

Strategy Idea

[중국 LLM 시리즈] 미국의 가격 공세 vs. 중국의 CapEx 러시



글로벌 투자전략 - 신흥국
Analyst 최설화
xuehua.cui@meritz.co.kr

- ✓ 중국기업 약진에 미국 LLM 가격 인하로 대응 → 미국 토큰 사용량 증가율 소폭 반등
- ✓ 중국 빅테크 2분기 실적에서 AI CapEx 확대 및 지출 속도 가속화, 미중 AI 컴퓨팅 경쟁 본격화
- ✓ 유상증자, IPO 등 외부자본 조달 확대 → 반도체 등 AI 인프라 수혜 논리 강화, 커황50 선호

미국 LLM 모델 가격 인하

7월 Kimi K3를 비롯한 중국 LLM 모델의 성능이 미국 프런티어 수준으로 올라오며, 미국 LLM들은 모델 가격 인하로 대응했다.

7월 30일 OpenAI는 GPT-5.6 Luna의 가격을 기존 대비 80% 인하해 1M 입력 토큰당 0.20달러(기존 1달러)로 낮췄고, 중간 등급 모델인 Terra도 입력 2달러 수준으로 조정했다. 8월 중순에는 최상위 모델인 Sol의 가격도 추가 인하하며 앤스로픽의 Opus 라인업을 정면으로 겨냥했다.

앤스로픽 역시 가격 정책을 전격 수정했다. Claude Sonnet 5의 가격을 기존 정가(입력 3달러/출력 15달러) 대비 각각 33%, 33% 인하한 입력 2달러/출력 10달러로 책정했다. 당초 이 요금은 8월 31일까지만 한시적으로 적용되고 9월 1일부터 다시 인상될 예정인 도입 프로모션이었으나, 앤스로픽이 8월 10일 이를 영구 적용(Permanent)하기로 공식 전환했다.

결과적으로 8월 이후 미국 진영의 가격 인하는 단기 프로모션에 그치지 않고, 프런티어 모델의 가격 기준점 자체가 하향 평준화되는 본격적인 구조적 치킨게임 국면 진입을 시사한다.

표1 미국 프런티어 LLM 가격 인하로 중국의 가성비 모델에 대응

기업명	모델명	기존 가격 (\$/Mtoken)		조정 후 가격 (\$/Mtoken)		인하폭 (%)		비고
		입력	출력	입력	출력	입력	출력	
Anthropic	Claude Sonnet5	3.0	15.0	2.0	10.0	-33	-33	8월 10일 인하
OpenAI	GPT-5.6 Terra	2.5	15.0	2.0	12.0	-20	-20	7월 30일 인하
	GPT-5.6 Luna	1.0	6.0	0.2	1.2	-80	-80	
	GPT-5.6 Sol	5.0	30.0	2.5	15.0	-50	-50	최상위모델, 한시적 인하

자료: 각 사, 메리츠증권 리서치센터

미국의 가격 방어전: 효과가 있었나?

필자는 8월 글로벌 토큰 사용량의 변화에 주목한다. 미국의 대표 모델들이 7월 말부터 전격적인 가격 인하에 돌입했기 때문이다.

OpenRouter에 따르면, 8월 일평균 토큰 사용량은 약 11조 개로 전월 대비 36.8% 증가했다. 동기간 진영별 증가율을 살펴보면 중국 모델이 21.7% 늘어난 반면, 미국 모델은 47.9% 증가했다(표 2). 미국의 가격 인하 조치가 트래픽 방어에 일정 부분 효과를 거두었음을 시사한다. 특히 경량 모델(Luna)의 단가를 무려 80% 인하한 OpenAI의 경우 8월 일평균 토큰 사용량이 전월 대비 146% 급증했다.

그러나 이번 방어전의 한계 역시 뚜렷하다. API 인하폭(-33%)이 적은 고성능 라인업 중심인 앤스로픽의 토큰 사용량은 8월에 오히려 21.7% 감소했다. 또한 딥시크는 8월에 가격을 인상했음에도 불구하고, 유사한 성능을 지닌 미국 모델 가격의 1/4로 절대 단가 우위를 유지한 덕분에 토큰 사용량은 7월대비 99.8% 늘었다. 한편 8월 중순 OpenAI의 최상위 모델인 GPT-5.6 Sol은 한시적 가격 인하로 사용량이 증가했으나, Kimi K3 역시 동반 상승세를 보이며 중국 모델 추월에는 실패했다. 플랫폼 내 점유율 기준으로 중국 모델의 비중은 56%로 여전히 과반을 차지하고 있다(표 2).

결국 사용자들은 가격의 '인하율'보다 토큰당 '절대 가격'에 더 민감하게 반응했다. 미국의 공격적인 단가 인하는 점유율의 급격한 이탈 속도를 늦추는 완충 효과는 있었으나, 원가 경쟁력의 절대적 열세를 뒤집지는 못했다. 중국 모델을 추월하려면 미국 모델들은 추가 가격 인하를 해야 하며, 이 경우 7월 말에 발간된 'China AI Attack'에서 주장한 것처럼 중국발 LLM 디플레이션 압력은 당분간 지속될 전망이다.

표1 7~8월 LLM별 일평균 토큰 사용량 비교

(십억토큰)	Anthropic	Google	Deepseek	OpenAI	MiniMax	xAI	Xiaomi	Tencent	Z.ai	Alibaba	Moonshot	기타	중국 합산	미국 합산	합계
7월	901	605	1,335	481	482	48	1,183	945	506	55	163	1,310	4,669	2,034	8,013
8월	705	860	2,667	1,182	262	62	954	1,192	532	64	211	2,269	5,682	3,009	10,959
증감폭	-21.7	42.2	99.8	146.0	-45.5	29.8	-19.4	26.1	5.0	15.9	29.3	73.2	21.7	47.9	36.8

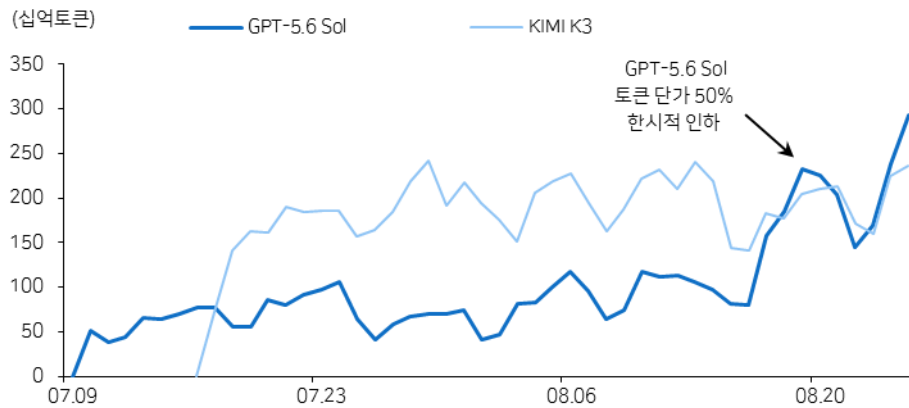
주: 8월은 23일까지의 일평균 기준으로 제시 / 자료: OpenRouter, 메리츠증권 리서치센터

표2 LLM별 일평균 토큰 사용 점유율

구분	Anthropic	Google	DeepSeek	OpenAI	xAI	MiniMax	Xiaomi	Tencent	Z.ai	Alibaba	Moonshot	기타	중국 합산
3Q25	18%	24%	14%	9%	15%	0%			2%	5%	1%	12%	22%
4Q25	14%	20%	6%	10%	26%	3%	1%		2%	3%	0%	14%	16%
1Q26	14%	18%	7%	10%	6%	10%	7%		5%	2%	4%	18%	34%
2Q26	15%	12%	14%	7%	1%	7%	6%	8%	4%	4%	3%	19%	46%
3Q26	9%	8%	21%	8%	1%	4%	12%	11%	6%	1%	2%	19%	56%

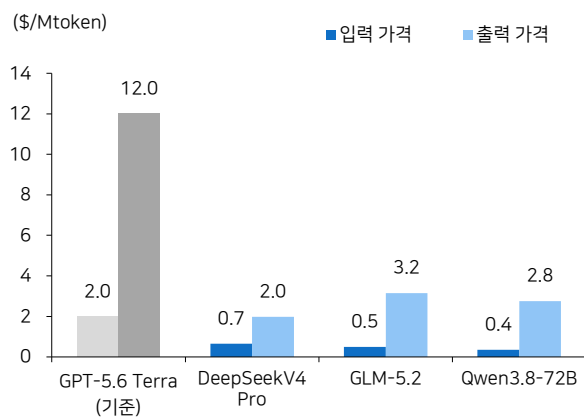
주: 3Q26은 2026.8.23일까지 누적 기준으로 제시
자료: OpenRouter, 메리츠증권 리서치센터

그림1 OpenAI의 최상위 모델 가격 인하 후 사용량 증가, 단 중국 모델 토큰 사용량은 견고



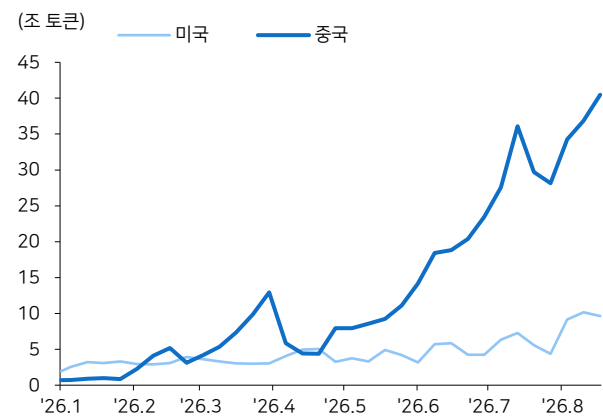
자료: OpenRouter, 메리츠증권 리서치센터

그림2 GPT-5.6 Terra와 비슷한 성능의 중국 LLM 가격은 1/4



자료: OpenRouter, 메리츠증권 리서치센터

그림3 미국과 중국 LLM 모델의 토큰 사용량 추이



자료: OpenRouter, 메리츠증권 리서치센터

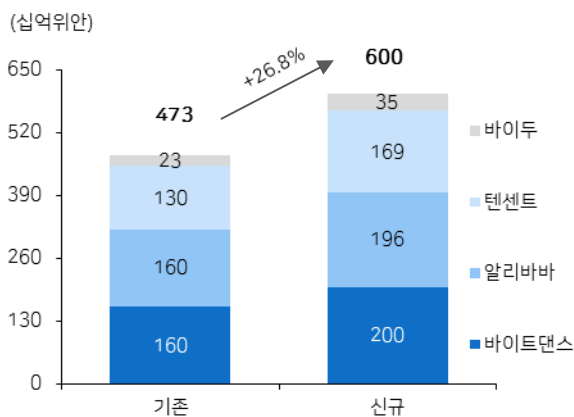
중국 빅테크 2분기 실적 발표 공통점: All in AI

2분기 중국 빅테크들의 실적 발표를 관통하는 공통점은 AI 자본지출의 전면적 확대이다. 텐센트의 2분기 자본지출은 전분기대비 65%, 전년동기대비 176% 급증한 528억 위안(약 78억 달러)을 기록했다. 알리바바 역시 2분기 CapEx가 전년동기대비 75% 증가한 677억 위안(약 100억 달러)에 달했고, 바이두도 전년동기대비 200% 증가한 114억 위안(약 17억 달러)을 집행하며 예외 없는 'AI 올인' 기조를 재확인시켰다.

이 같은 투자 확대는 연간 전망치에도 고스란히 반영되고 있다. 시장이 예상했던 2026년 중국 4대 빅테크의 합산 CapEx는 당초 4,730억 위안 수준이었으나, 2분기 실적 발표를 거치며 6,000억 위안으로 26.8% 상향 조정됐다. 이는 2025년 실제 자본지출 대비 65.5% 증가한 규모로, 2025년 연간 증가율(49.4%)을 웃돈다. 즉 중국 빅테크의 CapEx 사이클은 단순히 확대되는 방향성을 넘어, '증가 속도 자체가 가속화'되는 국면으로 진입했다고 평가할 수 있다.

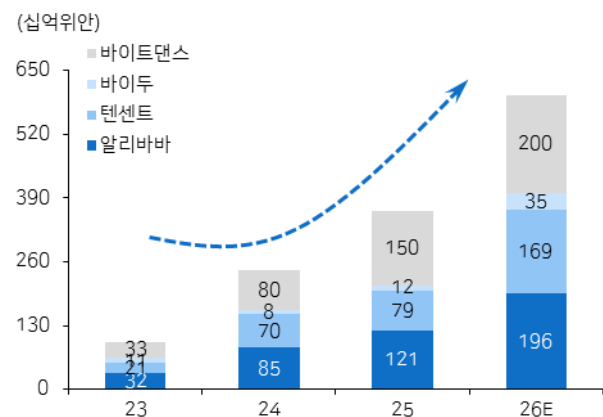
이러한 투자 가속화의 배경으로는 3가지를 꼽을 수 있다. 1) 올해 메모리를 포함한 글로벌 반도체 가격이 빠르게 상승하고 있고, 2) 자국 LLM 성능의 상향 평준화와 Agent 기반 워크로드 확산으로 토큰 사용량이 구조적으로 증가하고 있으며, 3) 국산 반도체 공급이 여전히 병목 요인이지만 점진적으로 개선되면서 실제 집행 가능한 투자 규모가 확대되고 있다는 점이다. 일각에서는 2027년 중국 빅테크의 자본지출 증가율이 올해 수준을 상회할 것이라는 낙관론도 힘을 얻고 있다.

그림4 2026년 중국 4대 빅테크들의 CapEx 전망치 +27%



주: 상장사들의 2분기 실적 발표 전후의 전망치 변화
자료: 각 사, 메리츠증권 리서치센터

그림5 2026년 중국 4대 빅테크의 연간 CapEx 가속화



자료: 각 사, Bloomberg, 메리츠증권 리서치센터

중국 LLM들의 유상증자 및 상장 러시

중국 기업들의 AI 자본지출 속도가 빨라지면서, 미국 하이퍼스케일러들처럼 CapEx 슈퍼 사이클을 뒷받침하기 위한 외부 자본 동원이 재현되고 있다. 알리바바의 유상증자, 비상장 LLM들의 상장 러시가 대표적이다.

알리바바 유상증자: 8월 23일 동사는 신주 7.1억주를 주당 112.7 HKD에 발행하는 800억 홍콩달러(약 102억 달러) 규모의 유상증자를 발표했다. 26일 결제가 완료될 예정인 이번 증자는 약 3배의 초과청약을 기록했으며, 주문의 40% 이상이 중동, 유럽, 아시아의 국부펀드 및 장기 투자자로부터 유입됐다. 알리바바는 조달 자금을 100% 풀스택 AI 역량 강화와 AI 인프라 확충에 투입할 것임을 명시했다. 알리바바는 본업의 부진으로 내부 현금으로만 컴퓨팅 투자 속도를 감당하지 못해 외부 자본시장으로 손을 뻗기 시작했다.

문샷AI IPO: 중국 비상장 LLM들의 상장도 가속화되고 있다. Kimi K3(2.8조 파라미터)로 잘 알려진 베이징 소재의 문샷AI는 오는 9월 말까지 홍콩거래소에 예비심사청구서를 제출하고 내년 1분기 상장을 추진 중이다. 기업가치는 500억 달러로, 딥시크 및 즈푸 대비 상대적으로 낮은 밸류에이션을 보이지만, 상장을 통한 컴퓨팅 자원 확보라는 목적은 동일하다.

딥시크 IPO: 지난 6월 74억 달러의 1차 펀딩을 거치고, 최근 추가적으로 74억 달러의 2차 펀딩에 착수했다. 당시 텐센트, JD닷컴, CATL 등 상업적 투자자들은 5년 락업과 의결권 없는 조건을 받아들인 반면, 중국 국가급AI 산업 투자기금은 의결권을 확보하고 락업 없이 투자한 것으로 전해져 이번 라운드의 실질적 주도권이 상업 자본보다 국가자본과 창업자 양원평에 있음을 시사한다. 딥시크는 올해 안에 중국 본토의 과창판 시장에 상장예비심사 신청을 목표로 준비 중에 있으며, 실제 상장은 2027년 2분기 유력하게 거론되고 있다.

즈푸 본토IPO: 동사는 이미 지난 1월 홍콩거래소에 상장을 완료했고 한때 시가총액이 830억 달러까지 치솟았다. 상장 이후에도 자금 수요는 줄지 않아, 6월 1일 이사회 결의를 통해 상하이 과창판에 A주를 추가 발행(전체 발행주식의 2~8%, 최대 150억 위안 조달)하는 이중상장 전략을 공식화했다. 조달 자금의 80%인 120억 위안은 범용 파운데이션 모델 R&D에, 30억 위안은 MaaS(Model-as-a-Service) 플랫폼 구축에 배정될 예정이다.

표3 중국 프런티어 LLM 기업들이 상장 일정 및 기업가치

기업명	상장시장	예비심사청구서 제출 계획	상장 시기(예상)	기업가치 (억달러)	26년 ARR (억달러)	비고
문샷AI	홍콩	26년 9월 말	1Q27	500	3~4	7월 Pre-IPO 밸류
DeepSeek	본토 과창판	26년 12월 말	2Q27	710	4~5	8월 Pre-IPO 밸류
Z.ai	본토 과창판	26년 9월	1Q27	716	10	26년 1월 홍콩증시 상장

주: 즈푸의 기업가치는 8/26일 홍콩주식의 시가총액 기준
 자료: 각 사, 현지 언론, 메리츠증권 리서치센터

중국 AI 투자전략: 여전히 AI 인프라(반도체·전력) > LLM 기업

8월 들어 나타난 변화를 정리하면 다음과 같다. 첫째, 미국 프런티어 모델들의 가격 인하에 토큰 사용량 증가율은 개선되고 있으나, 이것이 중국 모델의 절대적 점유율을 반전시킬 만큼의 위협 요인은 아직 아니다.

둘째, 중국 LLM 기업들은 모델 성능 개선과 토큰 사용량 증가에 발맞춰 AI 자본 지출을 가속화하고 있으며, 유상증자 및 IPO 러시를 통한 외부 자본 조달을 확대하고 있다. 이는 미국 하이퍼스케일러들이 CapEx 슈퍼사이클을 뒷받침하기 위해 외부 자본을 적극 동원했던 패턴과 궤를 같이 한다.

즉 현재 미중 AI 경쟁은 여전히 국가 대항전 성격의 컴퓨팅 확보 경쟁 구도로 진행되고 있으며, 이번에 조달되는 중국 AI 기업들의 자금 역시 결국 중국산 반도체를 포함한 컴퓨팅 인프라로 흘러갈 가능성이 높다. 지난 7월 발간된 'China AI Attack' 보고서에서 제시한 바와 같이, 중국 LLM 경쟁력 확대의 최대 수혜는 반도체, 전력 등 AI 인프라 쪽에 우선적으로 귀속될 것이라는 기존 견해를 유지한다. 아울러 중국 LLM 기업들의 지속적인 학습 수요는 해외 네오클라우드 기업들에도 수혜로 작용할 전망이다.

반면 중국 LLM 기업에 대한 투자의견은 여전히 중립을 유지한다. LLM 산업의 구조적 특징은 낮은 스위칭 비용에 있다. API 한 줄만 교체하면 모델 전환이 가능한 환경에서는 한 차례 최고 성능 모델을 선보였다 해도 그 우위가 영속되지 않으며, 프런티어 모델을 지속적으로 갱신해야 하는 부담이 상존한다. 이에 따라 LLM 기업들은 뚜렷한 수익성이 확보되기 전까지 외부 자금 조달을 이어갈 수밖에 없는 구조다. 또한 중국 LLM 경쟁은 아직 초기 단계로 최종 승자를 예단하기 이르며, 분기별 상장사들의 수익성 및 비즈니스 모델 변화를 꾸준히 점검할 필요가 있다.

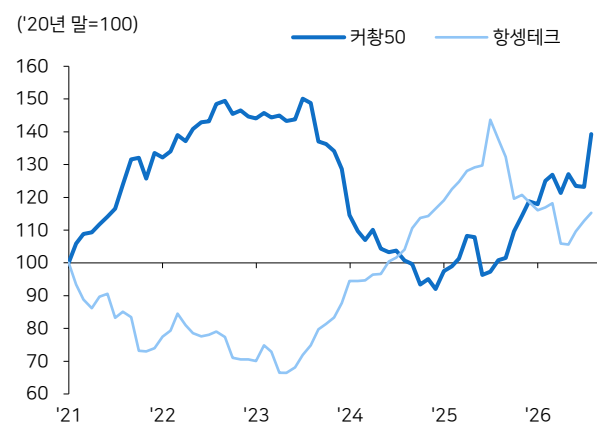
위와 같은 산업 변화를 감안해, 중국 주식시장 투자 대상을 압축한다면 반도체 중심의 커황50(Star50) 지수를 항셱테크 대비 선호한다.

그림6 커황50과 항셱테크 지수 추이



자료: Bloomberg, 메리츠증권 리서치센터

그림7 커황50과 항셱테크 지수의 12개월 예상 EPS 전망 추이



자료: Bloomberg, 메리츠증권 리서치센터

Compliance Notice

- 본 조사분석자료는 제3자에게 사전 제공된 사실이 없습니다.
- 당사는 자료작성일 현재 본 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
- 본 자료를 작성한 애널리스트는 자료작성일 현재 추천 종목과 재산적 이해관계가 없습니다.
- 본 자료에 게재된 내용은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 신의 성실하게 작성되었음을 확인합니다.

본 자료는 투자자들의 투자판단에 참고가 되는 정보제공을 목적으로 배포되는 자료입니다. 본 자료에 수록된 내용은 당사 리서치센터의 추정치로서 오차가 발생할 수 있으며 정확성이나 완벽성은 보장하지 않습니다. 본 자료를 이용하시는 분은 본 자료와 관련한 투자의 최종 결정은 자신의 판단으로 하시기 바랍니다. 따라서 어떠한 경우에도 본 자료는 투자 결과와 관련한 법적 책임소재의 증빙자료로 사용될 수 없습니다. 본 조사분석자료는 당사 고객에 한하여 배포되는 자료로 당사의 허락 없이 복사, 대여, 배포 될 수 없습니다.
