Minimax MAU 와 B2C 향 유료 유저 성장성 우상향




자료: MiniMax, 유안타증권 리서치센터
주: 평균 MAU (좌), AI-native 유료 유저 (우)

**[그림14]** Minimax 중국 외 글로벌 지역 매출 비중이 약 72% 이상




자료: MiniMax, 유안타증권 리서치센터

# 중국 AI 투자 포인트

## H200 확보 불확실성 속에서 피어난 중국 AI 전략

엔비디아 고성능 AI 칩 확보 불확실성이라는 지정학적 변수는 1) 중국 AI 칩 설계 자립과 2) 빅테크 기업들의 자체 칩 내재화라는 흐름을 가속화하였다. 2026년 1월 13일, 중국 규제당국은 엔비디아 H200 칩 구매 승인을 오직 연구기관 내에서만 사용가능한 경우로 제한하였다. 중국 정부의 H200 구매 제한 조치는 단기적인 AI 학습 성능 저하를 감수하더라도 화웨이, 캠브리콘 등 AI 칩 자립 생태계를 우선시하는 전략적 판단으로 해석하는 분석이 많았다.

그러나, 중국 산업계 전반에서 엔비디아 AI 칩에 대한 실질적 수요가 여전히 강하다는 점은 추후 보도를 통해 재차 확인되었다. 1월 23일, 중국 당국은 주요 빅테크 기업들에 대해 엔비디아 H200 칩 구매를 원칙적으로 승인한 것으로 확인되었다. 또한, 중국 내 암시장에서는 H200 GPU 8개가 탑재된 서버 한 대가 약 230만 위안 (약 4.62억원)에 거래되는 것으로 보도되었다. 이는 정상가 대비 약 +50% 수준의 프리미엄으로 추정되는 규모이다. 1월 28일에 승인된 엔비디아 H200 AI 칩 수입 물량이 수십만 개 규모로 추정된다는 추가 보도는 중국 내 엔비디아 고성능 AI 칩에 대한 수요가 여전히 견조함을 확인 가능하다. 규제 당국의 공식적인 입장은 없지만, 초기 물량이 알리바바, 텐센트, 바이트댄스 등 중국의 핵심 빅테크 기업에 배정되었다는 것이 이번 승인이 연구 목적을 넘어 실제 산업 내 AI 연산 병목을 해결하기 위한 정책으로 판단된다.

다만 2월 추가 보도에 따르면, 미국 상무부는 H200 수출에 대한 자체 검토 및 분석을 마쳤으나 국무부가 보다 강력한 제한과 승인 조건을 요구하며 절차가 지연되고 있다. 해당 이벤트들의 전개 과정을 보았을 때, 향후 엔비디아의 고성능 AI 칩 수입 불확실성은 지속될 전망이다. 그럼에도 불구하고 AI 연산 병목 완화를 위한 1) AI칩 국산화 가속과 2) AI칩 내재화 전략으로 중국 AI 기업에 대한 투자포인트는 유효하다.

[표 2] 엔비디아 중국향 AI 칩 수출통제 관련 주요 이벤트 전개과정

| | |
|---|---|
| 안보적 차원에서의 AI 칩 수출 제한 최초 시점 | |
| 2022년 8월 | 미 정부가 SEC 공시를 통해 엔비디아 A100 및 H 시리즈 중국, 러시아 수출에 라이선스 요건 부과 통보 |
| 2022년 10월 | 미국 상무부는 10/7 규정을 통해 중국의 첨단컴퓨팅 IC, 슈퍼컴퓨팅 관련 수출 통제 시행 |
| 2025년 엔베디아 AI 칩 수출 제한 전개과정 | |
| 2025년 1월 | 엔비디아 H200, AI Diffusion Rule 적용으로 중국에 대한 전면 수출 제한 |
| 2025년 3월 | 미 상무부, 중국 기업 54개를 수출통제 리스트에 추가 |
| 2025년 4월 | 엔비디아는 미국 정부로부터 H20 칩 수출에 대한 무기한 라이선스 신청이 필요하다는 통보를 받았다고 발표 |
| 2025년 6월 | 엔비디아는 중국 전용 B30 칩 개발. 가격은 6,500~8,000 달러로 H20 대비 저가 |
| 2025년 7월 | 엔비디아는 미국 정부가 H20 칩의 중국 판매를 승인했음을 발표 |
| 2025년 8월 | 엔비디아는 자사 칩에 백도어·킬스위치·모니터링 소프트웨어가 없다고 공식 부인 |
| 2025년 8월 | 트럼프 대통령은 엔비디아, AMD 의 중국 판매 AI 칩에 대해 매출의 15% 라이선스 수수료 부과를 요구 |
| 2025년 8월 | 미국은 AI 칩의 중국 우회 반입을 탐지하기 위해 칩에 추적 장치 삽입을 비밀리에 추진 중이라는 보도 |
| 2025년 8월 | 엔비디아는 Blackwell 기반 중국 전용 B30A 칩을 개발 중이며 성능은 B300의 약 절반 수준 |
| 2025년 9월 | 엔비디아의 반독점법 위반 혐의와 관련해 중국 국가시장감독관리총국이 추가 조사 착수 |
| 2025년 11월 | 백악관은 엔비디아의 최신 성능 AI 칩의 중국 판매를 허용하지 않겠다고 연방기관에 통보했다는 보도 |
| 미국 정부의 AI 칩 수출 허용 시작 | |
| 2025년 11월 | 트럼프 행정부는 엔비디아 H200 AI 칩의 중국 판매 승인 여부를 검토 |
| 2025년 12월 | 트럼프 대통령은 '승인된 중국 고객'에 한해 H200 판매를 허용할 것이라고 발언 |
| 중국 당국의 엔비디아 AI 칩 수입 제한 | |
| 2026년 1월 13일 | 중국 세관 당국은 엔비디아 H200 AI 칩에 대한 통관 금지를 지시 |
| 2026년 1월 15일 | 중국 당국 엔비디아 AI 칩 사용 자국 기업에 구매 총량을 제한하는 규제를 검토하고 있다는 보도 |
| 2026년 1월 17일 | 중국 규제로 인해 엔비디아 공급업체들이 PCB 등 H200 핵심 부품 생산 중단 |
| 중국 당국의 조건부 AI 칩 수입 승인 | |
| 2026년 1월 23일 | 중국 규제당국은 알리바바, 텐센트, 바이트댄스에 대해 H200 구매 준비를 원칙적으로 승인 |
| 2026년 1월 23일 | 중국 암시장에서 H200 GPU 8개가 탑재된 서버 한대 가격이 약 230만 위안으로 거래 (약 4.62억원, 프리미엄 50%) |
| 2026년 1월 28일 | H200 첫 수입 승인 초기 물량은 수십만 개 규모, 3대 주요 빅테크에 우선 배정될 것임을 확인 |
| 여전히 잔존하는 H200 확보 불확실성 | |
| 2026년 1월 29일 | 젠슨 황, 대만 내 인터뷰에서 중국 정부 H200 승인 여부와 관련해 구체적 언급 없이 "긍정적 결정을 기대한다" 언급 |
| 2026년 2월 4일 | 상무부가 아닌 미 국무부 안보 심사에 엔비디아 H200 대중 수출 지연 |

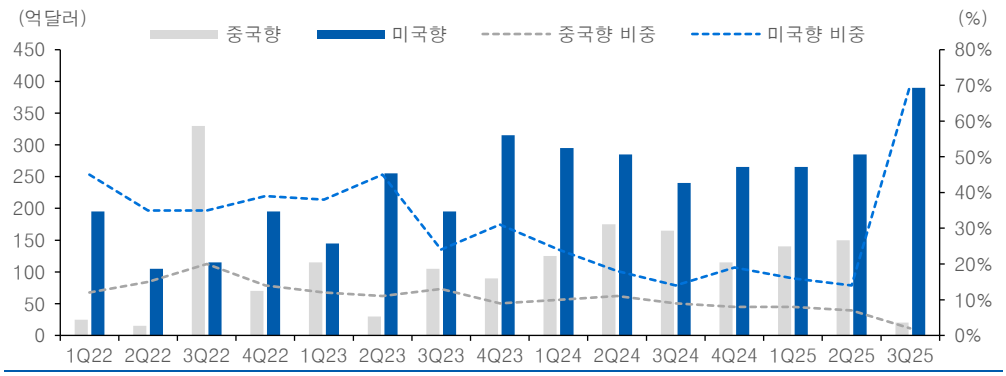
자료: 각 언론사 보도 취합, 유안타증권 리서치센터

### 미국 수출 규제가 만들어 낸 국산 AI 칩 가속화

엔비디아 CEO 젠슨 황이 지난 해 최소 3차례 중국을 방문하고, 올해 1월 상하이를 방문하며 대중국 수출 규제 완화를 위해 적극적인 태도를 보이는 것은 [그림15]를 보면 자명하다. 22년 3분기 엔비디아의 중국 본토 매출 비중은 약 20%(미국 35%)에 달한 것과 달리, 25년 2분기 6%에서, 3분기는 전분기 대비 5%p 추가 하락했다. 이는 장기화된 미국의 수출 규제 강화로 엔비디아의 대중국 매출 공백이 실제로 확대되었으며, 동시에 해당 수요가 중국의 국산 AI 칩으로 전환하는 기회가 될 수 있음을 시사한다.

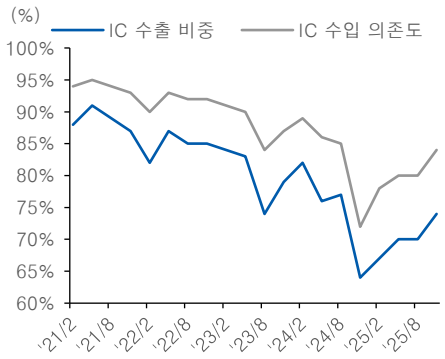
실제로 중국의 반도체 수입 의존도는 감소하고 반도체 중 AI칩 국산화 비중은 증가하였으며, AI 칩 국산 대체 흐름은 지속화될 전망이다. [그림16]을 보면, 2025년 11월 집계 기준 중국의 반도체 수입 의존도는 83%, 총 생산량 대비 수출 비중은 74%이다. 또한 [그림17]에서 확인 가능한 IDC 데이터에 따르면, 2025년 상반기 중국 AI 칩 국산화율은 약 37%로 점진적인 개선 추세가 확인된다. 특히 엔비디아 칩 수출 중단을 감안할 경우, 2025년 하반기에는 국산 AI 칩 점유율은 큰 폭으로 확대된 수치가 집계될 가능성이 크다.

[그림 15] 엔비디아 지역별 매출 분포



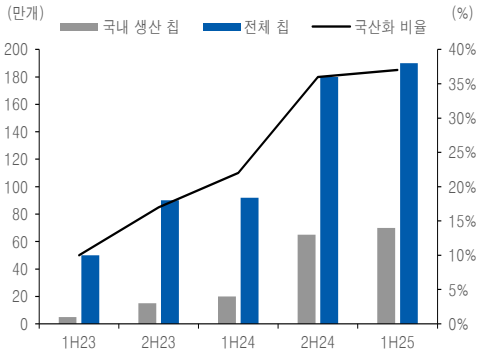

자료: NVIDIA IR, 유안타증권 리서치센터

[그림16] 21~2025년 중국 반도체 수출 비중 및 수입 의존도




자료: 중국 공업정보화부, 해관총서, 유안타증권 리서치센터
주: IC는 집적회로(Integrated Circuit)를 의미함

[그림17] 중국 AI 칩 출하량 및 국산화 비율




자료: IDC, 유안타증권 리서치센터

**중국 빅테크 기업의 AI 칩 내재화 전략: Case Study (1)_알리바바의 AI 하드웨어 밸류체인 완성**

중국 AI칩 국산화 흐름과 함께 주목해야 할 포인트는 빅테크 위주의 AI 칩 공급망 확보이다. 대표적인 사례로 알리바바 반도체 자회사 핑터우거(T-Head, 平頭哥)의 자체 AI칩 양산이 있다. 2026년 1월, 동사는 '전우(Zhenwu, 眞武) 810E' AI 칩 출시로 미국 수출 규제 상황 속에서 AI 하드웨어 자립을 이미 수행하고 있다. 전우 810E는 학습과 추론 연산 모두를 지원하도록 설계된 PPU(Parallel Processing Unit) 병렬 처리 장치로, 96GB HBM2e 고대역폭 메모리 탑재 및 AI 연산 성능 최대 60 TOPS(Tera Operation Per Second, 모델 추론 처리 능력 지표) 스펙을 보유하고 있다.

동사는 해당 자체 칩을 통해 Qwen 모델 및 클라우드 인프라와 최적화된 AI 업스트림 수직계열화를 완료하였다. 이는 엔비디아 A800 클러스터 대비 Qwen-14B 모델 속도가 약 18% 더 빠른 성능을 선보인다. 또한, GPUPU 아키텍처를 통해 기존 엔비디아 기반 범용 GPU와 유사한 병렬 컴퓨팅 방식을 채택하여 호환성이 높다는 장점도 존재한다. 동사는 전우 810E 출시로 구글에 이어 글로벌 두 번째 '파운데이션 모델 + 클라우드 + 자체 AI 칩' 시스템을 보유한 AI 풀스택 기업이다. 해당 칩은 알리바바 클라우드에 이미 10,000개 단위의 칩 클러스터로 구축되어 운용중이며, 이는 단순 테스트 수준을 넘어 실제 Qwen 모델 성능 고도화에 활용되고 있다.

시장 보급 또한 신속하게 진행되고 있다. 2026년 2월 SCMP(South China Morning Post) 보도에 따르면, 전우 810E는 400개 이상의 기업 고객을 확보했다는 사실이 공식 확인되었다. 중국 국가전력망, 중국과학원(CAS)과 같은 공공 및 연구기관뿐만 아니라 자율주행 기술 관련 샤오펑(Xpeng), SNS 플랫폼 시나웨이보(Sina Weibo) 등이 주요 레퍼런스로 언급된다. 또한, 전우 810E의 누적 출하량은 출시 2주만에 이미

10만개를 넘어섰으며, 이는 중국 내 토종 AI 칩 기업 캠브리콘의 판매량을 앞지른 수치이다.

중국 빅테크 자체 AI칩의 리스크 요인은 글로벌 AI 칩과의 성능 격차이다. 전우 810E는 엔비디아 H20에 대항하기에 충분한 스펙을 보유하지만 H200이나 블랙웰 아키텍처 기반 B200과 비교 시 약 1/6 수준의 성능 격차가 여전히 존재한다. 다만, 1) 중국 내 자체 생태계 구축 목표라는 방향과 2) 상대적으로 저렴한 AI 칩 공급 단가를 고려 시, 시장 점유율을 확보할 여력이 충분히 존재하는 것으로 판단된다. 전우 810E는 H20 칩과 유통 가격 경쟁을 이루고 있다는 사실에 근거하여 유닛당 단가는 약 10,000~12,000달러로 추정된다. 엔비디아의 H200과 B200의 단일 유닛 가격이 32,000~45,000달러 이상에 형성되는 것을 고려했을 때, 알리바바의 AI 칩은 1) AI 밸류체인 수직계열화를 통한 연산 병목 해소, 2) 저가 공세에 의한 AI 가속기 시장 침투율이라는 2가지 이점으로 AI 칩 내재화 전략은 중장기적 유효할 것으로 판단된다.

**[표 3]** 전우(Zhenwu, 真武) 810E 주요 스펙

| | 세부 내용 | 비고 |
|---|---|---|
| 메모리 용량 | 96GB HBM2e | 엔비디아 H20 AI 칩 96GB HBM3 적용<br>→ 메모리 세대 1단계 낮음 |
| 메모리 대역폭 | 700GB/s | - |
| 연산 성능 | 60 TOPS | - |
| 아키텍처 | 1) 자체 병렬 처리 장치 (PPU)<br>2) 독자 칩 네트워크 (ICN) | - |
| 전력 소비 | 약 400W | H20 (약 350~400W)과 유사한 데이터센터급 전력 설계 |
| 소프트웨어 스택 | 1) 전면 자체 소프트웨어 스택 구축<br>2) PyTorch·TensorFlow 호환 가능 | - |
| 지원 데이터 | FP16, BF16, INT8 | - |

자료: T-head, 유안타증권 리서치센터

**[표 4]** 글로벌 AI 칩 추정 단가 비교: 하드웨어 병목을 공급망 확보 및 저가공세로 해결

| | Alibaba Zhenwu 810E | Baidu Kunlun M100 | NVIDIA H200 (SXM) | NVIDIA B200 (Blackwell) |
|---|---|---|---|---|
| 유닛당 단가 (추정) | $10,000 ~ $12,000 | 약 $8,000 ~ $10,000 | $32,000 ~ $40,000 | $45,000 ~ $50,000 |
| 참고 레퍼런스 | Caixin Global 추정 | 로이터 보도 및<br>Trend Froce 추정 | TRG DataCenters,<br>Jarvislabs 참고 | 젠슨황 CEO 인터뷰 기반<br>Northflank 분석 |
| 메모리 용량 | 96GB HBM2e | 약 80~96GB (추정) | 141GB HBM3e | 192GB HBM3e |
| 메모리 대역폭 | 700 GB/s | 약 1,000 GB/s 내외 | 4.8 TB/s | 8.0 TB/s |
| AI 연산 성능 (FP8) | 약 60~80 TFLOPS (추정) | 약 100 TFLOPS 내외 | 1,979 TFLOPS | 9,000 TFLOPS |
| 주요 용도 | 추론 및 중소규모 학습 | 대규모 추론 (MoE 특화) | LLM 학습 & 추론 | 초대형 파운드리 모델 학습 |

자료: 유안타증권 리서치센터

### Case Study (2): 바이두의 쿤룬신 5개년 AI 칩 로드맵

바이두의 AI 반도체 자회사 쿤룬신(Kunlunxin, 昆仑芯)은 알리바바 핑터우거와 함께 중국 AI 칩 자립화의 양대산맥으로 꼽힌다. 쿤룬신은 2025년 11월에 개최된 'Baidu World 2025' 컨퍼런스에서 2030년까지의 '5개년 AI 칩 로드맵'을 공식 발표하였다. 동사는 개별 AI 칩의 성능 향상뿐만 아니라, 시스템 차원의 성능을 높이는 목표로 확장세를 보이고 있다.

[표 5] Baidu World 2025 컨퍼런스: 쿤룬신 5 개년 AI 칩 로드맵

| 타임라인 | 마일스톤 | 세부 내용 |
|---|---|---|
| 2026년 | M100 칩 출시 | 고성능 추론 최적화 칩. 특히 MoE(Mixture of Experts) 아키텍처 모델의 추론 효율 극대화 설계 상 |
| 2026년 | Tianchi 512 시스템 | 512개의 칩을 하나로 통합하는 대규모 컴퓨팅 유닛 (슈퍼노드) 출시 |
| 2027년 | M300 칩 출시 | 학습 및 추론 통합형 플래그십 칩. 조 단위 파라미터 멀티모달 모델 학습 지원 목표 |
| 2028년 | Tianchi 슈퍼노드 확장 | 1,000개 단위의 칩을 결합하는 차세대 인프라 플랫폼 구축 |
| 2030년 | 100만 카드 클러스터 | 쿤룬 칩 100만개를 단일 네트워크로 연결하는 초거대 AI 연산 클러스터 완성 |

자료: Baidu, 유안타증권 리서치센터

쿤룬신은 중국 내수 AI 칩 시장에서 엔비디아와 화웨이에 이어 점유율 3위를 기록하고 있다. 바이두는 2030년까지 바이거 5.0 플랫폼을 통해 쿤룬신 칩 100만개 규모의 단일 클러스터 구축을 목표로 하고 있다. 2028년에는 1,000개 단위의 Tianchi 슈퍼노드를 출시할 예정이다. 주요 고객 및 파트너인 국영기업 차이나 모바일은 약 10억 위안 이상 규모의 대규모 공급 계약 체결을 맺었으며, 국가 데이터 센터뿐만 아니라 시나 웨이보 등의 민간 플랫폼으로의 도입 또한 확대되고 있다.

## 모델 아키텍처 고도화를 통한 추론 비용 절감

### MoE 아키텍처를 활용한 추론 비용 감소

중국 AI 기업들은 비용을 하드웨어 인프라에서 해결하는 것이 아닌 모델 아키텍처 연산구조의 비효율을 해결하고자 한다. 미국의 수출 규제로 인해 제한된 AI칩 확보 가능성은 오히려 중국의 AI 모델 아키텍처 기술을 급격하게 발전시키는 환경을 만들어냈다. 대표 사례로 앞서 언급했던 MiniMax는 AI 추론 비용을 전년 대비 45% 절감하는 성과를 달성했다. 관련해서 우리가 주목할 점은 Mixture-of-Experts(MoE) 기반 아키텍처이다.

MoE란 모델 파라미터 전체를 활성화하는 대신 일부 Expert 서브모델만 선택적으로 작동시켜 동일한 파라미터 규모에서도 연산 비용을 낮추고 추론 효율을 높이기 위해 설계된 최적화 아키텍처를 의미한다. 기존 AI 모델들의 기반이 되는 트랜스포머 아키텍처는 모든 입력값에 대해 전체 파라미터를 사용하지만 MoE는 입력에 적합한 소수의 Expert만 활성화되기 때문에 동일한 파라미터 규모에서도 연산량(FLOPs)을 크게 줄여 추론 효율을 높인다. Minimax의 M 2.1 모델 또한 총 2,300억개의 파라미터를 보유하고 있으나, 실제 추론 시에는 100억개의 파라미터만 활성화되는 구조를 가진다.

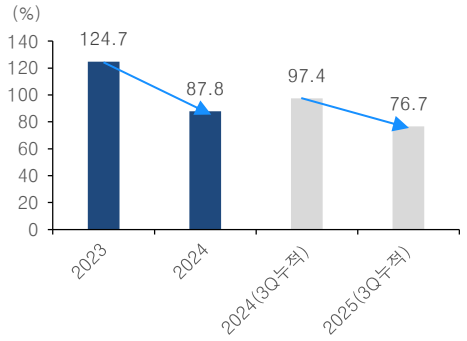
모델 MoE 적용은 실제 손익구조 개선으로 이어진다. 동사는 추론 비용 감소를 통해 매출원가율(COGS Ratio)을 빠르게 낮추고 있다. 2023년 124.7%에 달했던 매출원가율은 2024년 87.8%로 하락했고 2025년 3분기 누적 기준 76.7%까지 추가 개선되었다. 이에 따라 AI-Native 제품의 매출총이익률 또한 -380.2%에서 -8.1%로, 이후 +4.7%까지 회복되며 턴어라운드 국면에 진입했다. 이는 MoE 기반 추론 아키텍처가 장기적으로 마진 개선을 높이는 레버리지로 기능하고 있음을 의미한다.

**[표 6]** MiniMax 모델 Sparse 형태 MoE 아키텍처가 연산량 및 추론 비용에서 효율적

| 성능 지표 | Dense Model (Transformer) | Sparse Model (Mixture-of-Expert) |
|---|---|---|
| 활성 파라미터 비율 | 100% | 약 10% |
| 토큰당 FLOPs | Base (100%) | 약 30~40% 수준 |
| 동일 성능 대비 FLOPs | Base (1.0×) | 0.25~0.5× (2~4배 절감) |
| 추론 연산량 | 입력 길이에 비례 급증 | Expert 로 선형 증가 억제 |
| 추론 속도 | Base | 1.86배 |
| 컨텍스트 위도우 | 128K ~ 256K | 1M (학습) → 4M (추론) |
| GPU Utilization | ~50% 내외 | ~75% |

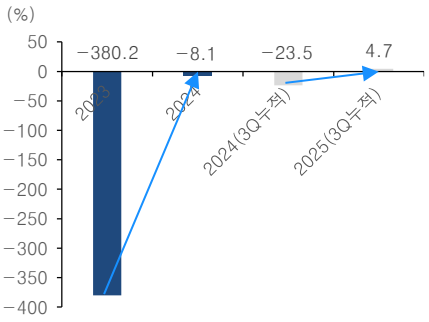
자료: ArXiv, 유안타증권 리서치센터

**[그림18]** 매출원가율(COGS Ratio) 감소 추이 확인 가능



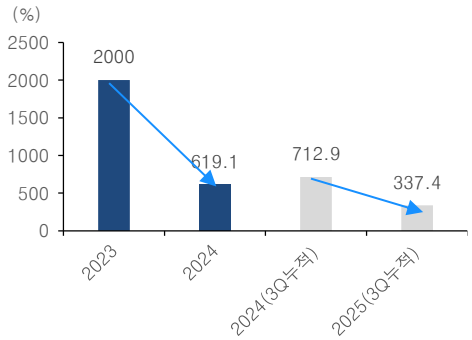

자료: Minimax, 유안타증권 리서치센터

**[그림19]** AI Native 제품의 매출총이익률은 상승 추세



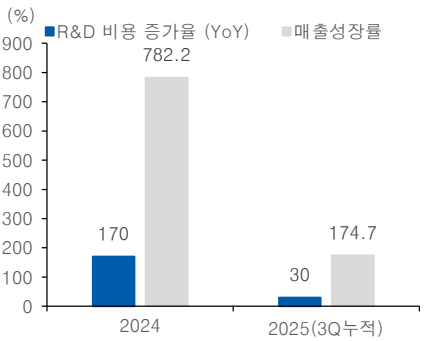

자료: Minimax, 유안타증권 리서치센터

**[그림20]** 매출액 대비 R&D 비용 감소 추이 확인 가능




자료: Minimax, 유안타증권 리서치센터

**[그림21]** R&D 비용 증가율과 매출성장률의 차이 존재




자료: Minimax, 유안타증권 리서치센터

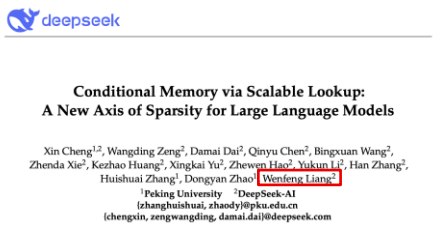
**MoE를 넘어선 Engram 모듈을 활용한 추가적인 비용 절감**

DeepSeek는 다가오는 2월 춘절 전후로 차세대 모델 V4를 공개할 예정이다. DeepSeek의 창시자 량원펑이 공동 저자로 참여한 논문 "Conditional Memory via Scalable Lookup"은 기존 트랜스포머와 MoE 아키텍처가 기억 용량 확장과 계산 효율성 사이에 Trade-off 문제로 불필요한 연산 비용을 소모하고 있음을 지적한다. 이 문제를 해결하기 위해 DeepSeek은 Engram 모듈을 제안한다. 중국 AI 아키텍처의 진화는 향후 AI 연산 비용 병목이 GPU 기반에서 아키텍처 설계 기반으로 전환될 수 있는 잠재 변수이기에 주목할 필요가 있다.

중국의 프론티어급 AI 모델들은 대다수 MoE 아키텍처 기반으로 설계되어 있다. DeepSeek의 V2, V3 모델 역시 MoE 아키텍처를 채택하고 있다. 그러나 MoE는 Expert 선택을 위한 라우팅 비용, 메모리 접근 오버헤드라는 새로운 병목을 내포하고 있다. MoE는 여전히 모델 파라미터가 기억(Memory)과 추론 2가지 연산을 동시에 수행해야 하는 한계점을 지닌다. 2가지 역할을 수행하면 상당한 연산 비효율성이 발생한다. 이로 인해 실제 추론 과정에서 필요하지 않은 파라미터까지 활성화되거나 불필요한 메모리 접근이 반복되며 연산 비용이 증가한다. 비즈니스 측면에서 이는 곧 고성능 GPU, HBM 확보를 위한 대규모 비용으로 직결된다.

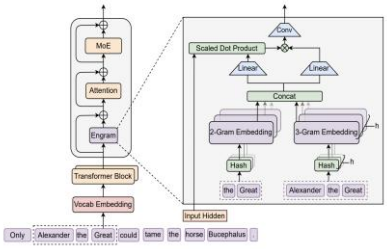
DeepSeek는 Engram 기반 조건부 메모리(Conditional Memory) 아키텍처를 제시한다. 해당 아키텍처는 입력값 길이나 모델 크기와 무관하게 일정한 시간과 연산량으로 필요한 정보를 상수 시간 조회 방식으로 수행하는 'O(1) 메모리 조회'로 대체하여, 모델 성능을 높이면서도 추가 연산 비용 없이 파라미터를 확장할 수 있도록 설계되었다. Engram은 동일 파라미터와 동일 FLOPs 조건에서 기존 MoE 추론 전반에서 의미 있는 성능 향상을 기록한다. 이는 모델 성능이 더 이상 GPU 연산량에 비례하지 않을 수 있음을 의미하며 대형 모델 경제성의 핵심 레버리지가 연산 인프라 공급에서 아키텍처 고도화로 이동할 수 있는 가능성을 제시한다.

**[그림22]** 량원펑(梁文锋) 이 공동저자로 포함된 DeepSeek 발간 논문




자료: ArXiv, DeepSeek, 유안타증권 리서치센터

**[그림23]** DeepSeek Engram 아키텍처 구조도



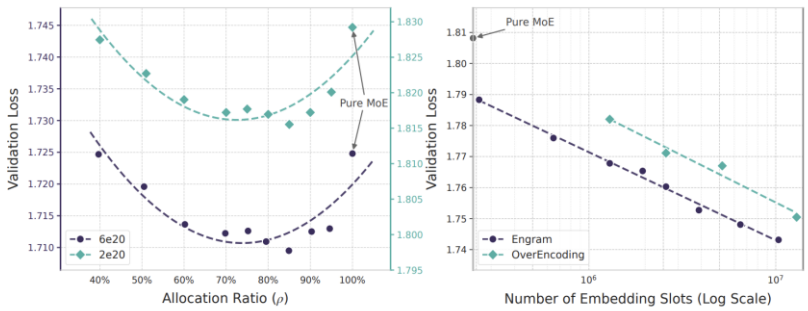

자료: ArXiv, DeepSeek, 유안타증권 리서치센터

**[표 6]** MoE vs Engram 성능 비교 (동일 파라미터 & 동일 FLOPs 기준)

| | 파라미터 수 | FLOPs (연산량) | MMLU (지식 이해) | CMMLU (중국 기준) | BBH (고난도 추론) | ARC Challenge | HumanEval (코딩) | MATH (수학) |
|---|---|---|---|---|---|---|---|---|
| MoE | 26.7B | 3.8B | 57.4 | 57.9 | 50.9 | 70.1 | 37.8 | 28.3 |
| Engram | 26.7B | 3.8B | 60.4 (+3.0) | 61.9 (+4.0) | 55.9 (+5.0) | 73.8 (+3.7) | 40.8 (+3.0) | 30.7 (+2.4) |

자료: ArXiv, DeepSeek, 유안타증권 리서치센터

**[그림 24]** (좌) MOE Engram 최적 조합, (우) Engram 스케일링 법칙




자료: ArXiv, DeepSeek, 유안타증권 리서치센터

[표6]은 MoE와 Engram 아키텍처를 동일 파라미터 수, 동일 FLOPs 조건에서 비교한 결과를 제시한다. 연산 인프라와 모델 크기가 동일함에도 Engram은 MMLU(+3.0p), CMMLU(+4.0p), BBH(+5.0p), ARC(+3.7p), HumanEval(+3.0p), MATH(+2.4p) 등 주요 벤치마크 전반에서 성능 개선을 보인다. 이는 성능 향상이 단순한 연산량 증가나 파라미터 확장 등 모델 스케일링으로부터 기인한 것이 아닌 파라미터 활용 효율이 개선되었음을 의미한다.

[그림24]는 Engram 아키텍처가 메모리 슬롯 수 및 할당 비율 변화에 따라 모델 손실(Loss)과 연산 비용의 관계를 어떻게 재구성하는지 시각적으로 확인 가능하다. 기존 MoE 구조에서는 Expert 수나 메모리 용량을 확장할수록 라우팅 비용과 메모리 접근 오버헤드가 증가하며, 일정 수준 이후에는 효율 개선 효과가 급격히 둔화되는 특성이 나타난다. 반면 Engram은 메모리 슬롯 확장 시에도 손실 감소 효과가 비교적 선형적으로 유지되며, 연산 비용 증가 없이 성능 개선이 지속되는 구간을 확보한다.

투자 관점에서 이는 중국 AI 모델의 비용 함수 자체가 변화할 가능성을 시사한다. 향후 모델 성능 개선의 한계 비용이 GPU 연산량이 아닌 메모리 구조 설계 및 아키텍처 혁신에 의해 결정될 수 있으며, 이는 CAPEX 부담을 완화하는 방향으로 AI 경제성을 재편할 가능성을 보인다.

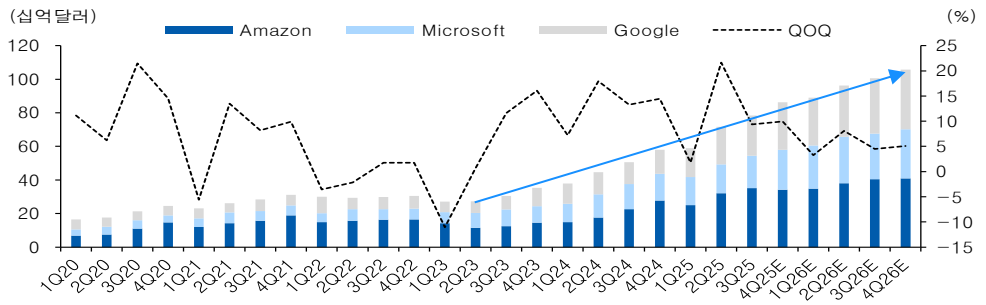
# 중국 AI 투자 전략

## 추론 수요 증가에 따른 AI 밸류체인 수직계열화 기업에 주목

알파벳, 마이크로소프트, 메타, 아마존 등 퍼블릭 클라우드 서비스를 제공하는 미국 4대 하이퍼스케일러가 4분기 실적발표에서 제시한 2026년 CAPEX 규모는 총합 약 6,500억 달러(약 951조원)에 달할 것으로 추정된다. 금번 CAPEX 확대는 전년대비(2025년) 약 60% 이상 증가한 수준이다. 미국 하이퍼스케일러 클라우드 시장은 여전히 AI 수요 증가를 선제적으로 반영하는 CAPEX 확대 국면으로 해석된다.

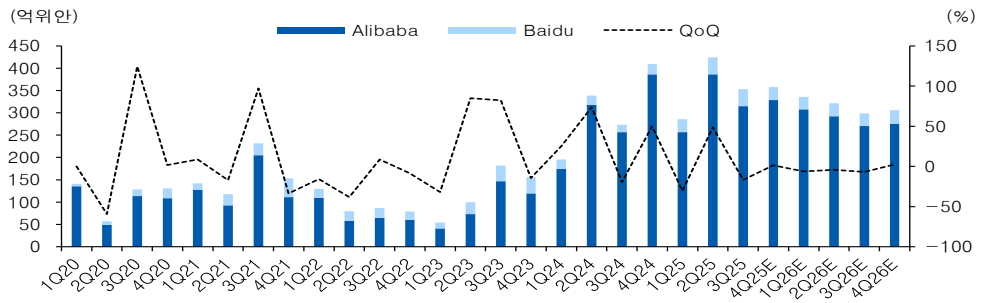
한편 알리바바, 바이두 등 중국의 주요 퍼블릭 클라우드 서비스 제공 기업의 2026년 추정 CAPEX를 확인해보면 성장 둔화 추세가 전망된다. 다만, 향후 추론 연산의 폭발적인 수요가 예상되는 가운데, 중국 메가 플랫폼에서 발생할 대규모 트래픽을 고려 시 클라우드 인프라 공급 기업에 주목할 필요가 있다.

[그림 25] 미국 클라우드 서비스 CAPEX 추이 및 전망 (AWS, Azure, GCP Service)




자료: Bloomberg, 유안타증권 리서치센터

[그림 26] 중국 클라우드 서비스 CAPEX 추이 및 전망 (Alibaba Cloud, Baidu Cloud)



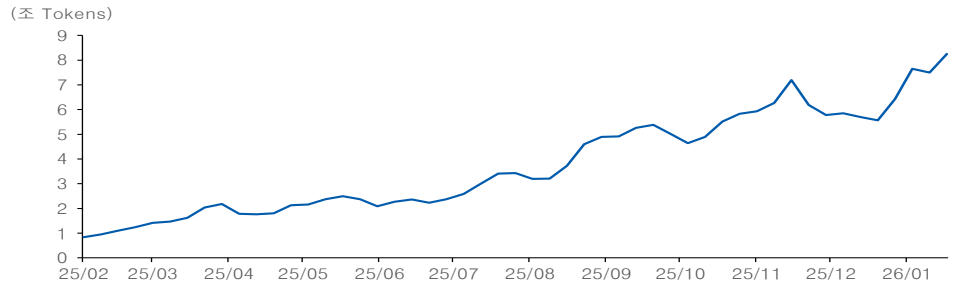

자료: Bloomberg, 유안타증권 리서치센터

## AI 연산 수요 성장축: 학습에서 추론으로

글로벌 AI 연산 수요가 학습에서 추론으로 이동하는 것에 주목할 필요가 있다. 글로벌 토큰 사용량은 2025년 초, 주간 1조 토큰 미만 수준에서 2026년 초, 8조 토큰 이상으로 8배 이상 급증하였다. IEA(국제에너지기구) 분석과 로이터 통신에 따르면, 100만개의 토큰 출력에 약 40~55달러의 연산 비용이 소요된다. 이를 적용하면 글로벌 토큰 사용량이 주당 8조 토큰 수준인 경우, 주간 추론 비용만으로도 약 3.2억~4.4억 달러, 연간 기준 약 170억~230억 달러(약 24.8조원~33.6조원) 수준의 비용이 발생한다.

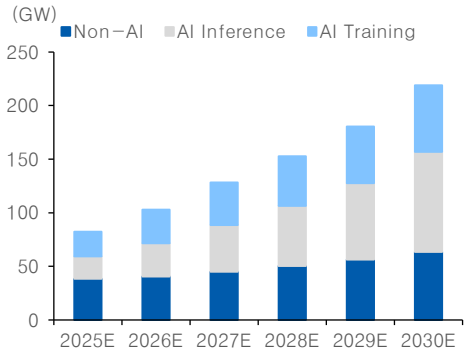
Mckinsey에 따르면, 글로벌 데이터센터 전력 수요 중 AI 추론 비중은 2030년 40% 이상으로 추정되며, 2025~2030년 AI 추론 수요의 연평균성장률은 35%로 AI 학습 수요 비중(약 22%)을 크게 상회한다. 이는 AI 서비스의 실사용 단계 진입에 따라 추론 트래픽이 가파르게 증가하고 있음을 보여준다.

**[그림 27]** 글로벌 대형 언어 모델 토큰사용량 (주간단위)



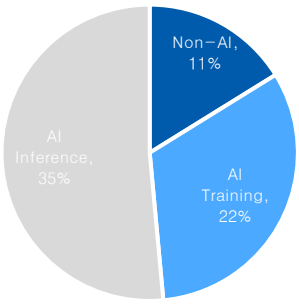

자료: OpenRouter, 유안타증권 리서치센터

**[그림28]** 2025~2030 글로벌 데이터센터 AI 수요: 추론 수요에 주목




자료: McKinsey & Company, 유안타증권 리서치센터

**[그림29]** '25~'30 글로벌 데이터센터 AI 수요별 연평균성장률(CAGR)




자료: McKinsey & Company, 유안타증권 리서치센터

### AI 밸류체인 수직계열화: 빅테크를 넘어 하이퍼스케일러로

'연산 인프라–모델–플랫폼'의 AI 밸류체인 수직계열화 완료 기업에 주목한다. 중국 AI 생태계의 특징은 알리바바, 바이두, 텐센트와 같은 기존 커머스, 검색엔진, 컨텐츠, 메신저 등 플랫폼 기반 비즈니스 업체들이 데이터센터, 클라우드, AI 칩 설계를 바탕으로 AI 하이퍼스케일러로 거듭나고 있다는 점이다. 자체 칩 설계를 통해 외부 하드웨어 의존도를 낮추고 인프라 구축 비용 절감 시, 서비스 단가 경쟁력을 확보할 수 있기 때문에 밸류체인 통합이 곧 AI 플랫폼 내 수익화의 핵심이 될 것으로 판단된다.

AI 밸류체인 수직계열화는 업스트림~미드스트림 연산 인프라(데이터센터, GPU, 클라우드)부터 파운데이션 모델, AI 사용자로 연결되는 다운스트림 플랫폼까지 하나의 기업이 통합 보유하는 것을 의미한다. 앞서 확인한 것처럼, AI 모델이 사용자 활용 단계로 진입하면서 연산 수요는 대규모 사전학습을 넘어 지속 반복적인 추론 사용량이 가파르게 증가하고 있다. 이는 AI 모델 성능보다 플랫폼에서 발생하는 트래픽을 자체 인프라 위에서 얼마나 효율적으로 처리할 수 있는지가 기업 경쟁력을 좌우할 것으로 판단된다. 이에 따라 중국 AI 기업들은 파운데이션 모델을 기존 플랫폼 트래픽을 활용한 수익화 수단으로 활용하는 전략을 추구한다.

**[표 7]** 중국 주요 빅테크 AI 밸류체인 수직계열화 현황 정리

| 기업 | AI 칩 내재화 | 클라우드 인프라 | 파운데이션 모델 | 사용자 플랫폼 | AI 수익구조 |
|---|---|---|---|---|---|
| Baidu | 쿤룬신<br>AI 칩 시리즈 | Baidu Cloud<br>(GPU 클러스터 서비스) | ERNIE<br>(검색/Agent/자율주행) | 검색, 앱<br>AI 어시스턴트 | 검색 광고, AI API<br>SaaS, Robotaxi |
| Alibaba | 핑터우거<br>AI 칩 자회사 보유 | Alibaba Cloud<br>(중국/아태 지역 1위) | Qwen 시리즈 | Taobao, Tmall<br>AliExpress, DingTalk | 커머스 수수료, 광고<br>클라우드, AI SaaS |
| ByteDance | 자체 AI 칩<br>초기 단계 | Volcano Engine<br>Cloud | Doubao/Seed<br>계열 LLM | Douyin, TikTok<br>(MAU ~10억 규모) | 광고, 인앱결제<br>커머스, AI 서비스 |
| Tencent | 자체 AI 칩<br>연구 검토 단계 | Tencent Cloud | Hunyuan<br>(위챗/게임/핀테크 통합) | Wechat, WeCom<br>(자체 게임 IP) | 게임, 광고, 결제,<br>미니프로그램 생태계 |
| Huawei | Ascend NPU 시리즈 | Huawei Cloud | Pangu | Harmony OS<br>(디바이스 운영체제) | B2B 중심<br>AI 칩 판매 |

자료: 유안타증권 리서치센터 / 주: 파란색 – 침투율 높음, 빨간색 – 침투율 낮음

**수직계열화 전략의 목표: 추론 비용의 자산화**

AI 밸류체인 수직계열화의 핵심은 추론 수요를 비용이 아니라 자산으로 전환하는 것에 있다. AI 플랫폼 사용자가 늘어날수록 토큰 사용량은 선행적으로 증가하지만, 증가분이 손익계산서에 반영되는 방식은 기업 형태에 따라 달라진다. 외부 클라우드에 의존하는 AI-Native 기업의 경우 추론 증가가 곧바로 GPU 사용 과금으로 이어지는 OPEX로 분류되며 플랫폼 트래픽 증가는 손익 악화로 이어진다. 반면, AI 전 밸류체인을 내재화한 기업은 추론 연산 증가가 내부 데이터센터 가동률 상승과 GPU 활용도 개선으로 이어지며 클라우드 매출로 인식된다.

대표적인 사례로, 알리바바 클라우드는 GPU를 모델 단위가 아닌 토큰 단위로 분산하여 추론 작업을 동시에 스케줄링하고 하나의 GPU에 다수 모델을 패킹해 자원을 최소화하는 Aegaeon 풀링 시스템을 도입해 GPU 활용을 극대화하였다. 그 결과, 자체 Qwen 모델에 필요한 GPU 수를 1,192개에서 213개로 82% 절감하였다. 이는 추론 수요가 증가할수록 연산 부담이 비용으로 누적되는 것이 아니라, CAPEX 기반 고정비를 분산시키며 단위당 비용을 낮추는 레버리지로 기능할 수 있음을 보여주는 사례다.

데이터센터 및 클라우드 서비스를 구축한 중국 메가 플랫폼 기업들은 모델을 외부 API로만 판매하는 것을 넘어서 자사 커머스, 컨텐츠, 메신저 트래픽 안에서 추론 연산이 늘어날수록 플랫폼 참여도와 클라우드 수익성이 동시에 개선되는 구조를 구축하고 있다. 수직계열화의 본질은 AI 추론을 비용에서 자산으로 전환하며, 추론 수요 폭증 국면에서 AI 풀스택 기업으로의 전환을 완성한 기업이 중장기적 수익성을 확보할 수 있다.

AI 밸류체인 수직계열화를 완료한 기업으로 알리바바(9988.HK)와 바이두(9888.HK)를 Top picks로 제시한다.


# 알리바바 (9988.HK)

# 알리바바 (9988.HK) N/R (I)

## 4,800억 위안 규모 투자 확대에서 드러나는 중국 AI 1위의 자신감

**2026년에도 클라우드 성장은 이어진다:** 3Q 클라우드 매출은 398억 위안(+34% YoY)으로 9개 분기 연속 두 자릿수 성장. 동사는 2026년 AI 및 클라우드 투자 가이던스를 4,800억 위안(기존 3,800억 위안 제시) 수준으로 상향 조정. 동사는 AI 특화 클라우드 점유율 1위(비중 약 35.8%) 인프라 보유 기업으로, 정부 주도 하에 빠르게 성장할 중국 AI 추론 연산 증가 국면에서 수혜 전망

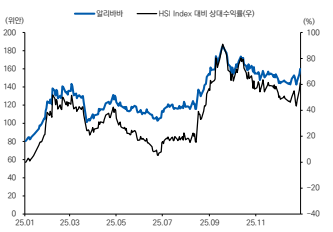
**자체 AI칩 클라우드 배포 유일 기업:** 동사는 자회사 핑터우거(T-Head)를 통해 AI 반도체 내재화함으로써 AI 전 밸류체인 수직계열화 완료. 1월 29일 공개한 AI 칩 전우(眞武) 810E는 엔비디아 H20과 유사한 수준의 성능을 제공하며 알리 클라우드 내부 연산에 활용 중

**Agentic AI를 활용한 본업의 외형성장:** 4Q25E 커머스 매출액은 1,663억 위안(QoQ +25%, YoY +45%)으로 큰 폭 성장 기대. 다만 2026년 AI 플랫폼 경쟁 심화에 따른 동사의 커머스 광고 및 결제 수익 둔화 우려가 유일한 리스크. 1월 15일, Qwen App 업데이트 발표로 400개 이상의 AI 에이전트 기능을 새롭게 출시. 이는 Taobao, Alipay, Fliggy, Amap 등 핵심 플랫폼 연계로 Customer단 트래픽 증가가 기대되며 수익화 전환의 기반이 될 것으로 판단

| | |
|---|---|
| **컨센서스** | **171.90 위안** |
| **현재주가 (02/11)** | **141.59 위안** |
| 시가총액 | 2,704.33 십억위안 |
| 총발행주식수 | 19,099 백만주 |
| 90일 평균 거래대금 | 13,527.34 백만위안 |
| 90일 평균 거래량 | 94,628,150 주 |
| 52주 고 | 170.43 위안 |
| 52주 저 | 90.70 위안 |
| 주요주주 | JP모간 체이스 외 1 인 10.5% |

| 주가수익률 (%) | 1M | 3M | 12M |
|---|---|---|---|
| 절대 | 10.3 | 3.6 | 48.8 |
| HSI Index | 4.0 | 2.5 | 28.0 |

주가 및 상대수익률




Forecasts and Valuation (GAAP 연결) (단위: 억위안, 위안, %, 배)

| 결산 (12월) | 2022A | 2023A | 2024A | 2025E | 2026E |
|---|---|---|---|---|---|
| 매출액 | 8,531 | 8,687 | 9,412 | 9,999 | 10,347 |
| 영업이익 | 696 | 1,004 | 1,134 | 1,368 | 864 |
| 당기순이익 | 622 | 728 | 800 | 1,563 | 1,000 |
| PER | 15.7 | 17.2 | 10.2 | 17.6 | 27.4 |
| PBR | 2.0 | 1.8 | 1.3 | 2.5 | 2.8 |
| EV/EBITDA | 8.6 | 7.9 | 4.0 | 11.8 | 11.8 |
| ROE | 6.6 | 7.5 | 8.1 | 13.1 | 9.7 |

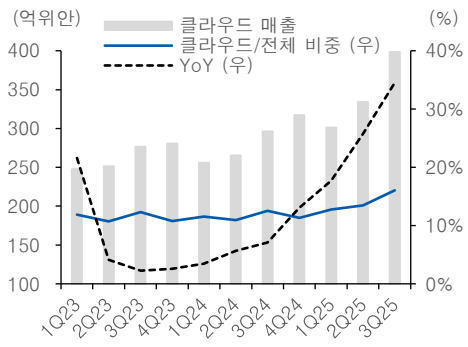
자료: 유안타증권 리서치센터

[표 1] 알리바바 AI Full-Stack Architecture

| 레이어 | 핵심 제품 | 주요 역할 | 전략적 의미 |
|---|---|---|---|
| AI Chip | 핑터우거(T-Head) 전우(鎮岳) 810E | AI 가속기 내재화<br>(엔비디아 H20 대체 가능) | AI 연산 병목 리스크 해소<br>(지정학적 요소, 지연 비용 등) |
| Cloud Platform | Alibaba Cloud | AI 학습 & 추론 인프라<br>(중국 내 최대 클라우드) | AI 연산 수요 증가의<br>직접적 수혜 |
| AI Foundation Model | Qwen (通义千问) | LLM, 멀티모달 API, 앱,<br>AI 에이전트 생태계 확장 | 플랫폼 확장성 및 락인 효과 |
| Application | Qwen App, Taobao, Alipay,<br>Amap, Fliggy | 검색/추천/광고/에이전트<br>기반 수익 전환 | 광고, 커머스의 AI 에이전트 기반<br>실제 거래 전환 |

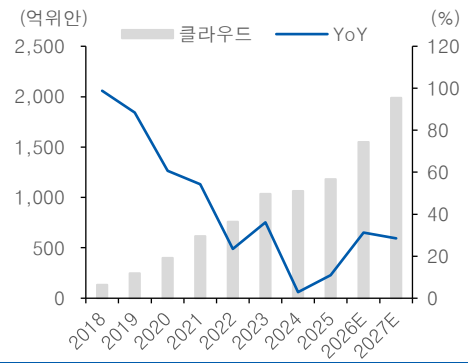
자료: 유안타증권 리서치센터

[그림1] 알리바바 클라우드 인텔리전스 분기 매출 및 비중 추이




자료: Bloomberg, 유안타증권 리서치센터

[그림2] 알리바바 클라우드 인텔리전스 연간 매출 및 성장률




자료: Bloomberg, 유안타증권 리서치센터

[그림 3] 1 월 15 일 소비자 대상 AI 어플리케이션 알리바바 Qwen App 업데이트 진행


자료: Alibaba Qwen, 유안타증권 리서치센터

[표 2] Valuation Table (클라우드 인텔리전스 서비스 Peer 기준) (단위: 십억위안, %, 배)

| | | Alibaba | Baidu | Tencent | Amazon | Microsoft | Alphabet | Oracle |
|---|---|---|---|---|---|---|---|---|
| 시가총액 | | 2,704.3 | 357.9 | 4,411.9 | 15,353.3 | 21,207.2 | 26,634.5 | 3,174.6 |
| 매출액 | 2024 | 941.2 | 134.6 | 609.0 | 4,593.4 | 1,770.7 | 2,520.2 | 382.2 |
| | 2025E | 999.9 | 132.5 | 657.6 | 4,959.1 | 2,002.5 | 2,370.0 | 410.4 |
| | 2026E | 1,034.7 | 129.1 | 751.2 | 5,555.8 | 2,271.7 | 2,794.5 | 462.9 |
| | 2027E | 1,146.0 | 134.7 | 831.2 | 6,199.9 | 2,622.5 | 3,223.9 | 601.5 |
| 영업이익 | 2024 | 113.4 | 21.9 | 208.1 | 574.6 | 927.7 | 927.2 | 127.7 |
| | 2025E | 136.8 | 24.7 | 211.2 | 555.7 | 902.7 | 910.8 | 178.6 |
| | 2026E | 86.4 | 13.9 | 246.3 | 687.2 | 1,052.5 | 1,093.6 | 194.8 |
| | 2027E | 134.3 | 15.6 | 285.7 | 848.0 | 1,211.7 | 1,361.6 | 239.0 |
| OPM (%) | 2024 | 12.0 | 16.2 | 34.2 | 12.5 | 52.4 | 36.8 | 33.4 |
| | 2025E | 13.7 | 18.6 | 32.1 | 11.2 | 45.1 | 38.4 | 43.5 |
| | 2026E | 8.3 | 10.7 | 32.8 | 12.4 | 46.3 | 39.1 | 42.1 |
| | 2027E | 11.7 | 11.6 | 34.4 | 13.7 | 46.2 | 42.2 | 39.7 |
| 당기순이익 | 2024 | 130.1 | 23.8 | 194.1 | 558.1 | 735.0 | 949.7 | 89.9 |
| | 2025E | 156.3 | 24.9 | 218.1 | 659.9 | 718.9 | 943.8 | 123.3 |
| | 2026E | 100.0 | 18.2 | 260.4 | 719.5 | 867.8 | 980.2 | 148.7 |
| | 2027E | 142.9 | 19.6 | 293.0 | 852.6 | 979.8 | 1,174.6 | 160.3 |
| PER (X) | 2024 | 10.2 | 10.1 | 18.7 | 39.0 | 37.8 | 23.9 | 30.2 |
| | 2025E | 17.6 | 14.3 | 21.6 | 23.6 | 30.8 | 29.2 | 26.7 |
| | 2026E | 27.4 | 17.2 | 17.4 | 22.6 | 24.7 | 26.8 | 21.7 |
| | 2027E | 18.7 | 15.2 | 15.4 | 18.9 | 21.7 | 22.6 | 20.0 |

자료: Bloomberg, 유안타증권 리서치센터 / 주: 1) 중국, 미국 단위 모두 십억위안으로 통일, 2) 2026/02/11 종가 기준

## 알리바바 (9988.HK) 재무제표 (GAAP 연결)

| 손익계산서 | | | | | (단위: 십억위안) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| **매출액** | **510** | **717** | **853** | **869** | **941** |
| 매출원가 | 282 | 421 | 539 | 550 | 586 |
| 매출총이익 | 227 | 296 | 314 | 319 | 355 |
| 판관비 | 79 | 137 | 152 | 146 | 157 |
| **영업이익** | **91** | **90** | **70** | **100** | **113** |
| EBITDA | 139 | 144 | 129 | 158 | 169 |
| 영업외손익 | -75 | -76 | 10 | 11 | 12 |
| 외환관련손익 | 0 | 0 | 0 | 0 | 0 |
| 이자손익 | - | - | - | - | - |
| 관계기업관련손익 | - | - | - | - | - |
| 기타 | -75 | -76 | 10 | 11 | 12 |
| 법인세비용차감전순손익 | 167 | 166 | 60 | 89 | 102 |
| 법인세비용 | 21 | 29 | 27 | 16 | 23 |
| 계속사업순손익 | 140 | 143 | 47 | 66 | 71 |
| 중단사업순손익 | 0 | 0 | 0 | 0 | 0 |
| 당기순이익 | 140 | 143 | 47 | 66 | 71 |
| **지배지분순이익** | **149** | **151** | **62** | **73** | **80** |
| 포괄순이익 | 151 | 132 | 48 | 96 | 94 |
| 지배지분포괄이익 | 151 | 131 | 46 | 97 | 94 |

주: 이익 산출 기준은 기존 k-GAAP과 동일. 즉, 매출액에서 매출원가와 판관비만 차감

| 현금흐름표 | | | | | (단위: 십억위안) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| **영업활동 현금흐름** | **181** | **232** | **143** | **200** | **183** |
| 당기순이익 | 149 | 151 | 62 | 73 | 80 |
| 감가상각비 | 42 | 48 | 48 | 47 | 45 |
| 외환손익 | 0 | 0 | 0 | 0 | 0 |
| 종속, 관계기업관련손익 | 6 | -8 | -14 | 8 | 8 |
| 자산부채의 증감 | - | - | - | - | - |
| 기타현금흐름 | -17 | 41 | 47 | 72 | 50 |
| **투자활동 현금흐름** | **-108** | **-244** | **-199** | **-136** | **-22** |
| 투자자산 | 194 | 399 | 489 | 577 | 544 |
| 유형자산 증가 (CAPEX) | - | -41 | -53 | -34 | -32 |
| 유형자산 감소 | 0 | 0 | 0 | 1 | 0 |
| 기타현금흐름 | -302 | -602 | -634 | -679 | -534 |
| **재무활동 현금흐름** | **21** | **29** | **27** | **16** | **23** |
| 단기차입금 | 5 | 4 | 9 | 7 | 13 |
| 사채 및 장기차입금 | 120 | 136 | 133 | 149 | 142 |
| 자본 | 871 | 1,075 | 1,073 | 1,113 | 1,102 |
| 현금배당 | 0 | 0 | 0 | 0 | -18 |
| 기타현금흐름 | -975 | -1,185 | -1,187 | -1,254 | -1,216 |
| 연결범위변동 등 기타 | 54 | -6 | -100 | -78 | -126 |
| **현금의 증감** | **147** | **10** | **-129** | **2** | **57** |
| 기초 현금 | 198 | 346 | 356 | 227 | 230 |
| 기말 현금 | 346 | 356 | 227 | 230 | 286 |
| NOPLAT | **74** | **81** | **53** | **75** | **80** |
| FCF | **135** | **190** | **89** | **165** | **151** |

자료: 유안타증권

주: 1. EPS, BPS 및 PER, PBR은 지배주주 기준임
2. PER등 valuation 지표의 경우, 확정치는 연평균 주가 기준, 전망치는 현재주가 기준임
3. ROE, OA의경우, 자본, 자산 항목은 연초, 연말 평균을 기준일로 함

| 재무상태표 | | | | | (단위: 십억위안) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| **유동자산** | **463** | **643** | **639** | **698** | **753** |
| 현금및현금성자산 | 331 | 321 | 190 | 193 | 248 |
| 매출채권 및 기타채권 | 20 | 27 | 33 | 32 | 31 |
| 재고자산 | 0 | 30 | 28 | 29 | 25 |
| 비유동자산 | 850 | 1,047 | 1,057 | 1,055 | 1,012 |
| 유형자산 | 138 | 219 | 250 | 253 | 262 |
| 관계기업등 지분관련자산 | 190 | 200 | 220 | 207 | 203 |
| 기타투자자산 | 16 | 15 | 21 | 18 | 28 |
| **자산총계** | **1,313** | **1,690** | **1,696** | **1,753** | **1,765** |
| 유동부채 | 242 | 377 | 384 | 385 | 422 |
| 매입채무 및 기타채무 | 178 | 280 | 286 | 282 | 300 |
| 단기차입금 | 5 | 4 | 9 | 7 | 13 |
| 유동성장기부채 | 0 | 10 | 0 | 5 | 16 |
| 비유동부채 | 191 | 229 | 230 | 245 | 231 |
| 장기차입금 | 120 | 136 | 133 | 149 | 142 |
| 사채 | 0 | 0 | 0 | 0 | 0 |
| **부채총계** | **433** | **607** | **613** | **630** | **652** |
| 지배지분 | 765 | 946 | 958 | 1,000 | 997 |
| 자본금 | 0 | 0 | 0 | 0 | 0 |
| 자본잉여금 | 344 | 394 | 411 | 417 | 398 |
| 이익잉여금 | 406 | 555 | 564 | 599 | 598 |
| 비지배지분 | 115 | 137 | 124 | 123 | 115 |
| **자본총계** | **880** | **1,084** | **1,082** | **1,123** | **1,113** |
| **순차입금** | **-377** | **-539** | **-502** | **-575** | **-586** |
| 총차입금 | 147 | 181 | 177 | 196 | 206 |

| Valuation 지표 | | | | | (단위: 위안, 배, %) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| EPS | 56.8 | 55.6 | 23.0 | 27.7 | 31.6 |
| BPS | 281.2 | 344.9 | 349.9 | 382.8 | 408.0 |
| EBITDAPS | 68.8 | 85.8 | 52.7 | 76.2 | 72.4 |
| SPS | 194.0 | 265.5 | 314.7 | 331.3 | 373.1 |
| DPS | 0.0 | 0.0 | 0.0 | 7.1 | 7.1 |
| PER | **30.8** | **54.0** | **15.7** | **17.2** | **10.2** |
| PBR | **4.9** | **4.3** | **2.0** | **1.8** | **1.3** |
| EV/EBITDA | **24.7** | **25.2** | **11.7** | **8.7** | **4.8** |
| PSR | **7.1** | **5.6** | **2.2** | **2.1** | **1.4** |

| 재무비율 | | | | | (단위: 배, %) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| **매출액 증가율 (%)** | **35.3%** | **40.7%** | **18.9%** | **1.8%** | **8.3%** |
| **영업이익 증가율 (%)** | **60.2%** | **-1.9%** | **-22.3%** | **44.1%** | **13.0%** |
| **지배순이익 증가율 (%)** | **70.0%** | **0.8%** | **-58.7%** | **16.9%** | **9.9%** |
| 매출총이익률 (%) | 44.6% | 41.3% | 36.8% | 36.7% | 37.7% |
| **영업이익률 (%)** | **17.9%** | **12.5%** | **8.2%** | **11.6%** | **12.0%** |
| 지배순이익률 (%) | 27.5% | 20.0% | 5.5% | 7.5% | 7.6% |
| EBITDA 마진 (%) | 27.4% | 20.1% | 15.1% | 18.2% | 17.9% |
| ROIC | 8.0 | 6.7 | 3.9 | 5.5 | 5.9 |
| ROA | 13.1 | 10.0 | 3.7 | 4.2 | 4.5 |
| ROE | **23.9** | **17.8** | **6.6** | **7.5** | **8.1** |
| 부채비율 (%) | 1674% | 1674% | 1632% | 1742% | 1848% |
| 순차입금/자기자본 (%) | -4289% | -4976% | -4640% | -5117% | -5270% |
| 영업이익/금융비용 (배) | 17.7 | 20.0 | 14.2 | 17.0 | 14.3 |


# 바이두 (9888.HK)

# 바이두 (9888.HK) N/R (I)

## 안정적인 AI 플랫폼 캐시카우, 고성장하는 AI 자율주행

**중국판 CUDA 플랫폼으로 성장:** 자체 AI 칩 쿤룬은 중국 AI 가속기 3세대까지 상용화에 성공. 2025년 쿤룬 AI칩 매출은 35억 위안 이상으로 추정. 동사가 자체 개발한 딥러닝 프레임워크 페이장은 3Q25 기준 개발자 2,333만명, 엔터프라이즈 고객 76만곳 확보, 플랫폼 내 100만개 이상의 모델 생성. 동사는 2026년 AI 클라우드 및 Ernie 모델 관련 매출 성장 목표를 기존 100%에서 200%로 상향 조정

**검색엔진 기반 AI-Native 광고 수익의 폭발적 성장:** 전체 모바일 검색의 70%가 AI 생성 컨텐츠가 적용. 또한 동사의 App MAU 90% 이상을 차지하며 AI 기반 사용자 경험 개선에 성공. AI 에이전트 기반 마케팅은 일간 3.3만개의 광고주가 활용중. 3Q25 기준, AI-Native 마케팅 매출은 28억 위안(+262% YoY)으로 핵심 광고 매출의 18%를 차지

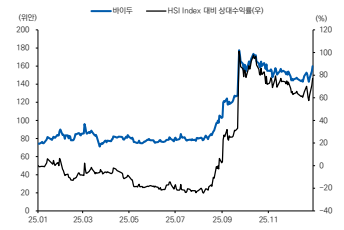
**중국 AI 유일 글로벌 자율주행 플랫폼:** 3Q25 기준, Apollo Go는 주간 25만회 이상의 완전 무인 자율주행 서비스 제공하며 누적 운행 건수 1,700만회 달성. '24년 우한 지역 내 흑자 전환을 시작으로 RT6 차량 가격(약 20만 위안대)을 대폭 낮춰 2025E 연간 기준, AI 자율주행 전체 지역 흑자 전환가능성 높음. 전세계 22개 도시에서 운영중이며, 중동(두바이, 아부다비) 진출을 시작으로 글로벌 추가 확장 시 AI 자율주행 기업으로서 리레이팅 가능성 유효

| | |
|---|---|
| **컨센서스** | **150.45 위안** |
| **현재주가 (02/11)** | **127.62 위안** |

| | |
|---|---|
| 시가총액 | 357.94 십억위안 |
| 총발행주식수 | 2,280 백만주 |
| 90일 평균 거래대금 | 1,756.01 백만위안 |
| 90일 평균 거래량 | 15,172,990 주 |
| 52주 고 | 143.95 위안 |
| 52주 저 | 69.42 위안 |
| 주요주주 | 블랙록 외 1인 11.4% |

| 주가수익률 (%) | 1M | 3M | 12M |
|---|---|---|---|
| 절대 | 2.9 | 11.9 | 65.1 |
| HSI Index | 4.0 | 2.5 | 28.0 |

주가 및 상대수익률




Forecasts and Valuation (GAAP 연결) (단위: 억위안, 위안, %, 배)

| 결산 (12월) | 2022A | 2023A | 2024A | 2025E | 2026E |
|---|---|---|---|---|---|
| 매출액 | 1,237 | 1,346 | 1,331 | 1,291 | 1,347 |
| 영업이익 | 159 | 219 | 213 | 139 | 156 |
| 당기순이익 | 76 | 203 | 238 | 182 | 196 |
| PER | 21.1 | 16.5 | 10.1 | 17.2 | 15.2 |
| PBR | 1.2 | 1.2 | 0.8 | 1.2 | 1.2 |
| EV/EBITDA | 7.1 | 4.4 | 2.9 | 10.0 | 10.0 |
| ROE | 3.5 | 8.4 | 9.1 | 5.3 | 6.0 |

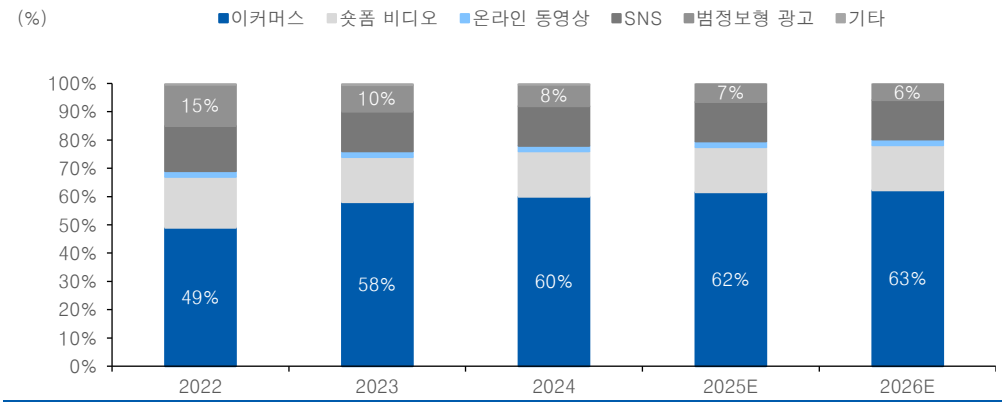
자료: Bloomberg, 유안타증권 리서치센터

**[표 1]** 바이두 AI Full-Stack Architecture

| 레이어 | 핵심 제품 | 주요 역할 | 전략적 의미 |
|---|---|---|---|
| AI Chip | 쿤룬 P800,<br>M100/M300 시리즈 (예정) | AI 가속기 내재화 | AI 연산 병목 리스크 해소<br>(지정학적 요소, 지연 비용 등) |
| Cloud Platform | Baidu AI Cloud<br>바이거(百舸) 5.0 | AI 학습 & 추론 인프라 | AI 연산 수요 증가의<br>직접적 수혜 |
| AI Framework | PaddlePaddle (飞桨) | 딥러닝 기반 엔터프라이즈<br>AI 개발 프레임워크 | 칩-모델 최적화 과정<br>개발자 락인 |
| AI Foundation Model | ERNIE (文心) | LLM, 멀티모달 API, 앱,<br>Agent 생태계 확장 | 플랫폼 확장성 및 락인 효과 |
| Application | 1) 검색엔진 AI 컨텐츠<br>2) AI-Native 마케팅<br>3) Apollo Go AI 자율주행 | AI 검색, 광고,<br>자율주행 상용화 | AI → 사용자 체류 시간 증가 →<br>광고, 모빌리티 수익화 |

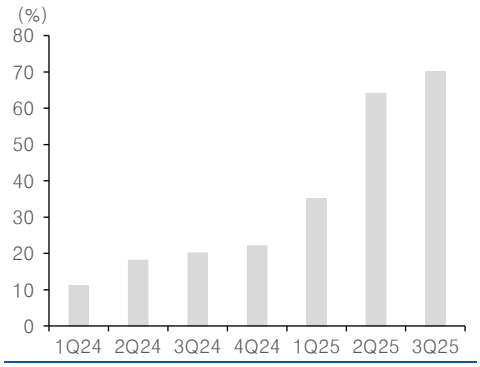
자료: 유안타증권 리서치센터

**[그림 1]** 중국 광고 유형별 시장 점유율 변화



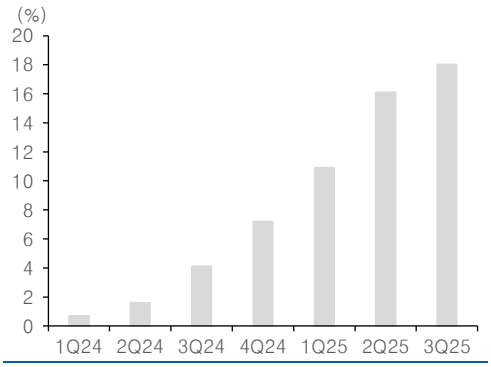

자료: QuestMobile, 유안타증권 리서치센터 / 주: 범정보형 광고는 검색 엔진 내 배너 광고가 하나의 예시임

[그림2] 바이두 검색 내 AI 생성 컨텐츠 비중




자료: Baidu, 유안타증권 리서치센터

[그림3] 바이두 AI-Native 마케팅 비중 증가 추이




자료: Baidu, 유안타증권 리서치센터

[표 2] Valuation Table (클라우드 인텔리전스 서비스 Peer 기준) (단위: 십억위안, %, 배)

| | | Baidu | Alibaba | Tencent | Amazon | Microsoft | Alphabet | Oracle |
|---|---|---|---|---|---|---|---|---|
| 시가총액 | | 357.9 | 2,704.3 | 4,411.9 | 15,353.3 | 21,207.2 | 26,634.5 | 3,174.6 |
| 매출액 | 2024 | 134.6 | 941.2 | 609.0 | 4,593.4 | 1,770.7 | 2,520.2 | 382.2 |
| | 2025E | 132.5 | 999.9 | 657.6 | 4,959.1 | 2,002.5 | 2,370.0 | 410.4 |
| | 2026E | 129.1 | 1,034.7 | 751.2 | 5,555.8 | 2,271.7 | 2,794.5 | 462.9 |
| | 2027E | 134.7 | 1,146.0 | 831.2 | 6,199.9 | 2,622.5 | 3,223.9 | 601.5 |
| 영업이익 | 2024 | 21.9 | 113.4 | 208.1 | 574.6 | 927.7 | 927.2 | 127.7 |
| | 2025E | 24.7 | 136.8 | 211.2 | 555.7 | 902.7 | 910.8 | 178.6 |
| | 2026E | 13.9 | 86.4 | 246.3 | 687.2 | 1,052.5 | 1,093.6 | 194.8 |
| | 2027E | 15.6 | 134.3 | 285.7 | 848.0 | 1,211.7 | 1,361.6 | 239.0 |
| OPM (%) | 2024 | 16.2 | 12.0 | 34.2 | 12.5 | 52.4 | 36.8 | 33.4 |
| | 2025E | 18.6 | 13.7 | 32.1 | 11.2 | 45.1 | 38.4 | 43.5 |
| | 2026E | 10.7 | 8.3 | 32.8 | 12.4 | 46.3 | 39.1 | 42.1 |
| | 2027E | 11.6 | 11.7 | 34.4 | 13.7 | 46.2 | 42.2 | 39.7 |
| 당기순이익 | 2024 | 23.8 | 130.1 | 194.1 | 558.1 | 735.0 | 949.7 | 89.9 |
| | 2025E | 24.9 | 156.3 | 218.1 | 659.9 | 718.9 | 943.8 | 123.3 |
| | 2026E | 18.2 | 100.0 | 260.4 | 719.5 | 867.8 | 980.2 | 148.7 |
| | 2027E | 19.6 | 142.9 | 293.0 | 852.6 | 979.8 | 1,174.6 | 160.3 |
| PER (X) | 2024 | 16.5 | 10.2 | 18.7 | 39.0 | 37.8 | 23.9 | 30.2 |
| | 2025E | 14.3 | 17.6 | 21.6 | 23.6 | 30.8 | 29.2 | 26.7 |
| | 2026E | 17.2 | 27.4 | 17.4 | 22.6 | 24.7 | 26.8 | 21.7 |
| | 2027E | 15.2 | 18.7 | 15.4 | 18.9 | 21.7 | 22.6 | 20.0 |

자료: Bloomberg, 유안타증권 리서치센터 / 주: 1) 중국, 미국 단위 모두 십억위안으로 통일, 2) 2026/02/11 종가 기준

## 바이두 (9888.HK) 재무제표 (GAAP 연결)

| 손익계산서 | | | | | (단위: 십억위안) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| **매출액** | **107** | **124** | **124** | **135** | **133** |
| 매출원가 | 55 | 64 | 64 | 65 | 66 |
| 매출총이익 | 52 | 60 | 60 | 70 | 67 |
| 판관비 | 18 | 25 | 21 | 24 | 24 |
| **영업이익** | **14** | **11** | **16** | **22** | **21** |
| EBITDA | 23 | 20 | 26 | 33 | 32 |
| 영업외손익 | -9 | 0 | 6 | -3 | -7 |
| 외환관련손익 | 1 | 0 | 1 | -1 | -1 |
| 이자손익 | -2 | -2 | -3 | -5 | -5 |
| 관계기업관련손익 | 2 | 1 | 2 | 4 | 1 |
| 기타 | -9 | 1 | 6 | -2 | -2 |
| 법인세비용차감전순손익 | 23 | 11 | 10 | 25 | 29 |
| 법인세비용 | 4 | 3 | 3 | 4 | 4 |
| 계속사업순손익 | 19 | 8 | 8 | 22 | 24 |
| 중단사업순손익 | 0 | 0 | 0 | 0 | 0 |
| 당기순이익 | 19 | 8 | 8 | 22 | 24 |
| **지배지분순이익** | **22** | **10** | **8** | **20** | **24** |
| 포괄순이익 | 24 | 10 | 8 | 19 | 23 |
| 지배지분포괄이익 | - | - | - | - | - |

주: 이익 산출 기준은 기존 k-GAAP과 동일. 즉, 매출액에서 매출원가와 판관비만 차감

| 현금흐름표 | | | | | (단위: 십억위안) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| **영업활동 현금흐름** | **24** | **20** | **26** | **37** | **21** |
| 당기순이익 | 22 | 10 | 8 | 20 | 24 |
| 감가상각비 | 6 | 6 | 7 | 8 | 7 |
| 외환손익 | 1 | 0 | 1 | -1 | -1 |
| 종속, 관계기업관련손익 | 2 | 1 | 2 | 4 | 1 |
| 자산부채의 증감 | - | - | - | - | - |
| 기타현금흐름 | -7 | 3 | 8 | 5 | -9 |
| **투자활동 현금흐름** | **-28** | **-31** | **-4** | **-50** | **-9** |
| 투자자산 | 203 | 211 | 200 | 241 | 243 |
| 유형자산 증가 (CAPEX) | -5 | -11 | -8 | -11 | -8 |
| 유형자산 감소 | 0 | 0 | 0 | 0 | 0 |
| 기타현금흐름 | -225 | -231 | -195 | -281 | -243 |
| **재무활동 현금흐름** | **4** | **3** | **3** | **4** | **4** |
| 단기차입금 | 8 | 4 | 14 | 13 | 11 |
| 사채 및 장기차입금 | 60 | 68 | 63 | 57 | 52 |
| 자본 | 189 | 217 | 238 | 263 | 284 |
| 현금배당 | 0 | 0 | 0 | 0 | 0 |
| 기타현금흐름 | -253 | -286 | -312 | -329 | -342 |
| 연결범위변동 등 기타 | 1 | 19 | -7 | -18 | -18 |
| **현금의 증감** | **2** | **11** | **18** | **-28** | **-1** |
| 기초 현금 | 34 | 37 | 48 | 65 | 38 |
| 기말 현금 | 37 | 48 | 65 | 38 | 37 |
| NOPLAT | **12** | **7** | **12** | **19** | **18** |
| FCF | **19** | **9** | **18** | **25** | **13** |

자료: 유안타증권

주: 1. EPS, BPS 및 PER, PBR은 지배주주 기준임
2. PER등 valuation 지표의 경우, 확정치는 연평균 주가 기준, 전망치는 현재주가 기준임
3. ROE, OA의경우, 자본, 자산 항목은 연초, 연말 평균을 기준일로 함

| 재무상태표 | | | | | (단위: 십억위안) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| **유동자산** | **183** | **213** | **213** | **230** | **169** |
| 현금및현금성자산 | 36 | 37 | 53 | 25 | 25 |
| 매출채권 및 기타채권 | 9 | 10 | 12 | 11 | 10 |
| 재고자산 | 0 | 0 | 0 | 0 | 0 |
| 비유동자산 | 149 | 167 | 178 | 177 | 259 |
| 유형자산 | 31 | 39 | 34 | 39 | 41 |
| 관계기업등 지분관련자산 | - | - | - | - | - |
| 기타투자자산 | 10 | 27 | 32 | 33 | 45 |
| **자산총계** | **333** | **380** | **391** | **407** | **428** |
| 유동부채 | 68 | 74 | 80 | 76 | 81 |
| 매입채무 및 기타채무 | 37 | 41 | 38 | 38 | 41 |
| 단기차입금 | 8 | 4 | 14 | 13 | 11 |
| 유동성장기부채 | 7 | 11 | 7 | 6 | 8 |
| 비유동부채 | 72 | 82 | 74 | 68 | 63 |
| 장기차입금 | 60 | 68 | 63 | 57 | 52 |
| 사채 | 0 | 0 | 0 | 0 | 0 |
| **부채총계** | **141** | **156** | **153** | **144** | **144** |
| 지배지분 | 183 | 211 | 223 | 244 | 264 |
| 자본금 | 0 | 0 | 0 | 0 | 0 |
| 자본잉여금 | 47 | 74 | 80 | 87 | 92 |
| 이익잉여금 | 135 | 145 | 148 | 161 | 180 |
| 비지배지분 | 9 | 12 | 14 | 19 | 20 |
| **자본총계** | **192** | **224** | **238** | **263** | **284** |
| **순차입금** | **-156** | **-156** | **-106** | **-182** | **-147** |
| 총차입금 | 83 | 92 | 91 | 85 | 79 |

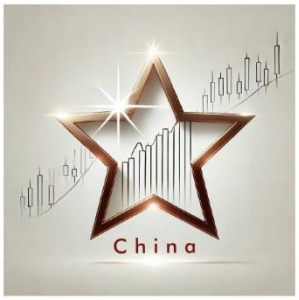
| Valuation 지표 | | | | | (단위: 위안, 배, %) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| EPS | 66.7 | 29.7 | 21.7 | 55.8 | 66.4 |
| BPS | 534.4 | 612.0 | 646.7 | 694.8 | 763.1 |
| EBITDAPS | 71.8 | 58.4 | 75.3 | 104.4 | 60.9 |
| SPS | 317.8 | 361.1 | 355.6 | 383.6 | 381.7 |
| DPS | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| **PER** | **31.1** | **27.2** | **21.1** | **16.5** | **10.1** |
| **PBR** | **2.6** | **1.5** | **1.2** | **1.2** | **0.8** |
| **EV/EBITDA** | **14.5** | **9.4** | **7.0** | **4.1** | **2.7** |
| **PSR** | **4.4** | **2.6** | **2.2** | **2.2** | **1.6** |

| 재무비율 | | | | | (단위: 배, %) |
|---|---|---|---|---|---|
| 결산(12월) | 2020A | 2021A | 2022A | 2023A | 2024A |
| **매출액 증가율 (%)** | **-0.3%** | **16.3%** | **-0.7%** | **8.8%** | **-1.1%** |
| **영업이익 증가율 (%)** | **127.4%** | **-26.7%** | **51.3%** | **37.4%** | **-2.7%** |
| **지배순이익 증가율 (%)** | **992.5%** | **-54.5%** | **-26.1%** | **168.8%** | **17.0%** |
| 매출총이익률 (%) | 48.5% | 48.3% | 48.3% | 51.7% | 50.3% |
| **영업이익률 (%)** | **13.4%** | **8.4%** | **12.9%** | **16.2%** | **16.0%** |
| 지배순이익률 (%) | 17.8% | 6.1% | 6.1% | 16.0% | 18.2% |
| EBITDA 마진 (%) | 21.7% | 15.7% | 21.0% | 24.2% | 23.9% |
| ROIC | 4.5 | 2.5 | 3.6 | 5.5 | 5.0 |
| ROA | 7.1 | 2.9 | 2.0 | 5.1 | 5.7 |
| **ROE** | **13.0** | **5.2** | **3.5** | **8.4** | **9.1** |
| 부채비율 (%) | 4305% | 4086% | 3842% | 3221% | 2797% |
| 순차입금/자기자본 (%) | -8123% | -6962% | -4469% | -6928% | -5171% |
| 영업이익/금융비용 (배) | 4.6 | 3.1 | 5.5 | 6.7 | 7.5 |

## Appendix

- 이 자료에 게재된 내용들은 본인의 의견을 정확하게 반영하고 있으며 타인의 부당한 압력이나 간섭 없이 작성되었음을 확인함. (작성자: 백종민)
- 당사는 자료공표일 현재 동 종목 발행주식을 1%이상 보유하고 있지 않습니다.
- 당사는 자료공표일 현재 해당 기업과 관련하여 특별한 이해관계가 없습니다.
- 당사는 동 자료를 전문투자자 및 제 3자에게 사전 제공한 사실이 없습니다.
- 동 자료의 금융투자분석사와 배우자는 자료공표일 현재 대상법인의 주식관련 금융투자상품 및 권리를 보유하고 있지 않습니다.
- 종목 투자등급 (Guide Line): 투자기간 12개월, 절대수익률 기준 투자등급 4단계(Strong Buy, Buy, Hold, Sell)로 구분한다
- Strong Buy: +30%이상 Buy: 15%이상, Hold: -15% 미만 ~ +15% 미만, Sell: -15%이하로 구분
- 업종 투자등급 Guide Line: 투자기간 12개월, 시가총액 대비 업종 비중 기준의 투자등급 3단계(Overweight, Neutral, Underweight)로 구분
- 2014년 2월21일부터 당사 투자등급이 기존 3단계 + 2단계에서 4단계로 변경

YUANTA TOP FOCUS




유안타증권
Yuanta Securities (Korea)