IBKS QUANT IDEA NOTE

엔비디아 실적만큼 중요해질 지표들

Quant 권순호 | 02) 6915-5667 | snowkonn@ibks.com

AI산업 성장 지도, Gemini 3.0으로 변곡점 맞이

OpenAI 와 Nvidia 밸류체인 내의 천문학적 자본 거래는, 역설적으로 AI 산업의 ROIC(투하자본이익률)에 대한 의문을 증폭시키며 버블 우려를 낳았다. 이 국면을 반전시킨 트리거는 구글의 Gemini 3.0 이다. 핵심은 칩(TPU, Tensor Processor Unit)부터 모델, 플랫폼으로 이어지는 수직계열화다. 구글은 이를 통해 성능의 확장과 비용 효율화를 동시에 달성할 수 있을 것이라는 기대감이 시장에 선반영 되기 시작하며, 수익성에 대한 우려를 일부 상쇄해 나가는 흐름을 보이고 있다.

이제 시장의 시선은 단순한 하드웨어 공급(Nvidia)에서 투자금을 상회하는 이익을 위한, 효율적인 서비스 구동으로 확장되고 있다. 물론 엔비디아의 실적과 하이퍼스케일러의 설비투자(CAPEX)는 여전히 산업 성장의 기반이 되는 중요한 지표다. 하지만 향후 수직계열화를 통해 실질적인 마진을 남길 것으로 기대되는 구글 클라우드의 매출 성장세와 TPU 와 같은 ASIC(Application-Specific Integrated Circuit)의 성장이, 'AI 수익화'를 검증하는 가장 확실한 잣대로서 그 중요성을 가질 수 있을 것으로 판단한다.

핵심 키워드 및 주목해야 할 변수 (Key Factors)

이번 국면 전환을 이해하기 위한 4 가지 핵심 키워드는 ▲AI 버블(수익성 우려) ▲Gemini 3.0(성능 검증) ▲TPU(비용 효율화, ASIC) ▲구글의 수직계열화(AI 산업 수익성 입증)이다.

향후 해당 지표나 기대에 대한 시장 민감도가 높아질 것으로 판단한다. 이에 따라 기존 하드웨어(Nvidia) 실적 외에, 다음 4 가지 대체 지표가 시장의 방향성을 가늠하는 주요 지표로 작용할 가능성이 높다.

① AI 밸류체인 신용 위험 지표 (Credit Risk):

지표: 오라클, 코어위브 CDS

② 구글 클라우드(GCP) 매출 및 이익률 (Service Demand):

지표: 구글 클라우드 부문의 분기별 매출 성장률(YoY) 및 영업이익률(OPM).

③ 브로드컴(Broadcom) AI 관련 매출 (ASIC Proxy):

지표: 브로드컴 내 AI 사업부 매출 추이. 12 월 10 일(현지시간) 실적 발표 예정

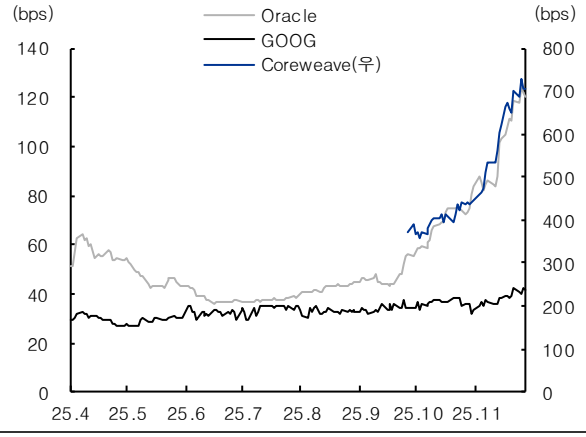
④ 파운드리 및 메모리 실적 (Infra Fundamentals):

지표: TSMC 의 HPC(고성능 컴퓨팅) 매출, 이익 변화, 삼성전자/SK 하이닉스 실적

IBK기업은행 금융그룹 IBK투자증권

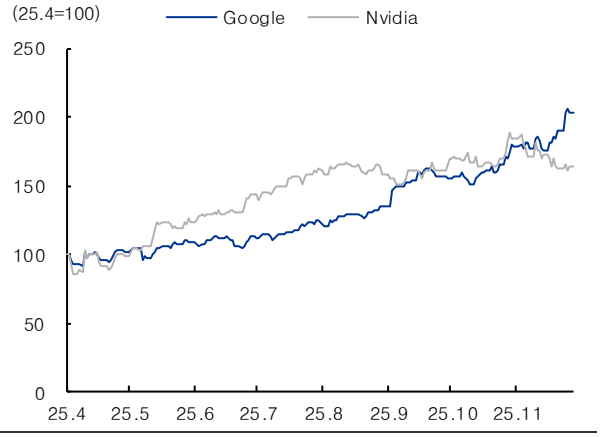
본 조사분석자료는 당사 리서치본부에서 신뢰할 만한 자료 및 정보를 바탕으로 작성한 것이나 당사는 그 정확성이나 완전성을 보장할 수 없으며, 과거의 자료를 기초로 한 투자참고 자료로서 향후 주가 움직임은 과거의 패턴과 다를 수 있습니다. 고객께서는 자신의 판단과 책임 하에 종목 선택이나 투자시기에 대해 최종 결정하시기 바라며, 본 자료는 어떠한 경우에도 고객의 증권 투자 결과에 대한 법적 책임소재의 증빙 자료로 사용될 수 없습니다.

그림 1. 구글, 오라클, 코어위브 CDS 추이



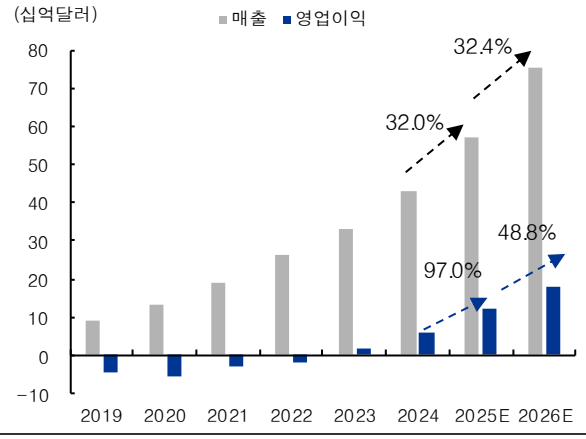
자료: Bloomberg, IBK 투자증권

그림 2. 구글, NVIDIA 주가 추이



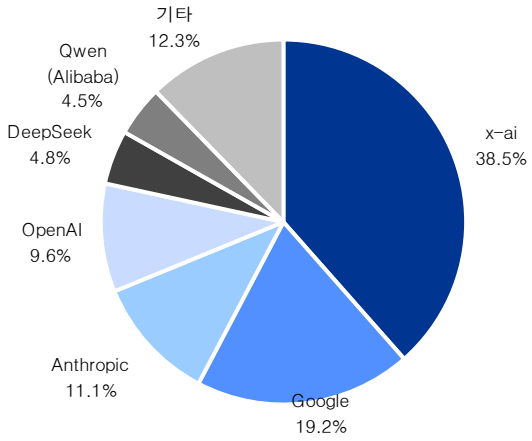
자료: Bloomberg, IBK 투자증권

그림 3. 구글 클라우드 부문 매출, 영업이익 추이



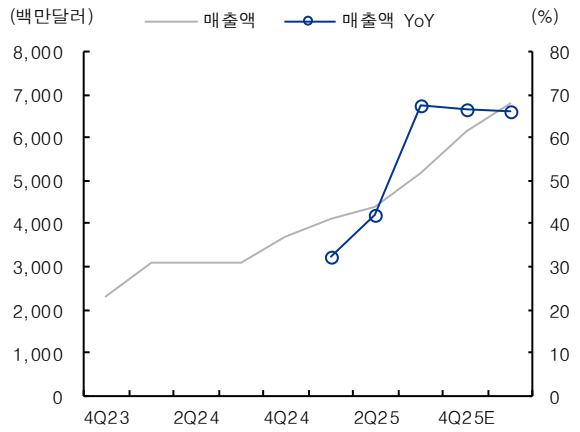
자료: Bloomberg, IBK 투자증권

그림 4. LLM API 누적 토큰 기준 사용량 비중



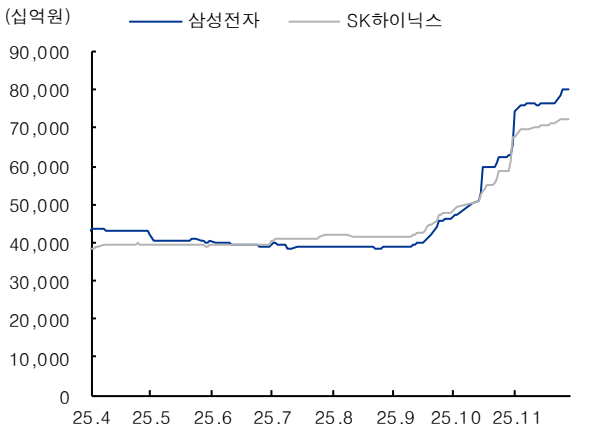
자료:OpenRouter, IBK 투자증권
주:Grok-4.1 Fast(x-ai 의 LLM)은 최근 무료 정책을 사용하며 API 토큰 사용량 급증

그림 5. 브로드컴 AI 관련 매출(ASIC, 데이터센터 네트워크 장비)



자료: Bloomberg, IBK 투자증권

그림 6. 삼성전자, SK 하이닉스 '26 년 실적 추정치 추이



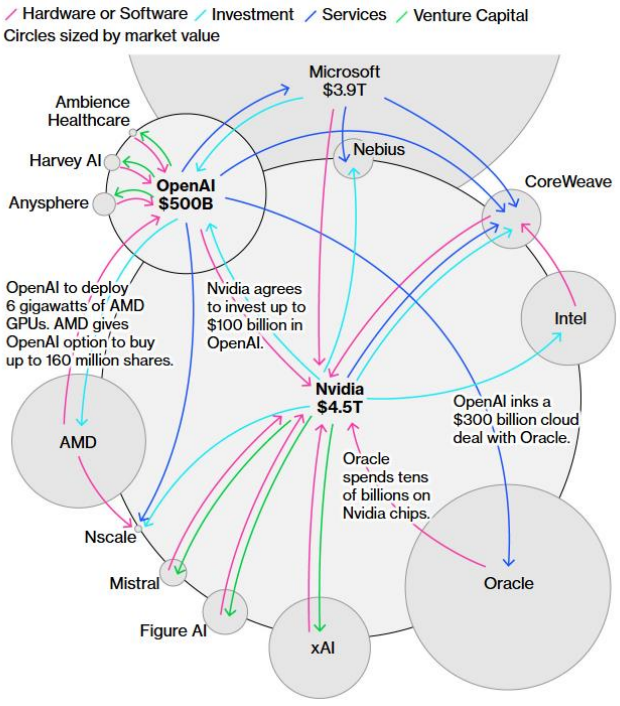
자료: Quantiwise, IBK 투자증권

AI 버블에 대한 우려는 현재 진행형

현재 시장의 불안은 AI 관련 업체들에 대한 주가 조정에 대한 우려를 넘어, 엔비디아–오픈AI–하이퍼스케일러–전용 인프라 사업자로 이어지는 자본의 순환 구조 자체에 대한 경계로 확장되고 있다. AI 밸류체인 내부에서는 엔비디아가 칩을 공급하고, 오픈AI와 코어위브(CoreWeave) 같은 인프라 사업자가 이를 기반으로 데이터센터를 구축하며, 다시 빅테크와 클라우드 고객이 장기 공급 계약·지분 투자·선급금의 형태로 자금을 되돌려 주는 구조가 반복되고 있다. 문제는 이 과정에서 집행되는 설비투자(CAPEX) 규모가 과거 IT 사이클과 비교해도 이례적으로 크다는 점이다.

주요 빅테크 4개사(아마존·메타·구글·마이크로소프트)의 향후 2년간 AI 관련 CAPEX는 7,500억달러 이상으로 추정되며, 글로벌 데이터센터 인프라 투자는 2024년 기준 약 2,900억달러에서 2030년 1조달러 수준으로의 확대가 예상된다. 반면, 매출은 아직 그만큼 따라오지 못하고 있어, 일부 기업의 매출 대비 CAPEX 비율과 영업현금흐름 대비 CAPEX 비율이 빠르게 상승하고 있다. 이 점이 "AI 인프라의 ROIC를 회수하지 못할 수 있다"는 우려의 핵심이다.

그림 7. AI 관련 기업 자금 순환 구조



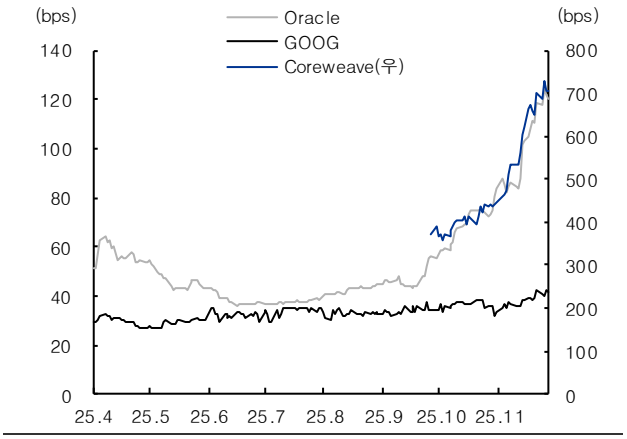
자료: Bloomberg

이러한 구조적 불안은 크레딧 시장에 직접적으로 반영되고 있다. 대표적인 사례가 오라클(Oracle)의 CDS다. 2025년 9~11월 사이 오라클 5년물 CDS 스프레드는 약 50bp에서 120bp 수준으로 두 배 이상 급등하며, 동기간 메타·알파벳 대비 훨씬 가파른 확대로 기록됐다. 최근 크레딧 리서치에서는 오라클의 AI 데이터센터 투자가 향후 2~3년간 배당·자사주 매입을 제외한 잉여현금흐름의 90% 이상을 잠식할 것으로 추정하면서, 채무 증가와 장기 리스 의무, 데이터센터 임대 계약 등을 이유로 AI 하이퍼스케일러 중 가장 약한 고리로 평가하고 있다.

코어위브(CoreWeave)는 엔비디아 GPU를 기반으로 한 AI 특화 클라우드·인프라 서비스 제공 업체로, 주로 대규모 LLM 학습·추론과 고성능 컴퓨팅(HPC) 워크로드를 위한 GPU 인프라를 임대·운영하는 비즈니스를 한다. 매출과 수주잔고는 오픈AI·마이크로소프트 등 소수 고객에 집중되어 있고, 데이터센터 구축·전력 계약·GPU 조달에 필요한 자금 대부분을 회사채·프라이빗 크레딧·담보대출 등 외부 차입에 크게 의존하고 있다는 점에서 자금 순환 고리의 취약한 지점으로 인식된다. 이와 같은 높은 레버리지 구조와 집중된 수요 의존도를 반영하듯, 코어위브의 CDS 스프레드는 상장 이후 단기간에 하이일드 구간으로 급격히 확대되며, 시장이 해당 기업의 현금창출력과 부채 상환능력에 대해 높은 위험프리미엄을 요구하고 있음을 보여준다.

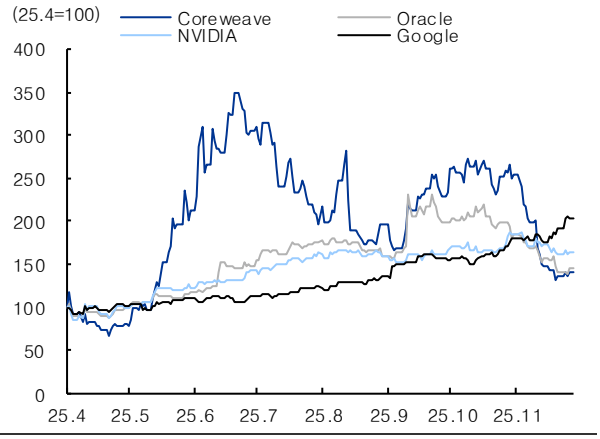
오라클과 코어위브의 CDS 스프레드는 각각의 개별 기업 수익성과 재무구조를 반영하는 지표이면서 동시에, AI 인프라 밸류체인 전체에 대해 시장 참여자가 요구하는 자본 건전성·수익성 기준을 집약적으로 보여주는 proxy로 해석할 수 있다. 즉, 두 기업의 CDS가 확대되는 국면은 단순히 특정 기업의 펀더멘털 악화만을 의미하는 것이 아니라, 고CAPEX·고레버리지 구조를 가진 AI 데이터센터 투자 전반에 대해 크레딧 시장이 요구하는 기대수익률(위험프리미엄)이 빠르게 상향 조정되고 있음을 시사한다. 엔비디아 실적과 빅테크의 AI 매출·CAPEX와 더불어, 오라클·코어위브의 CDS 추이를 AI 인프라 레버리지에 대한 시장의 허용 한계와 버블 논쟁의 온도를 측정하는 지표로 상시 모니터링할 필요가 있다.

그림 8. 주요 기업 CDS 추이



자료: Bloomberg, IBK 투자증권

그림 9. 주요기업 주가 추이



자료: Bloomberg, IBK 투자증권

플래그십 모델 우위와 AI 투자 정당성: Gemini 3.0 Pro 의 시사점

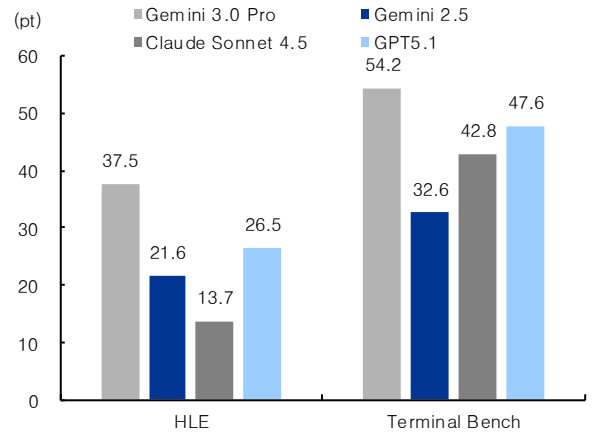
Gemini(구글 LLM)의 성능 확장은 단순한 모델 성능 개선을 넘어, AI 기술 고도화가 실제 수익성과 연결될 수 있는 가능성을 보여주는 사례다. 특히 Gemini 3.0 Pro를 중심으로 한 경쟁력 강화는 Google Cloud 내에서 모델 API 사용량과 B2B 활용도를 확대시키면서, 대규모 AI·데이터센터 CAPEX가 장기적으로 회수 가능한 투자라는 점에 일정 부분 근거를 제공한다. 이는 구글 개별 기업의 우위 확인에 그치지 않고, AI 인프라 투자가 구조적으로 ROIC를 확보하기 어렵다는 "AI 버블" 논쟁에 대해 일정 부분 완충 역할을 할 수 있는 요인으로도 해석할 수 있다.

플래그십 모델 우위: Gemini 3.0 Pro

Gemini 3.0 Pro는 구글이 제시한 공식 자료와 외부 평가에서 모두 기존 Gemini 2.5 대비 추론, 코딩, 멀티모달, 장문 컨텍스트 등 주요 영역에서 성능이 크게 개선된 것으로 소개되고 있다. 이전 세대 모델들이 특정 벤치마크에서 국지적 우위를 보이는 수준이었다면, Gemini 3.0 Pro는 다양한 범주의 벤치마크에서 일관되게 상위권을 기록하며, 플래그십 LLM 경쟁 구도에서 "최상단 레벨에 진입했다"는 인식을 형성한 것이 특징이다.

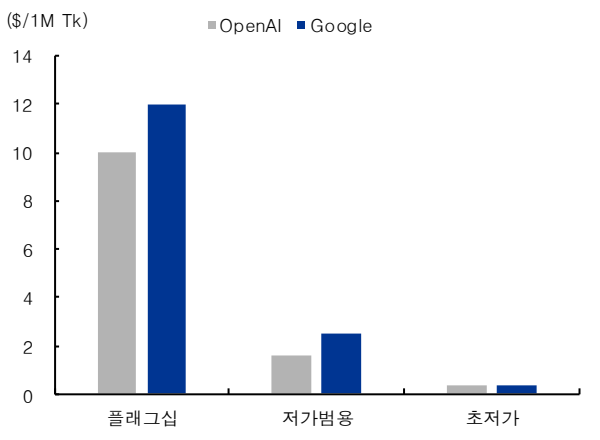
지금까지 일반 소비자 측면에서는 ChatGPT, 기업·개발자 측면에서는 GPT 기반 코파일럿, Claude 기반 분석·코딩 도구가 사실상의 표준 선택지로 자리잡고 있었다. 구글 Gemini는 검색·안드로이드·유튜브와의 결합 가능성에도 불구하고, 체감 품질과 도구 에코시스템 측면에서 후발 주자 이미지가 강했다. Gemini 3.0 Pro는 이러한 구도를 바꿀 "전환점 후보"라는 점에서 중요하다.

그림 10. 주요 플래그십 모델 성능 비교



자료: Google, IBK 투자증권
주 1) HLE는 Humaniti Last Exam 으로 고급 추론과 문제 해결 능력 테스트
주 2) Terminal Bench는 Code Agent 에서 프로그래밍 능력을 테스트 하는 지표

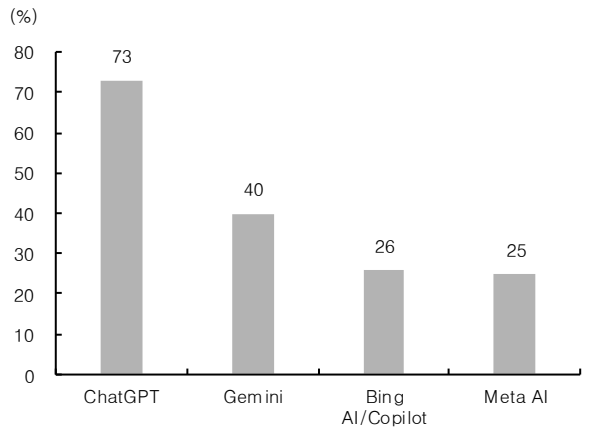
그림 11. 주요 LLM 모델 API 가격 비교



자료: OpenAI, Google, IBK 투자증권
주: 출력 토큰 기준 가격

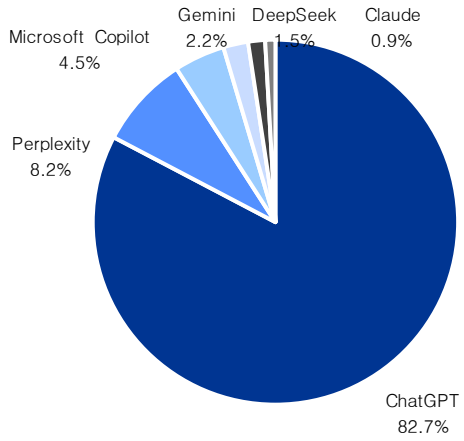
투자 관점에서 중요한 것은 이러한 성능 우위가 단일 API 상품에 국한되지 않고, 구글의 전체 서비스 포트폴리오에 공통으로 적용되는 "코어 자산"이라는 점이다. Gemini 3.0 Pro는 외부 개발자와 기업이 사용하는 모델 API이자, 엔터프라이즈용 AI 플랫폼(Vertex AI)과 구글 자체 서비스(검색, 안드로이드, 워크스페이스 등)에 공통으로 들어가는 인공지능 엔진 역할을 수행한다. 따라서 플래그십 모델 경쟁에서의 우위는 단순한 기술적 리더십을 넘어, 향후 수년간 다양한 상품·서비스 레이어에서 수익성을 뒷받침하는 기반 인프라로 기능할 수 있다.

그림 12. 일반 소비자의 LLM 사용 경험 비율(B2C)



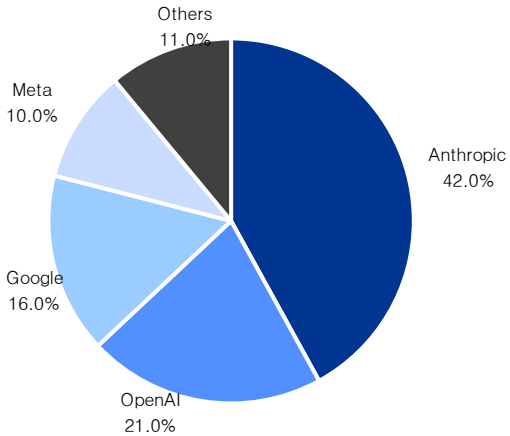
자료: S&P Global, IBK 투자증권
주: 25 년 10 월 까지의 설문 결과

그림 13. 웹사이트 방문 트래픽 기준 점유율(B2C)



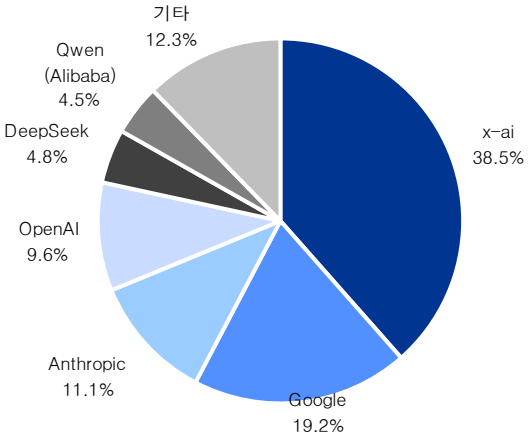
자료: Statcount, IBK 투자증권
주: 25 년 7 월 기준 트래픽

그림 14. 기업 대상 LLM 시장 점유율(B2B, 미국 기업 대상)



자료: MenloVentures, IBK 투자증권
주: 25 년 10 월 까지의 설문 결과

그림 15. LLM API 누적 토큰 기준 비중(API 토큰 기준)



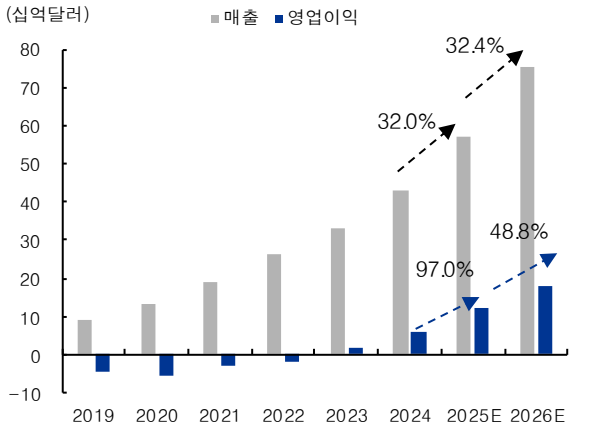
자료: OpenRouter, IBK 투자증권
주:Grok-4.1 Fast(x-ai 의 LLM)은 최근 무료 정책을 사용하며 API 토큰 사용량 급증

Google Cloud 구조: 모델 성능이 매출·마진으로 이어지는 경로

Google Cloud의 수익 구조는 크게 인프라·플랫폼(IaaS/PaaS), AI 인프라 및 모델 API(Vertex AI, Gemini API), 그리고 워크스페이스·애플리케이션 레이어로 나눌 수 있다. 최근 실적 발표에 따르면 Google Cloud 매출은 전년 대비 30%대 중반 성장률을 기록하며 그룹 내에서 가장 빠른 성장 축으로 자리잡았고, 영업이익률 역시 20%를 상회하는 수준까지 개선되고 있다. 경영진과 외부 분석들은 이러한 성장과 마진 개선의 핵심 요인으로 GCP 내 AI 인프라 및 Gemini 기반 생성형 AI 솔루션 수요를 반복적으로 언급하고 있다. 기존에는 이 AI 수요가 상대적으로 "보수적인 수준의 모델 품질을 가진 Gemini 1.x·2.x를 바탕으로, 구글 클라우드·워크스페이스 고객을 중심으로 한 제한된 성장"에 가까웠다면, Gemini 3.0 이후에는 GPT·Claude를 주로 사용하던 신규 워크로드 일부까지 흡수할 수 있는 가능성이 열렸다는 점이 다르다.

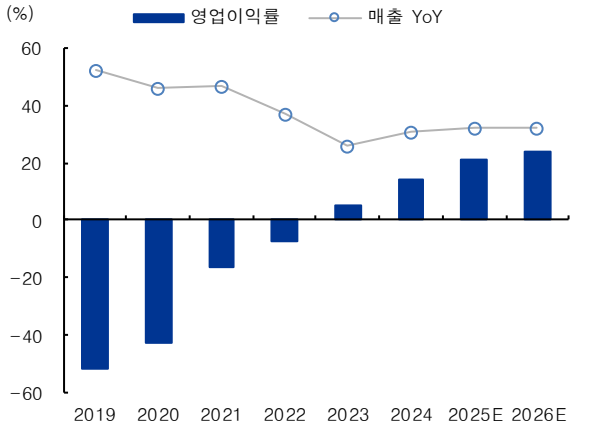
Gemini 3.0 Pro는 이 Cloud 구조 속에서 두 축으로 수익화된다. 첫째, Vertex AI 및 Gemini API를 통해 텍스트 생성, 코드 보조, 에이전트, 검색·요약 등 다양한 워크로드를 처리하는 모델로 활용되면서, 토큰·쿼리 기반 과금 구조를 통해 클라우드 매출로 직접 연결된다. 둘째, 워크스페이스·협업 도구·컨택센터 솔루션 등 엔터프라이즈 제품군 위에 "Gemini 옵션"이 얹히면서, 좌석당 단가를 높이는 업셀(상위 요금제 전환) 구조로 작동한다. 플래그십 모델의 품질이 높아질수록, 고객 입장에서는 추가 비용을 지불할 유인이 커지고, 이는 사용량 증가와 고객당 매출 확대에 동시에 기여하는 형태다. 특히 코딩·에이전트 관련 워크로드에서 성능 열위가 해소될 경우, 지금까지 Azure(OpenAI), AWS(Anthropic)로 배분되던 기업들의 투자금 중 일부가 Google Cloud로 이동할 가능성이 열리면서, 향후 1~2년간 클라우드 성장률과 AI 관련 매출 구성비에 추가적인 상향 압력을 줄 수 있다.

그림 16. 구글 클라우드 부문 매출, 영업이익 추이



자료: Bloomberg, IBK 투자증권

그림 17. 구글 클라우드 부문 영업이익률과 매출 성장률 추이



자료: Bloomberg, IBK 투자증권

또한 Google Cloud는 이미 일정 수준의 매출과 영업이익을 기록한 상태에서 자체 LLM(Gemini)과 인프라(TPU)를 결합하는 수직 계열화 구조를 구축하고 있다는 점에서, "AI 투자→클라우드 매출·이익"이 실제로 작동하는지 판단할 수 있는 대표적인 관측 지표가 될 가능성이 높다. 동일한 AI·데이터센터 CAPEX라 하더라도, 기존 클라우드 사업의 이익 기반 위에서 모델·인프라를 수직 통합해 추가 수요를 흡수하는 구조와, 별도의 사업부가 레버리지에 의존해 설비를 확대하는 구조는 ROIC와 변동성 측면에서 다른 결과를 낳을 수밖에 없다. 향후에는 개별 기업의 성공 사례라기보다는, "수직 계열화된 AI 클라우드 모델이 어느 정도의 수익성을 안정적으로 입증하는지"가 AI 산업 전반에 대한 기대와 버블 논쟁을 가늠하는 핵심 지표 중 하나로 시장에서 활용될 가능성이 크다.

TPU: 높은 CAPEX 속 효율성 우위. 효율성을 갖춘 ASIC에 대한 관심도 확대

구글의 TPU(Tensor Processing Unit)는 행렬·텐서 연산에 특화된 전용 ASIC으로, 범용 GPU 대비 유연성·범용성은 떨어지지만, LLM 학습·추론처럼 워크로드가 비교적 명확한 영역에서는 성능/전력/비용 측면에서 유리한 구성을 만들 수 있는 구조다. 구글은 여러 세대의 TPU를 거치며 자체 칩과 모델을 함께 최적화해 왔고, 최근 세대(v6·v7)는 대규모 LLM·추천 모델을 전제로 설계되었다고 밝히고 있다. 결과적으로, 일정 규모 이상의 트래픽에서는 범용 GPU만 사용할 때보다 단위 컴퓨트 비용을 낮출 여지가 생기고, 플래그십 모델 가격과 저가형(Flash 계열) 가격 모두를 비교적 공격적으로 가져갈 수 있는 기반이 된다.

다만 이런 구조가 "절대적인 우위"를 의미하는 것은 아니다. TPU는 구글 내부 워크로드와 Google Cloud 생태계에 최적화되어 있는 만큼, 다른 하이퍼스케일러나 범용 고객이 그대로 활용하기에는 제약이 크다. 또한, CAPEX 규모 자체가 매우 크기 때문에, 일정 수준 이상의 수요·가동률이 전제되지 않으면 오히려 부담 요인으로 작용할 수 있다. 그럼에도 불구하고, 구글처럼 특정 워크로드에 맞춰 ASIC을 설계·내재화하려는 움직임은 "범용 GPU만으로는 AI 인프라 비용 구조를 방어하기 어렵다"는 인식이 확산되고 있음을 시사한다.

이런 맥락에서 보면, 앞으로 AI 인프라 투자의 초점은 단순 GPU 증설을 넘어, 각 하이퍼스케일러와 대형 서비스 사업자가 어떤 식으로 자체 ASIC 또는 맞춤형 칩을 도입해 비용 구조를 재편하는지로 이동할 가능성이 크다.

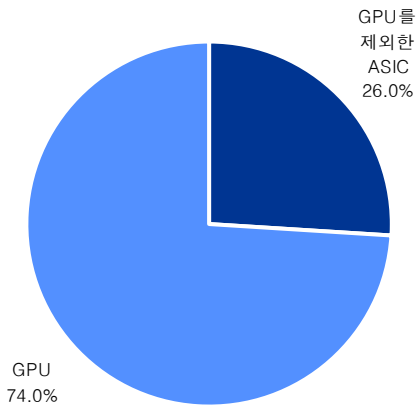
효율적 ASIC 에 대한 본격적인 시장 침투 확대 기대감 유효

AI 데이터센터 가속기 칩 시장에서 GPU 이외의 ASIC 비중은 이미 무시하기 어려운 수준까지 올라와 있다. 현재 기준으로 보면 GPU를 제외한 추론·훈련용 ASIC이 전체 데이터센터 가속기 칩 시장의 약 4분의 1 수준(약 20~30%대)을 차지하고 있고, 추론용 칩만 놓고 보면 ASIC 비중이 2025년 24%에서 2030년에는 80%까지 확대될 수 있다는 전망도 제시되고 있다. 단기적으로는 여전히 GPU가 CAPEX의 중심을 차지하지만, 중장기적으로는 "학습은 GPU + 일부 TPU(시장 점유율 현재 3% 정도로 추정), 대규모 반복 추론은 워크로드 특화 ASIC 중심"이라는 이원화 구조로 갈 것이라는 기대가 이미 시장에 상당 부분 반영되고 있는 셈이다.

이러한 변화의 배경에는 추론 워크로드의 성격이 있다. LLM·검색·추천·광고 입찰 등 대규모 온라인 서비스에서의 추론은, 모델 구조가 어느 정도 고정된 상태에서 동일하거나 유사한 연산을 매우 많이 반복 수행하는 구조다. 이 영역에서는 범용성이 높은 GPU보다, 연산 구조와 데이터 이동 패턴을 전제로 설계된 ASIC이 전력 밀도·지연 시간 측면에서 유리할 수 있다. 다시 말해, 모바일·PC에서 NPU(Neural Processing Unit)를 도입해 전력 효율을 높이려는 시도와 유사한 논리가 데이터센터 레벨에서 벌어지고 있고, 그 결과로 "효율형 ASIC"이 전체 가속기 풀에서 차지하는 비중이 점차 커지고 있다는 흐름이다.

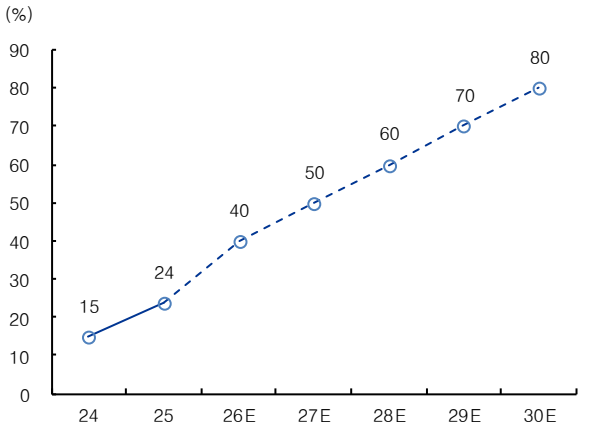
이런 관점에서 보면, 앞으로 AI 인프라 관련 지표를 해석할 때 단순히 GPU 매출과 이익에 대한 실적뿐만 아니라 이외의 ASIC에 대한 실적을 통해서 확인 할 수 있을 것으로 판단한다. 상장회사 기준으로 "비-GPU AI 칩의 확산 속도"를 추적할 수 있는 브로드컴의 AI 관련 매출, 그중에서도 XPU의 매출 비중을 대표적인 대용치 로 제시한다.

그림 18. 데이터 센터용 가속기 시장에서 GPU 와 그 외 칩 점유율



자료: MordorIntelligence, IBK 투자증권

그림 19. AI 추론 시장에서 ASIC 시장 점유율 추정



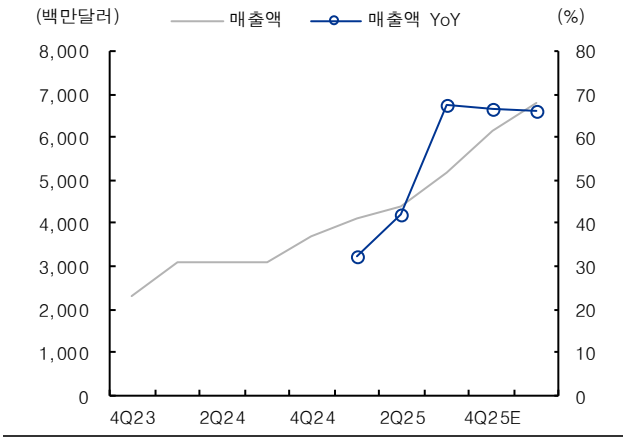
자료:InfinitiX, IBK 투자증권

브로드컴은 전통적으로 스위치·라우터용 네트워크 칩과 스토리지·통신용 반도체에 강점을 가지고 있었지만, 최근 몇 년 사이에는 대형 클라우드 고객을 대상으로 한 AI 관련 매출이 성장의 주된 축으로 부상하고 있다. 회사가 분리해서 제시하는 "AI 반도체 매출"에는 (1) 데이터센터 내부에서 GPU·ASIC을 연결하는 고속 네트워크 칩과 (2) 특정 하이퍼스케일러용으로 설계된 커스텀 XPU/ASIC(훈련·추론용 가속기)이 함께 포함된다. 최근 실적 기준으로 이 AI 관련 매출은 전년 동기 대비 두 자릿수 후반~세 자릿수에 가까운 성장률을 반복적으로 기록하고 있고, 반도체 부문 전체에서 차지하는 비중도 빠르게 높아지고 있다.

특히 눈여겨볼 부분은, 브로드컴 경영진이 AI 매출 중 상당 부분이 "커스텀 XPU 및 AI 가속기"에서 발생하고 있다고 밝히고 있다는 점이다. 이는 대형 고객들이 범용 GPU만 추가로 도입하는 것이 아니라, 자체 워크로드에 맞춘 맞춤형 칩을 병행 도입하고 있음을 의미한다. 실적 발표 및 가이던스에서는 향후 몇 년간 특정 소수 하이퍼스케일러를 대상으로 하는 커스텀 칩 수주가 장기 계약 형태로 이어지고 있고, 이로 인해 AI 관련 매출이 연간 기준 수십억 달러 단위로 추가 확대될 수 있다는 코멘트가 반복해서 등장하고 있다. 매출 절대 규모뿐 아니라, "AI 매출 중 XPU 비중이 얼마나 빨리 올라가는지"가 비-GPU형 가속기의 침투 속도를 가늠할 수 있는 핵심 포인트다.

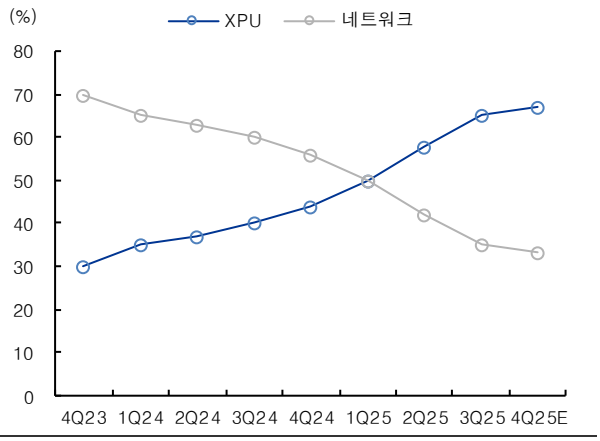
정리하면, 한편에서는 TPU와 같은 내부용 ASIC을 통해 하이퍼스케일러가 자체 비용 구조를 최적화하고, 다른 한편에서는 브로드컴과 같은 외부 공급자를 통해 커스텀 XPU·네트워크 칩이 빠르게 채택되면서, 데이터센터 가속기 시장의 구조는 점차 다변화되고 있다. ASIC 비중이 추론 영역에서 빠르게 50% 이상, 장기적으로는 80% 수준까지 확대될 수 있다는 전망은, 이러한 흐름이 단순한 기술 발전을 넘어 CAPEX와 매출 흐름에 이미 반영되기 시작했다는 신호로 볼 수 있다. 브로드컴의 AI·XPU 관련 매출과 그 비중 변화는, 이 구조적 변화를 확인할 수 있는 상장 데이터 포인트라는 점에서, GPU 중심의 AI 투자 사이클을 평가할 때 함께 모니터링할 필요가 있는 지표로 볼 수 있다.

그림 20. 브로드컴 AI 관련 매출 추이



자료: Bloomberg, IBK 투자증권

그림 21. AI 매출 추이 중 XPU, 네트워킹 비중 추정치 추이



자료:Broadcom, IBK 투자증권
주:분기별 컨콜 스크립트에서 XPU 매출 비중 언급을 통해 추정

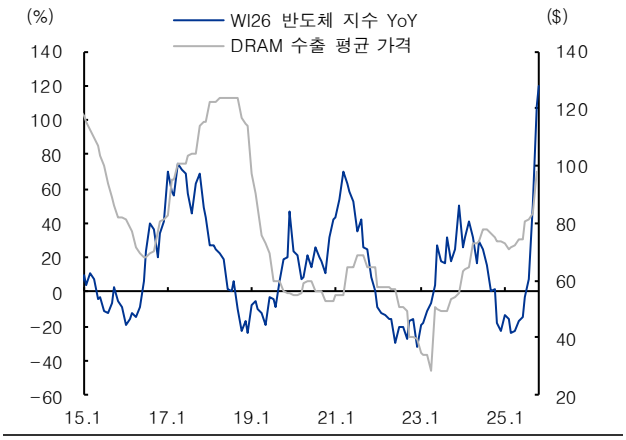
AI 업황 전반을 반영하는 우리나라 대형 반도체 기업들

2025년 7월까지는 시장의 관심이 거의 전적으로 엔비디아향 HBM에 쏠려 있었다. 이후 8월부터는 HBM 증설·믹스 변화가 레거시 DRAM까지 파급되면서, 일반 서버·PC·모바일용 메모리 가격까지 동반 상승하는 '풍선 효과'가 나타나고 있다. 그 결과 국내 대형 메모리 2개사에 대한 2026년 실적 추정치는 HBM 매출 증가와 전통 DRAM 가격 반등을 동시에 반영하며 빠르게 상향 조정되는 구간에 진입했다. 국내 2개 반도체 업체는 AI 서버뿐 아니라 비(非)AI IT 수요까지 함께 담는다는 점에서, 개별 GPU/ASIC 업체보다 'AI + 글로벌 IT 수요'가 중첩된 업황을 보여주는 지표로 보는 것이 적절하다.

현재 주가에는 상반된 두 가지 기대와 우려가 동시에 녹아 있다. 한쪽에서는 AI 버블 논란이 확대될 경우 빅테크 CAPEX가 조정되고, 엔비디아·오픈AI 중심의 설비투자 속도가 둔화될 것이라는 리스크가 반영되어 있다. 다른 한쪽에서는 TPU를 포함한 ASIC 기반 학습·추론 규모가 확대되면서, GPU 의존도가 일부 줄어들더라도 고대역폭 HBM과 서버 DRAM에 대한 구조적 수요는 유지·확대될 수 있다는 기대가 살아 있다. 요약하면, CAPEX 규모 축소 우려와 '대체 칩 확산을 통한 구조적 성장'이라는 두 축이 동시에 국내 메모리 실적 전망에 투영되고 있는 국면이다.

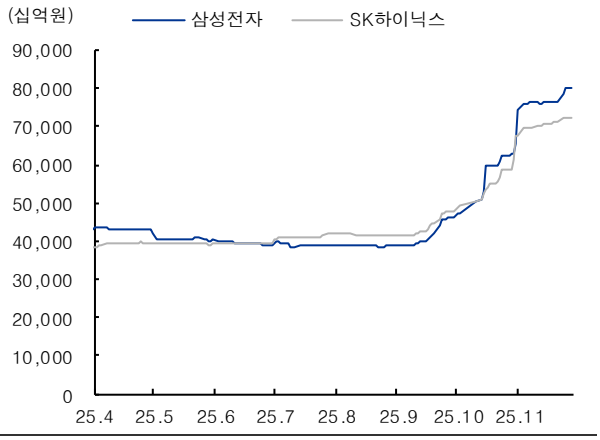
이러한 맥락에서 국내 대형 메모리 2개사의 2026년 이후 실적 추정치와 제품 믹스 변화는, 이번 보고서에서 다룬 여러 지표 가운데 가장 '종합적인 업황 지표'에 가깝다. 엔비디아 실적과 빅테크 CAPEX가 AI 설비투자의 절대 규모를 보여준다면, 브로드컴의 AI 매출은 비(非)GPU ASIC 확산 속도를, 오라클·코어위브 CDS는 레버리지 한계를 드러낸다. 여기에 삼성전자·SK하이닉스의 이익 전망과 DRAM 가격·수출 단가 추세를 함께 보면, TPU·ASIC 확산과 AI CAPEX 조정 가능성, 전통 IT 수요 회복이 최종적으로 메모리 업황과 이익 레벨에 어떻게 정리되고 있는지를 한 번에 확인할 수 있다. 이 의미에서 한국 대형 메모리 2개사는 '엔비디아 실적만큼 중요해질 지표들'을 보완하는 마지막 체크포인트라 할 수 있다.

그림 22. 한국 수출 평균 DRAM 가격과 반도체 지수



자료: Bloomberg, IBK 투자증권

그림 23. 삼성전자, SK 하이닉스 '26 년 실적 추정치 추이



자료: Blooomberg, IBK 투자증권