

Market Issue

하이퍼스케일러로의 분산

- 메타의 클라우드 사업 진출로 capex 우려 촉발. 그러나 AI 기업들로부터 들어오는 수주 잔고(RPO)는 증가세. 반도체 상승 흐름은 결국 재개될 전망
- 다만 높아진 AI 비용에 대한 기업들의 대응은 계속 모니터링 필요. 토큰 비용 축소 전략 확산으로 인한 AI 모델 기업들의 재무환경 악화가 리스크. 수주잔고 증가세 위협 요인
- 하이퍼스케일러로의 분산 필요. 상대적으로 가성비 모델을 보유한 구글이 유리

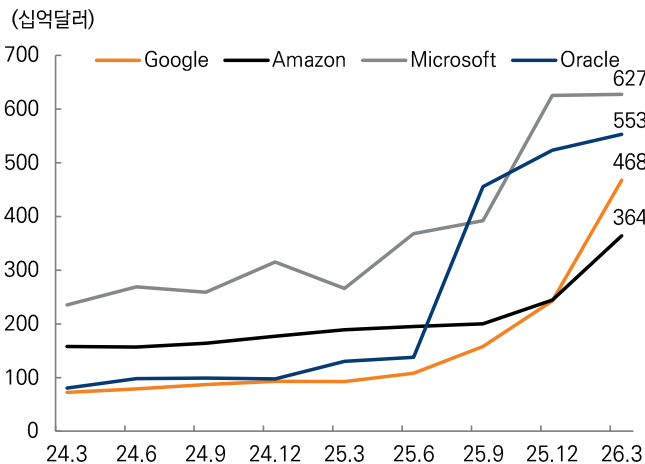
메타의 잉여 컴퓨팅 캐파 활용 계획: 다른 클라우드 capex에는 제한적

메타가 잉여 컴퓨팅 캐파를 활용해 클라우드 사업에 진출할 계획이 있다고 발표했다. 자체적으로 필요한 데이터센터 캐파를 제외한 여분을 판매하겠다는 계획이다. 해당 계획은 메타에는 긍정적이다. 데이터센터를 더 효율적으로 활용해 지금까지의 AI 투자에 대한 수익성을 높일 수 있기 때문이다. 계획이 현실화되면 메타는 AI 관련 수익원으로 광고 사업 개선 외에 클라우드 사업까지 제시할 수 있어 수익성 우려를 완화시킬 수 있게 됐다.

반대로 반도체를 비롯한 capex 수혜 분야에는 부정적으로 작용했다. 메타가 과잉 컴퓨팅 캐파를 판매한다는 의미는 메타가 더 이상 이전처럼 공격적으로 capex 증강에 나서지 않을 가능성을 시사한다. 현재 보유한 데이터센터 캐파만으로 자체적으로 추진하고 있는 AI 모델 개발은 이어갈 수 있다는 의미이기 때문이다.

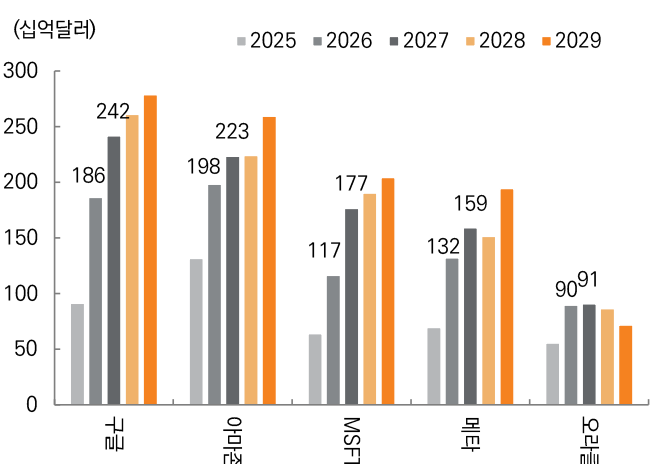
그러나 메타의 계획은 구글, 아마존, 마이크로소프트 등 클라우드 기업들의 capex 계획에는 영향이 제한적일 것으로 판단된다. 해당 기업들은 메타와는 달리 본업이 클라우드인 만큼, OpenAI, Anthropic 등으로부터 받은 컴퓨팅 계약(RPO)를 처리하기 위한 capex를 주로 집행하고 있다. AI 수요가 늘어나는 이상 RPO도 지속적으로 늘어날 가능성이 높고, RPO가 늘어나는 이상 capex 증가세도 지속될 전망이다.

그림 1. 기업들의 RPO 증가 추이



자료: Bloomberg, 미래에셋증권 리서치센터

그림 2. 하이퍼스케일러별 Capex 추이



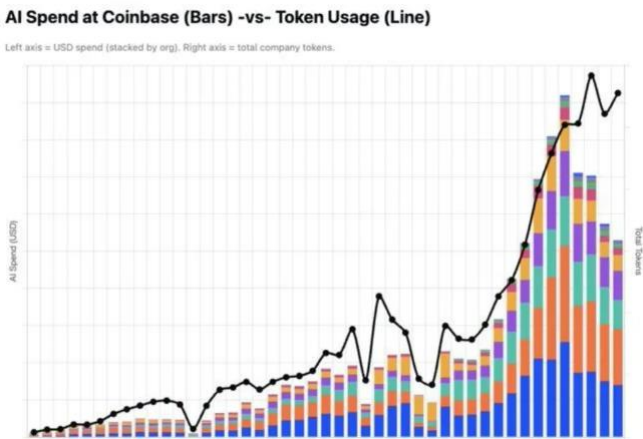
자료: Factset, 미래에셋증권 리서치센터

기업들의 토큰 비용 축소 전략: OpenAI 등 LLM 기업들에 위협적

메타 이외에도 AI의 비용에 대한 기업들의 고민은 지속될 전망이다. 애플의 중국산 칩 사용 시도도 이와 같은 고민의 일환이다. 최근에는 토큰 비용도 문제로 부상했다. 지금까지 기업들은 AI 사용량을 늘리는데(tokenmaxxing) 집중했지만 AI모델 사용료인 토큰 비용이 급격히 오르며 부담이 늘어났다. Uber는 올해 AI 예산을 4개월안에 소진했고 아마존도 AI 사용량을 권장하기 위해 만든 토큰 사용량 리더보드를 폐지했다. 실제로 미국에서 AI를 가장 많이 사용하는 상위 1% 기업들의 직원당 AI 비용은 전년대비 약 4배 상승했다.

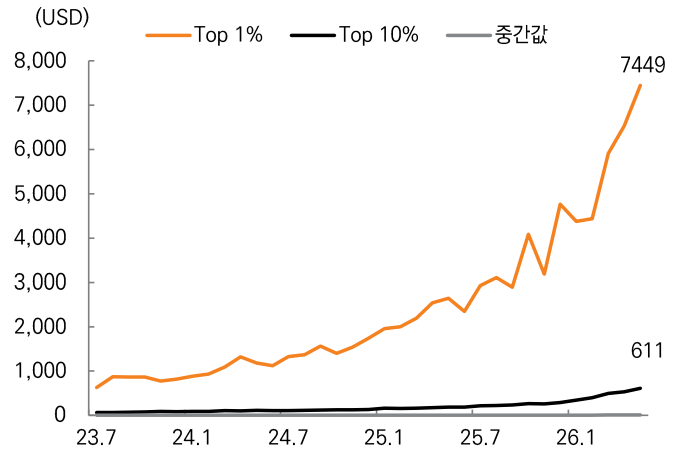
이와 같은 이유로 AI 사용의 비용 감소 방안에 기업들의 초점이 맞춰지는 분위기다. 더 저렴한 LLM 모델을 도입하는 흐름이 대표적이다. Coinbase CEO는 일반 업무에 사용하는 디폴트 모델을 GLM 5.2와 Kimi K2.7 모델로 채택하면서 전체 AI 사용량을 늘리면서 비용은 감소시키는 효과를 봤다고 발표했다. 복잡한 추론 작업이 필요할때에는 성능이 가장 뛰어난 frontier 모델을 사용하지만 대다수의 반복적 일반 작업에는 성능대비 가성비가 높은 모델을 사용하는 ‘모델 routing’ 전략이다. 실제로 AI를 사용하는 기업 수는 늘어나면서먼서도 저가형 모델 사용 증가로 인해 기업들이 지불하는 평균 토큰 가격은 감소 중이다.

그림 3. Coinbase의 AI 비용(바)과 토큰 사용량(검은 선)



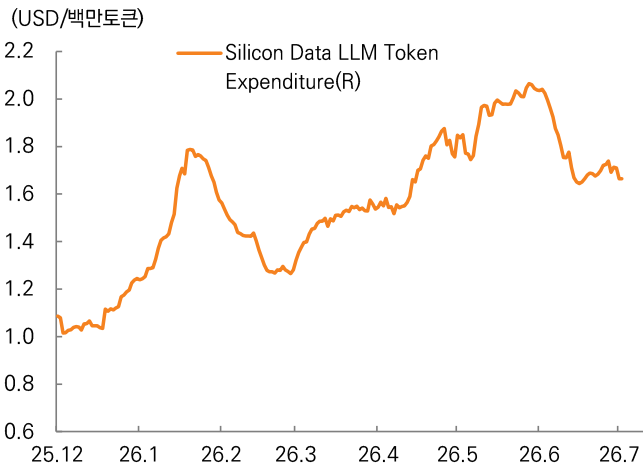
자료: X, 미래에셋증권 리서치센터

그림 4. 늘어나는 직원당 토큰 비용 추이



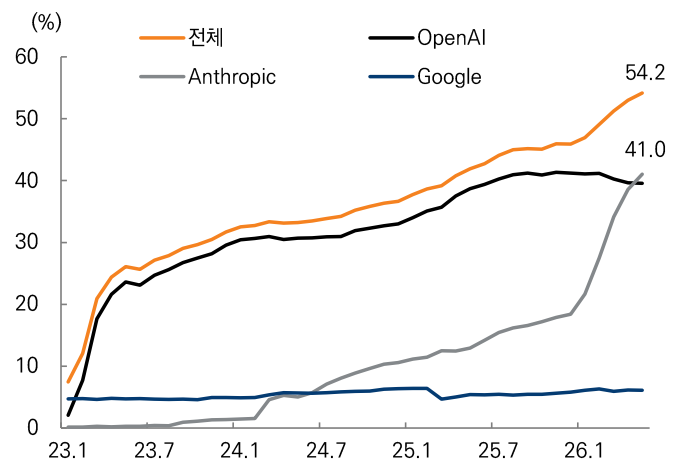
자료: RampAI, 미래에셋증권 리서치센터

그림 5. 기업들의 평균 토큰 가격은 하락 중



자료: Artificial Intelligence Index, 미래에셋증권 리서치센터

그림 6. 미국 기업들의 AI 답습율: AI 수요자체는 증가 중



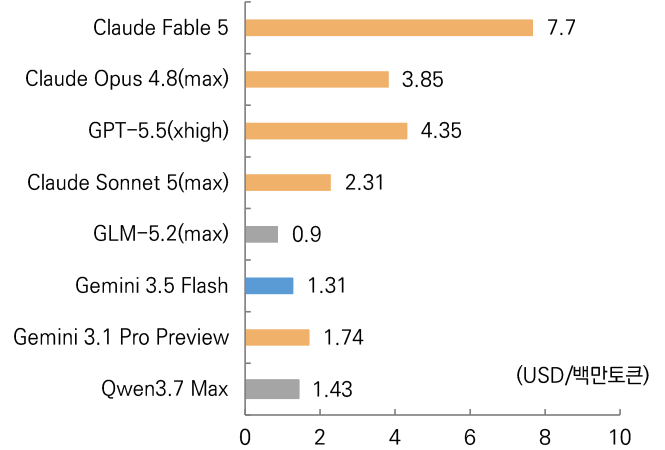
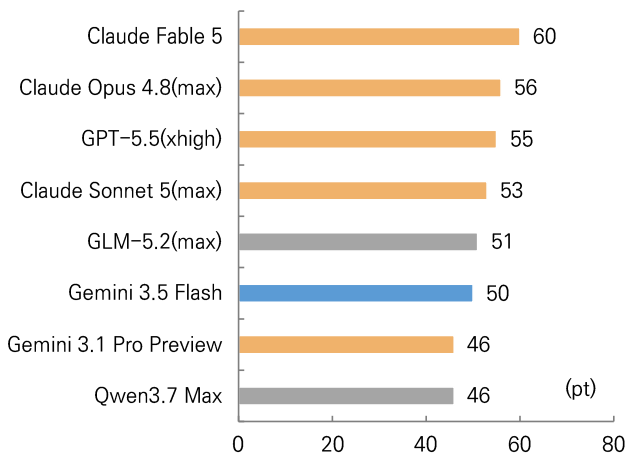
자료: Ramp AI, 미래에셋증권 리서치센터

기업들 사이에서 routing 전략이 확산되면 OpenAI, Anthropic과 같은 frontier 모델들을 보유한 기업들의 피해가 예상된다. 기업들의 tokenmaxxing 구간에 비해 사용량이 하락할 수 밖에 없기 때문이다. 여러 모델들을 비교해볼 경우 성능은 Claude Fable 5, GPT-5.5가 가장 뛰어나지만 가격 측면에서는 불리하다. 특히 중국 Z.ai가 출시한 GLM 5.2 모델은 성능 측면에서는 GPT-5.5 성능에 조금 못미치지만 토큰당 비용은 월등히 적다. Snowflake CEO도 GLM 5.2 모델이 클로드와 유사한 성능을 보이지만 가격은 월등히 낮아 가성비가 높다는 평가를 내렸다. 실제로 미국-중국 경제안보리부 위원회의 보고서에 의하면 80%의 미국 AI 스타트업이 중국 모델을 사용하고 있다고 추정했다.

Routing 전략으로 인한 LLM 모델 기업들의 재무환경 악화가 리스크다. 모델 기업들 모두 향후 매출 추정치에서 토큰 사용 관련 매출 비중이 상당하다. 해당 매출이 악화되면 향후 빅테크 기업들에게 컴퓨팅 파워를 계약할 여력이 하락한다. 잠재적으로 빅테크 기업들의 수주잔고(RPO)를 줄일 수 있는 요인이다. 특히 frontier 모델 성능에서는 엔트로픽에 뒤처지고 가격 경쟁력은 저가형 중국 모델에 뒤처지는 OpenAI가 더 위험하다는 판단이다. OpenAI는 원하는 밸류에이션을 확보하는데 실패해 IPO 연기를 고려 중이기도 하다.

그림 7. LLM 모델들의 성능 비교

그림 8. LLM 모델별 비용 비교

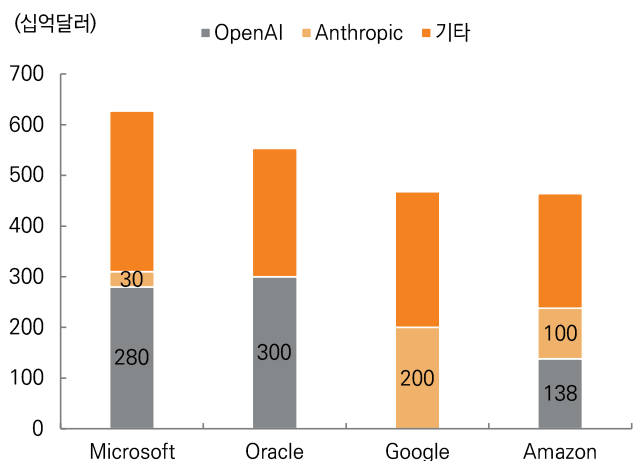
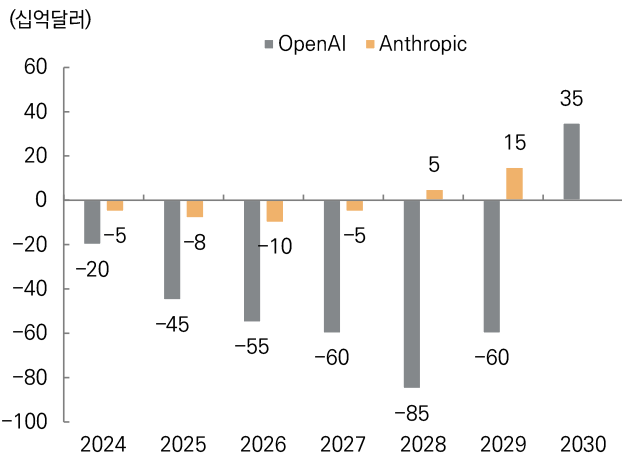


자료: Artificial Intelligence Index, 미래에셋증권 리서치센터

자료: Artificial Intelligence Index, 미래에셋증권 리서치센터

그림 9. OpenAI와 엔트로픽의 예상 손익 추이

그림 10. 하이퍼스케일러 기업들의 수주잔고 비중



자료: WSJ, 미래에셋증권 리서치센터

자료: The Information, 미래에셋증권 리서치센터

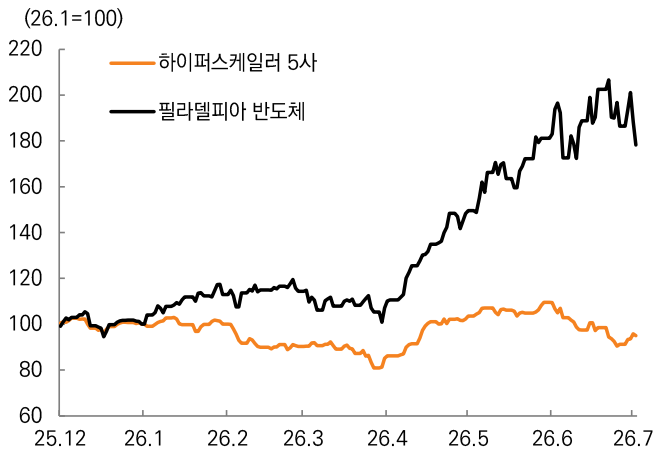
하이퍼스케일러로의 분산 필요: 구글이 상대적으로 유리

메타 사업 계획 발표에도 불구하고 반도체 기업들의 강세 흐름은 지속될 전망이다. 다만 7월 빅테크 실적발표에서 수주잔고(RPO)와 capex 증가세가 재확인되기 전까지 경계심은 유지될 수 있다. 이번 이벤트를 계기로 하이퍼스케일러로의 분산이 필요하다는 판단이다. 만약 우려대로 capex가 예상보다 위축되면 하이퍼스케일러들은 수익성이 개선되는 효과가 기대되고, 기업들 사이 routing 전략이 확산되도 하이퍼스케일러들에게는 유리하게 작용할 수도 있다. 반도체와 달리 올해 주가 상승도 제한적이었던 만큼 가격 부담도 제한적이다.

하이퍼스케일러 기업들에게는 routing 전략이 더 유리할 수 있다. 기업들이 저렴한 모델을 사용하더라도 토큰은 프로세스 해야한다. 하이퍼스케일러 입장에서는 사용자가 Fable5 모델로 토큰을 사용하든 GLM 5.2 모델을 사용하든 차이가 없다. 오히려 저렴한 모델 사용으로 전체 토큰 볼륨이 늘어나면 하이퍼스케일러 기업들의 매출에는 긍정적인 수 있다.

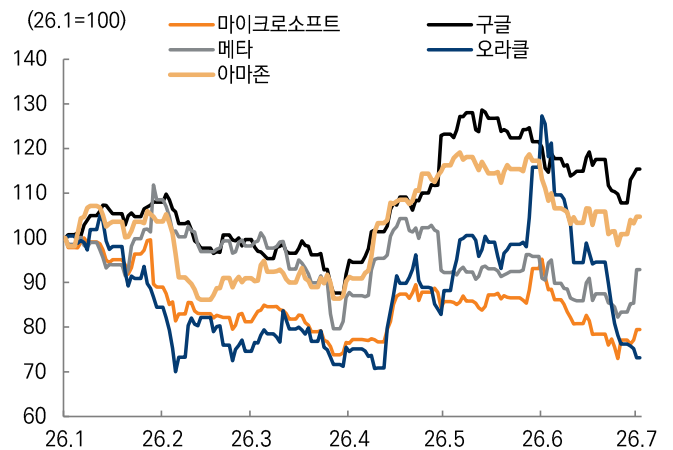
하이퍼스케일러 중에서는 구글이 가장 유리해 보인다. 이미 클라우드 매출 증가율이 가장 높은데다 OpenAI에 대한 노출도도 가장 적다. 또한 Gemini 모델은 기업들의 토큰 비용 감소 흐름에서도 유리하다. 성능 대비 가격이 낮아 가성비가 높기 때문이다. Gemini 3.5 Flash 모델은 지난달 연간 개발자 컨퍼런스(I/O)에서 가성비 모델로 소개되었는데, 토큰 가격은 클로드나 GPT frontier 모델에 비해 월등히 낮다. 이와 같은 가격 경쟁력은 향후 미국 정부 차원에서 중국 LLM 사용을 제한할 경우 더욱 부각될 수 있다. 현재 미국 정부 기관에서는 중국 LLM 모델 사용이 금지되어 있고, 버지니아, 텍사스, 뉴욕 등 여러 주에서도 정부 기기에 중국 모델의 사용을 금지된 상황이다.

그림 11. 하이퍼스케일러 기업들과 반도체 기업들의 주가 추이



자료: Bloomberg, 미래에셋증권 리서치센터

그림 12. 하이퍼스케일러 기업들의 주가 추이



자료: Bloomberg, 미래에셋증권 리서치센터

Compliance Notice

- 당사는 자료 작성일 현재 해당 회사와 관련하여 특별한 이해관계가 없음을 확인합니다.
- 당사는 본 자료를 제3자에게 사전 제공한 사실이 없습니다.
- 본 자료를 작성한 애널리스트는 자료작성일 현재 본 자료에서 매매를 권유한 금융투자상품 및 권리를 보유하고 있지 않습니다.
- 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.