

# IT 장비/소재

## 경험의 확장 (feat. AI)

SK증권 리서치센터



Analyst  
이동주

natelee@sks.co.kr  
3773-9026

### 생성형 AI의 하방 전개

생성형 AI 인 Chat GPT 의 등장으로 방대한 양의 학습과 추론을 위한 HBM 과 패키징 시장이 급속도로 커졌다. 또한 Chat GPT 는 AI 와 사용자 경험 간의 직접적인 연결고리를 확인할 수 중요한 계기가 되었다. 앞으로 전개되는 AI 기술 개발의 방향은 바로 '경험의 확장'에 있다. 경험의 확장은 모바일, TV, 가전, 모빌리티, 로봇 등 엣지 디바이스를 중심으로 연결이 강화되고 생성형 AI 접목이 늘어남을 시사한다.

### 온디바이스 AI, 관점의 전환

온디바이스 AI 가 새로운 테크 트렌드임을 부인할 수 없다. 서비스 부재, 정확성에 대한 의구심, 디바이스 BoM cost 부담 등이 약점으로 지적되지만 저지연·개인화·보안 측면에서는 유리하다. 무거운 작업은 클라우드, 가볍고 빠른 작업은 온디바이스로 처리하는 하이브리드 AI 방식으로 진화할 것이다. 또한, 테크 트렌드는 소비자의 요구보다는 Tech 업체가 부가가치 창출을 위해 주도적으로 이끌어 가는 영역임을 기억할 필요가 있다. LLM 모델의 경량화 이후 NPU 기능을 강화한 칩셋 공개가 한창이다. 2024 년부터 모바일과 PC 는 인공지능 기능을 강화한 자체 SoC 기반의 신제품이 대거 출시될 전망이다.

### 소부장 투자 전략

- 온디바이스 AI: 부품 단에 직접적인 영향은 NPU 기능 고도화에 있다. AP 및 CPU/GPU 업체로 수혜가 집중된다. Qualcomm 이 최대 수혜 업체이다. 국내 소부장에서 일부 낙수 효과를 기대할 수 있는 영역은 소켓, 기판, IP 쪽이다. 그 중에서 리노공업, 심텍, 티엘비, 쉐리타스반도체를 주목한다.
- 클라우드 AI: 중국 규제, 대체재 성격의 온디바이스 개화 등センチ먼트 이슈로 주가는 단기 조정을 거쳤다. 그러나 대용량 데이터의 연산에는 강력한 클라우드 시스템은 필수다. 엣지와 클라우드의 역할은 확실히 구분된다. 생성형 AI 서버 시장은 엣지 대비 10 배 이상 크다. 주요 서버 업체 CAPEX 기초 변화에도 AI 투자는 줄지 않는다. 클라우드 관련 소켓, 기판, HBM 관련 업체를 주목한다. ISC, 티에프이, 삼성전기, 이오테크닉스, 피에스케이홀딩스를 선호한다.

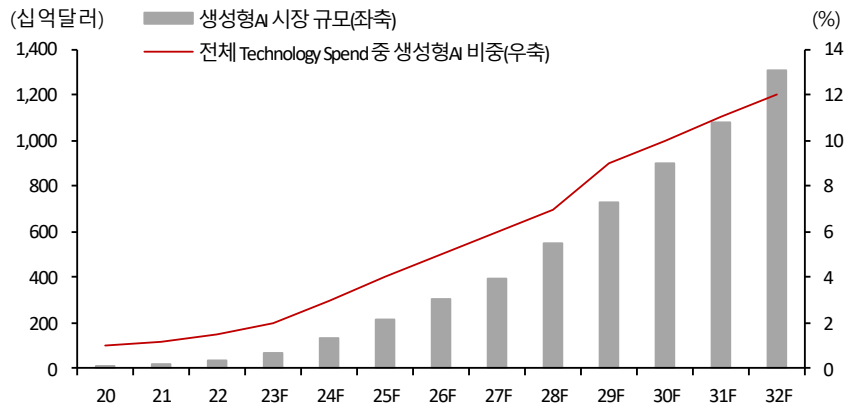
## 1. 생성형 AI의 하방 전개

### (1) Chat GPT가 불러온 바람

2022년 11월, Open AI사의 Chat GPT의 출시는 반도체 업계에 큰 지각 변동을 일으켰다. GPT는 딥러닝 기술을 기반으로 인간과 유사한 텍스트를 생성하는 자연어 기술이다. 생성형 AI로 대표되는 GPT4는 대규모 언어 모델(LLM)로 인간과 대화가 가능한 수준에 이른다. 다만 고도화된 학습과 추론을 위해 병렬 연산에 유리한 GPU가 쓰이게 되는데 방대한 양의 데이터를 한 번에 처리할 수 있는 HBM 수요도 함께 폭발적으로 증가하고 있다.

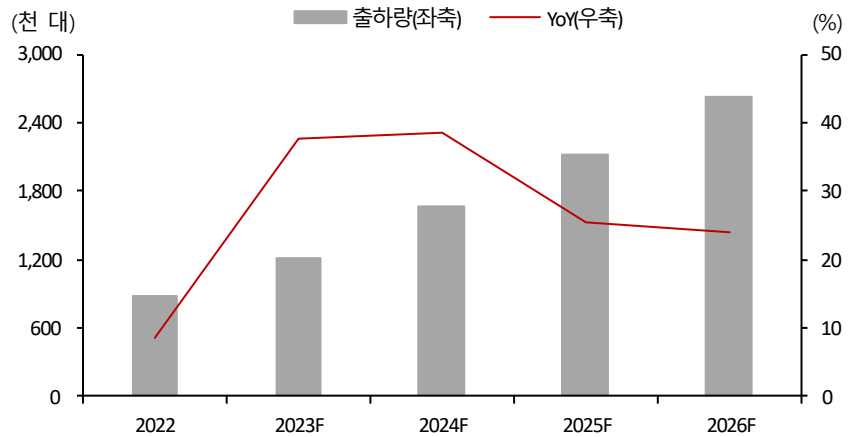
인프라, 디바이스, 서비스 등을 포함한 생성형 AI 시장 규모는 2022년 400억달러에서 2032년 1조 3,040억달러 수준으로 가파른 성장이 예상된다. 클라우드 기반의 AI Server 출하량도 2022년 88만대에서 2026년 263만대 수준까지 급성장할 것으로 기대된다.

생성형 AI 시장 규모 전망



자료: 블룸버그 인텔리전스, SK 증권

Global AI Server shipments



자료: Trendforce, SK 증권

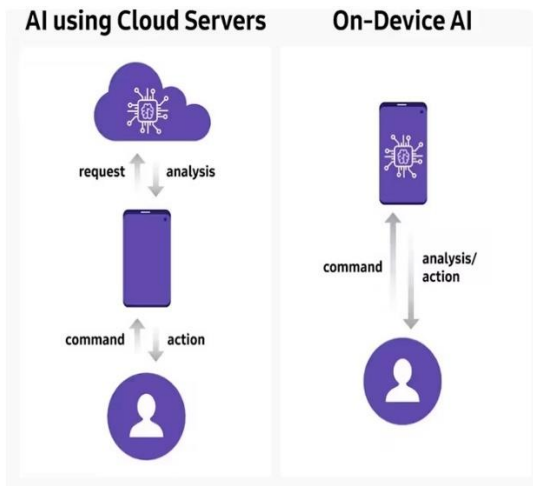
## 1. 생성형 AI의 하방 전개

### (1) Chat GPT가 불러온 바람

온디바이스 AI가 시장의 뜨거운 감자다. 온디바이스 AI는 클라우드 시스템을 통한 연결 없이도 기기 자체적으로 AI 연산이 가능한 시스템이다.

이를 가능케 한 배경에는 LLM 모델의 발전이 있다. 오픈 AI의 GPT-4, 구글의 'PaLM2' 등 현재 클라우드 시스템을 통한 Chat GPT 영역에서는 최소 1,000억 개에서 많게는 1조개의 파라미터가 쓰이면서 모델 자체가 굉장히 무겁다. GPU, 다수의 HBM, 대면적 기판 등까지 고려하면 물리적으로, 가격적으로 엣지 단에 적용은 불가능하다. 그러나 메타에서 경량화 모델인 'LLaMa2'를 오픈소스로 공개하면서 엣지 디바이스 단에도 생성형 AI를 위한 길이 열렸다. LLaMa2는 70억개, 130억개, 700억개의 파라미터를 가지는 3가지 모델로 출시됐는데 70억개는 엣지 단에 적용 가능한 경량화 모델이다. sLLM으로 불리는 이런 경량화 모델은 파라미터가 보통 60-100억개 사이다. 이와 함께 NPU 성능을 강화한 디바이스용 칩셋들도 연이어 출시되면서 AI 디바이스 개화에 힘을 실어주고 있다. Qualcomm의 스냅드래곤 8세대 3과 Mediatek의 Dimensity 9300이 대표적이며, 70-100억개 파라미터까지 지원한다. 최초 생성형 AI 폰으로 꼽히는 구글의 Pixel 8 Pro는 자체 AP인 Tensor G3가, 최근 출시된 비보의 X100 시리즈에서 Dimensity 9300이 처음 탑재되었다.

Cloud AI와 온디바이스 AI 시스템 비교



자료: 삼성전자

Qualcomm, on device AI 챗봇 서비스



자료: 언론보도, Qualcomm

Qualcomm, on device AI 이미지 생성 서비스



자료: CES2024, SK 증권

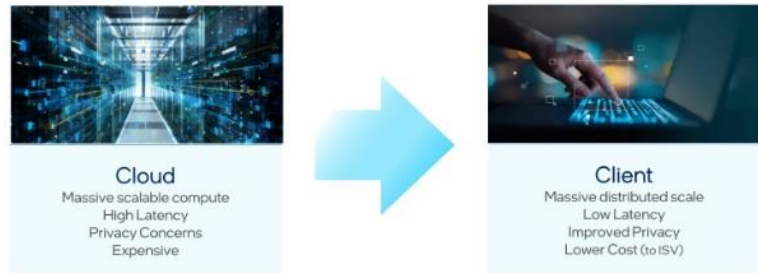
삼성전자, on device AI 실시간 번역 서비스



자료: CES2024, SK 증권

Intel, AI workloads 비교

AI Workloads: Cloud and Client



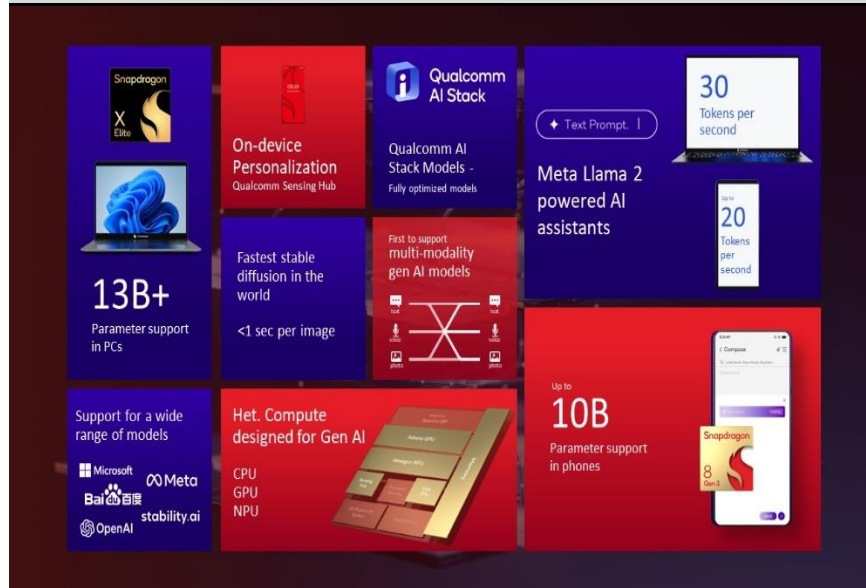
자료: Intel

LLM Model 비교

	OpenAI 'GPT-4'	Google 'Palm2'	Meta 'Llama2'	Google 'Gemini Ultra'
출시	2023.03	2023.05	2023.07	2023. 12
매개변수	비공개 (500억~1조 개 추정)	3,400억개	70억개 130억개 700억개	비공개 (약 1조개 추정)
MMLU 벤치마크 점수	86.4%	78.3%	68.9%	90.0%

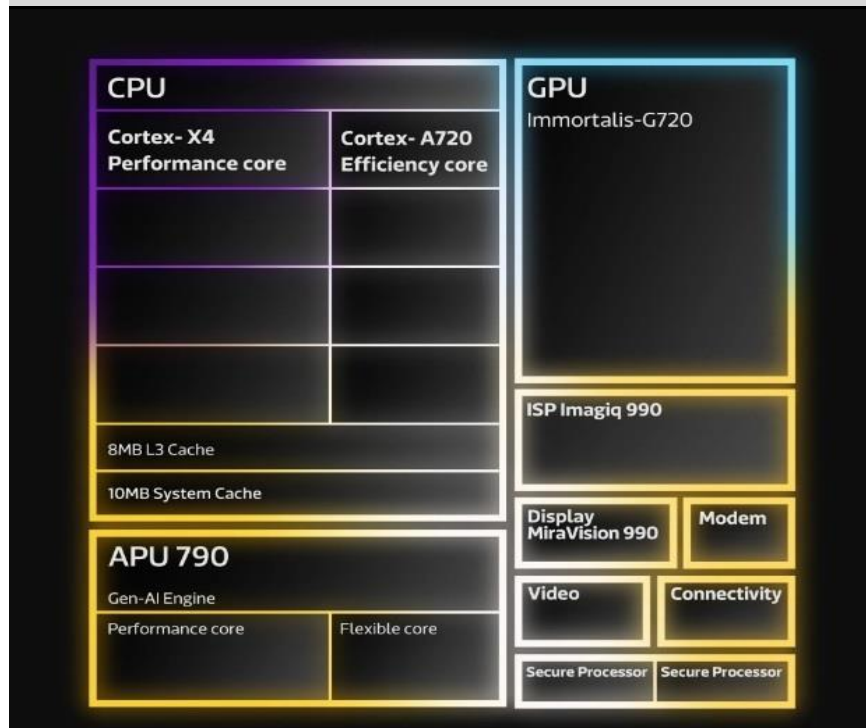
자료: 각 사, SK증권

Qualcomm Summit AI 관련 전략



자료: Qualcomm

Mddiak 신형 AI 칩 'Dimensity 9300'



자료: Mediatek

모바일을 넘어 랩탑에도 생성형 AI 칩셋 적용이 확대될 것으로 보인다. 인텔 코어 울트라(메테오레이크)는 인텔 최신 기술의 집약체이다. CPU 4 나노 EUV 공정, GPU 5나노 공정, I/O 와 SoC 6나노 공정으로 칩렛 구조로 포베로스 패키징 기술이 적용되었다. 여기에 AI 연산을 위한 NPU 를 탑재한 것이 특징적이다. 후속 제품인 루나레이크는 2024 년말 출시 예정으로 전작 대비 GPU 와 NPU 성능이 3 배 향상될 것이라고 밝혔다. PC 용 차세대 프로세서인 애로우레이크도 AI 기능을 강화해 2024 년 중 출시될 전망이다.

Qualcomm 이 공개한 스냅드래곤 X 엘리트에는 자체 개발한 오라이온 CPU 가 처음으로 탑재했다. 특히 NPU 탑재를 통해 130 억개 파라미터까지 지원한다고 밝혔다. 관련 제품이 탑재된 랩탑 및 PC 는 2024 년 하반기부터 출시될 예정이다.

엔비디아는 랩탑 및 PC 용 GPU 신제품을 발표했다. 'RTX 4060 슈퍼', 'RTX 4070 Ti 슈퍼', 'RTX 4080 슈퍼' 3 가지 제품으로 게이밍용은 물론 LLM 처리 및 AI 작업 수행도 지원한다.

AMD 도 PC 용 프로세서 라이젠 8000G 시리즈를 공개했다. 특히 상위 2개 제품에는 PC 프로세서 최초로 NPU 를 내장시켰다.





### (3) 엣지 디바이스 AI 대거 탑재

모바일과 랩탑 중심으로 인공지능 기능을 강화한 자체 SoC 기반의 신제품이 2024년부터 대거 출시될 전망이다. 이번 CES 2024 행사에서도 이를 짐작해 볼 수 있었다.

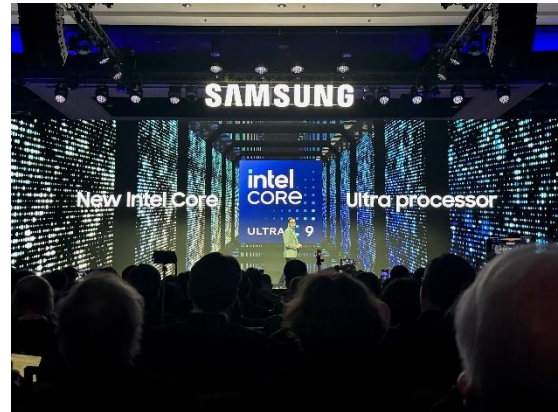
삼성전자는 'AI For All'이라는 주제에 걸맞게 주요 제품의 AI 기능 강화에 초점을 맞추었다. 갤럭시북 4 시리즈에는 인텔의 코어 울트라(메테오레이크) 프로세서를 탑재해 향상된 AI 퍼포먼스를 시현했다. TV에서도 NPU가 2배 빨라진 NQ8 AI 3세대 프로세서를 선보였다. 8K로 upscaling은 물론, 화면 속 피사체 모션 감지와 음성 구분까지 가능해져 몰입감을 높였다. 이 외에도 부스를 통해 향후 AI 서비스에 관한 청사진도 제시하였다. 'Live caption in calls'는 통화 연결시 말하거나 듣기가 어려운 경우, 음성을 텍스트로, 텍스트를 음성으로 변환시키는 기능이다. 'Audio Subtitles'는 TV 내 실시간 번역 기능으로 음성과 텍스트를 동시에 제공하며 30개 언어까지 지원한다. 저지연 Real-time에 강점이 있는 온디바이스 AI가 유리한 영역이다. 1월 17일에는 갤럭시 S24 언팩 행사를 통해 삼성전자 최초 생성형 AI 폰을 선보일 계획이다. 자체 LLM 가우스 모델과 AI SoC를 통해 동시 통역 통화 등 새로운 경험을 제공할 것으로 보인다.

CES 2024, 삼성전자 Media Day 주제



자료: CES2024, SK 증권

삼성전자 갤럭시북 4, Intel Core Ultra 탑재



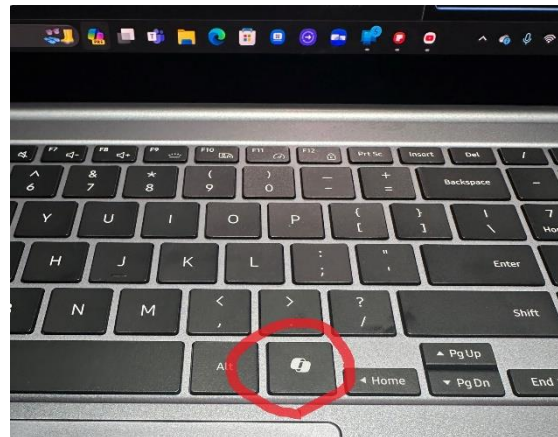
자료: CES2024, SK 증권

삼성전자, 갤럭시 북 4 AI 서비스



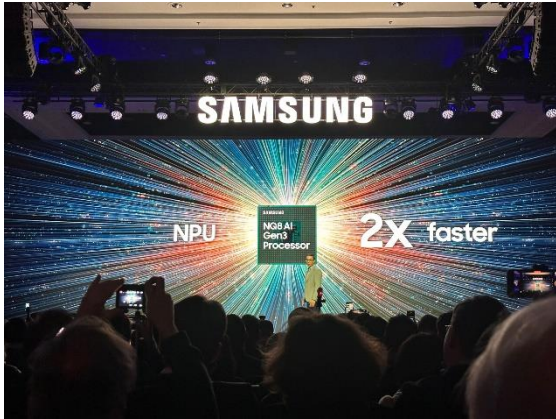
자료: CES2024, SK 증권

삼성전자, 갤럭시북 AI 전용 키



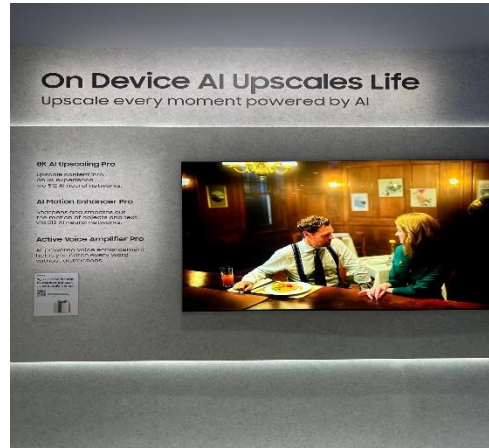
자료: CES2024, SK 증권

삼성전자, TV 용 NQ8 AI processor



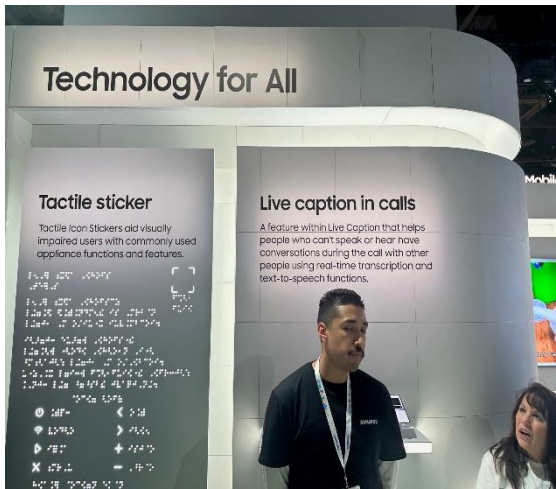
자료: CES2024, SK 증권

삼성전자 TV, AI 접목



자료: CES2024, SK 증권

삼성전자, Live caption Calls



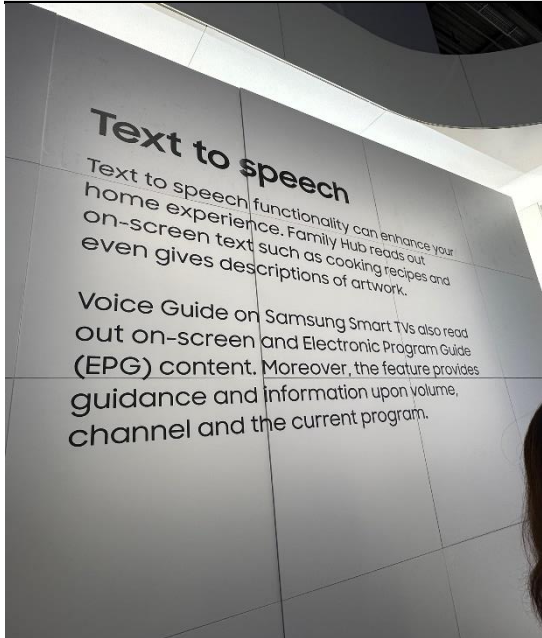
자료: CES2024, SK 증권

삼성전자, Live caption Calls



자료: CES2024, SK 증권

삼성전자, Text to Speech



자료: CES2024, SK 증권

삼성전자, Audio subtitles



자료: CES2024, SK 증권

LG 전자 역시 AI 관련 주제도 많았지만 AI 보다는 무선 투명 OLED 4K TV 인 '시그니처 OLED T'와 전장 관련 청사진에 많은 관심이 쏠렸다. 미래 모빌리티 컨셉인 'LG able'은 모빌리티의 커스터마이징, 편의 향상, 휴식 공간지향 등 컨셉카를 통해 전략을 소개했다.

해외 업체 중 단연 눈에 띄었던 곳은 Qualcomm 이다. NPU 성능을 강화한 스냅드래곤 8 세대 3 을 통해 모바일 내 생성형 AI 사례를 보여주었다. 제시된 문장 속 빈칸에 어울릴 만한 단어를 넣고 읽으면 AI 가 이것을 이미지화 시킨다. 놀라운 것은 같은 문장에 똑같은 단어를 넣어도 생성되는 이미지는 일관성을 가지되 조금씩 차이를 보였다는 점이다. 70-100 억개의 파라미터의 모델과 이를 지원하는 칩셋으로도 온디바이스 내 생성형 AI 구현이 가능하다는 것이 확인된 셈이다. Google 역시 외부 전시장을 활용해 Pixel 8 Pro 로 비슷한 경험을 제공했다.

차량 내 LLM 적용도 눈길을 끌었다. 아마존웹서비스(AWS)은 알렉사 LLM 을 BMW X1 차량에 적용, 운전자가 차량에 관한 정보를 대화를 통해 습득할 수 있는 경험을 마련했다.

LG 전자, 무선 투명 OLED 4K TV



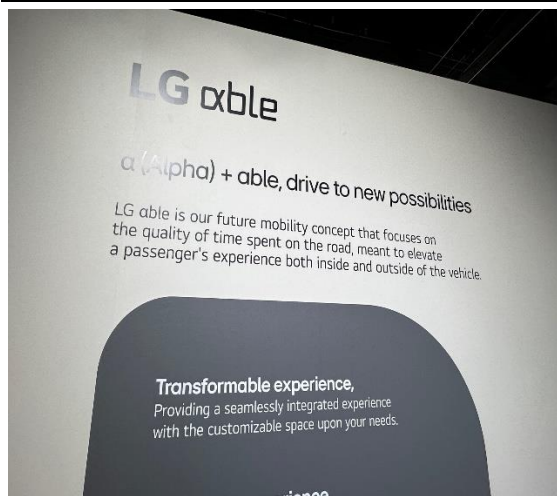
자료: CES2024, SK 증권

LG 전자, 무선 투명 OLED 4K TV



자료: CES2024, SK 증권

LG 전자, 미래 모빌리티 컨셉 'LG able'



자료: CES2024, SK 증권

LG 전자, 미래 모빌리티 컨셉 'LG able'



자료: CES2024, SK 증권

Qualcomm, 생성형 AI 사례 1



자료: CES2024, SK 증권

Qualcomm, 생성형 AI 사례 2



자료: CES2024, SK 증권

Qualcomm, 생성형 AI 사례 3



자료: CES2024, SK 증권

Google, 생성형 AI 사례 4



자료: CES2024, SK 증권



AWS, BMW 차량 내 LLM 서비스



자료: CES2024, SK 증권

Google, 생성형 AI 벽지



자료: CES2024, SK 증권

### (3) 엣지 디바이스 생성형 AI 접목

온디바이스 AI가 새로운 tech trend로 떠오르는 배경에는 클라우드 기반의 비용 부담 증가, 개인화&보안화, 저지연 등이 있다. 반대로 서비스의 부재, 클라우드 대비 낮은 정확성, 엣지 디바이스 단의 BoM cost 부담은 여전히 약점이다. 결국 클라우드와 엣지는 각 시스템 체계에서 장점을 극대화 시킬 수 있는 하이브리드 AI로 진화되어 나갈 것이다. 빅데이터 분석과 같은 까다로운 작업은 클라우드 AI를 통해, 빠르게(실시간 통역) 혹은 간단히(이미지 및 영상 보정) 처리해야 하는 작업은 엣지 AI를 이루어질 것이다. on-device AI 시장도 2024년을 기점으로 개화할 것으로 예상된다.

특히 Tech Trend 변화는 결국 소비자의 필요 조건에 의한 요구가 아니라 Tech 업체가 부가가치 창출을 위해 주도한다는 점을 기억할 필요가 있다. 스마트폰의 카메라가 듀얼, 트리플, 쿼드러플까지 가는 과정을 상기해보면 소비자의 요구보다는 스마트폰 제조사가 Spec 상향을 주도했고 단순 스마트폰을 넘어 성능 좋은 카메라까지 경험이 확장되었고, 결국 관련 supply chain의 부가가치가 창출로 귀결되었다. On-device AI 역시도 비슷한 관점에서의 접근이 필요하다.

Qualcomm, Intel, Meta, Google, 삼성전자 등 주요 빅테크 업체가 하나 같이 on-device AI를 얘기하고 있고 sLLM, 칩셋, 디바이스까지 기반 기술 준비를 끝내고 하나 둘 시장에 모습을 드러내고 있다. On-device AI만이 가질 수 있는 장점 그리고 이후 등장하게 될 서비스들은 또 한번 스마트폰의 새로운 경험 그리고 supply chain의 부가가치를 높일 수 있는 계기가 될 것으로 전망한다.

## 2. Cloud AI 인프라

### (1) AI Server 투자 여전히 강하다

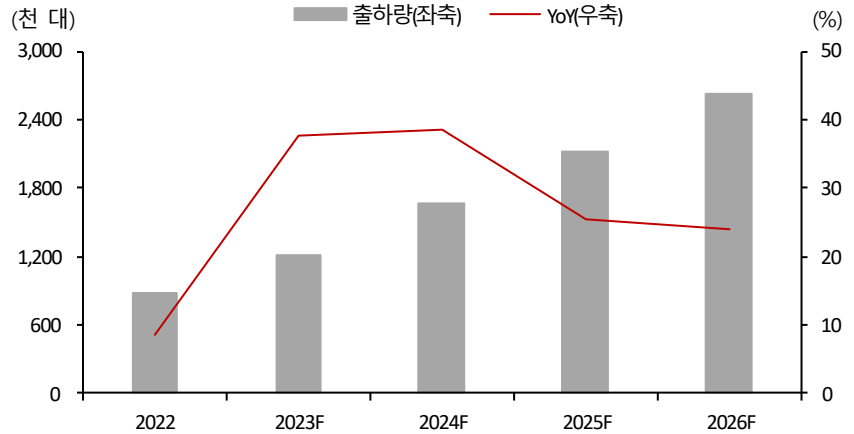
클라우드 AI 시장도 여전히 뜨겁다. 클라우드 AI 의 장점은 고도화된 학습과 추론이 가능하다는 점이다. 빅데이터 분석이나 데이터 결과값에 대한 신뢰도가 훨씬 높다. 시장 규모나 요구 되는 투자 규모(서버 vs 엣지 디바이스)도 훨씬 더 크다. 클라우드 기반의 AI Server 출하량은 2022년 88만대에서 2026년 263만대 수준까지 급성장할 것으로 기대된다. 주요 서버 업체들의 CAPEX는 2023년 줄어들었지만 AI Server 시장 확대에 2024년부터는 재차 증가할 것으로 보인다.

관련 업체 AI 가속기 시장 경쟁도 여전히 치열하다.

Nvidia는 지난해 11월 H200을 공개, 2Q24 출시 예정이다. H100 대비 출력 속도, 용량, 대역폭이 2배 증가한 제품이다. AMD도 지난해 12월 차세대 GPU인 MI300X를 공식 출시했다. 두 제품 모두 LLM 등 생성형 AI 구동을 위한 전용 AI 칩이다.

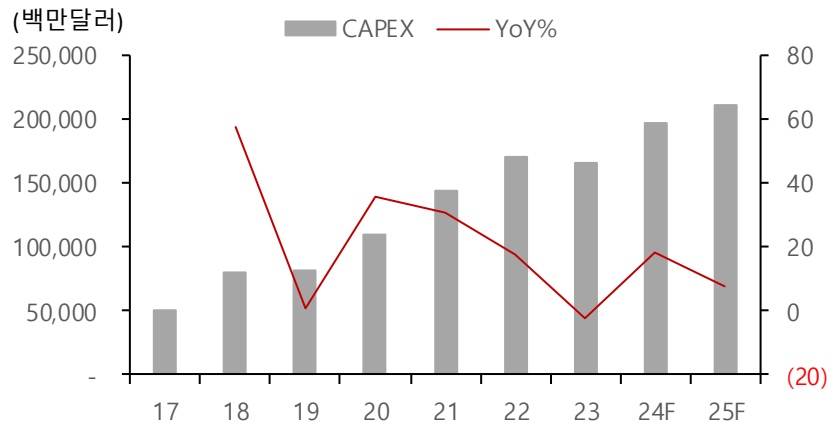
Intel은 올해 신형 AI 가속기인 '가우디3'를 출시할 계획이다. 가우디3는 전작 대비 처리 속도는 4배, HBM 탑재 용량은 1.5배 증가하며 LLM 최적화에 힘을 쏟았다. 현재 AI 칩 시장을 끌어가고 있는 Nvidia, AMD에 이어 Intel까지 시장 진입이 본격적으로 이루어진다면 수요 단에서의 가격 이슈는 점차 해소될 것으로 보인다.

Global AI Server shipments



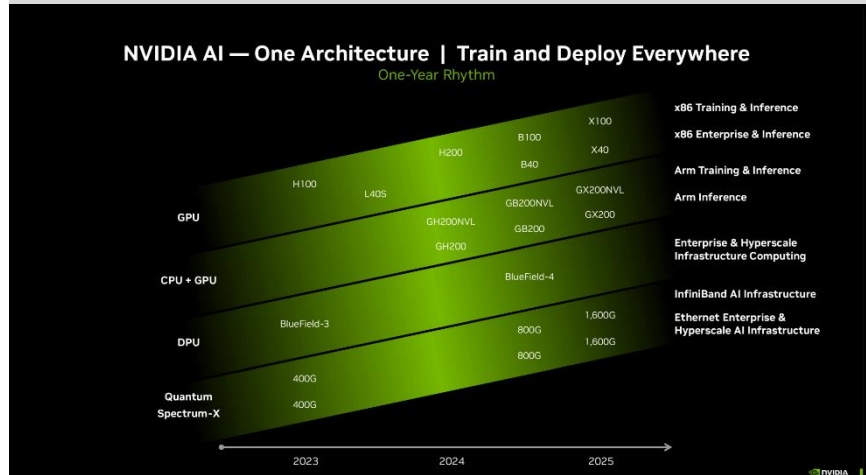
자료: Trendforce, SK 증권

주요 서버 업체 CAPEX 추이 및 전망



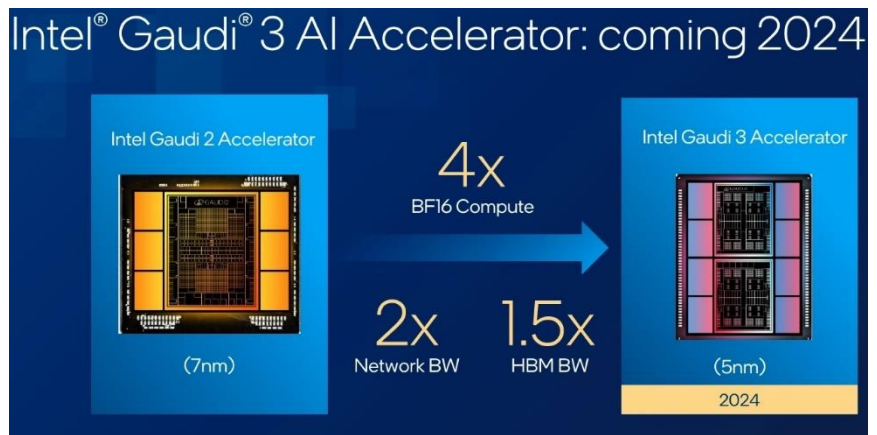
자료: Bloomberg, SK 증권

Nvidia, AI chip Roadmap



자료: Nvidia

Intel, 신형 AI 가속기 'Gaudi3'



자료: Intel

## (2) 투자 기회는 여전히 Cloud AI에 많다

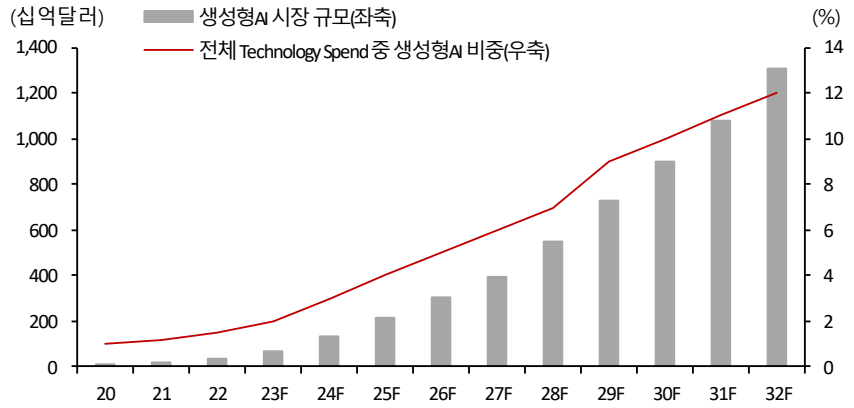
블룸버그 인텔리전스에 따르면 전세계 생성형 AI 시장 규모는 2022년 400억달러, 2023년 670억달러, 2027년 4,000억달러까지 증가한다. 하드웨어와 소프트웨어가 포함된 수치로 Cloud AI 인프라 투자만 구분해보면 2022년 320억달러에서 2027년 830억달러까지 2.5배 늘어난다. 2024년은 62조원 시장으로 추정된다.

On device AI 시장 규모는 어떨까? 옛지 단에는 이미 AI 기능을 하는 칩셋들이 많이 탑재되어 있다. TV에도 업스케일링을 도와주는 AI 전용 칩이 있고, 맞춤형 콘텐츠를 추천하는 AI 셋탑박스, 카메라 사진과 영상 품질 향상을 위해 스마트폰에도 NPU가 쓰이고 있다. 자연어를 이해하고 추론까지 할 수 있는 sLLM으로 한정해서 보게 되면 시장은 이제 막 열리기 시작했다. 스마트폰 내 on device AI로 파생되는 시장 규모는 2024년 5조원 내외로 추정된다. AI PC의 본격적인 개화 시점은 2025년, 가전 및 주변기기의 경우 고성능 칩 탑재보다는 basic 수준의 연산이 가능한 AI 칩 탑재 혹은 클라우드 환경 연결 강화를 통해 인공지능 서비스가 주로 활용될 것으로 예상된다.

On Device AI는 새롭게 열리는 시장임이 틀림없고 초기 시장 규모도 5-6조원 내외로 적지 않은 수준이다. SoC 칩, DRAM, NAND, 수동부품, 기판 등에서 부가가치가 창출되는데 그 중 80% 이상이 SoC 칩에서 나타난다. 클라우드 단에서 HBM, Packaging 등 새로운 파생 영역이 나타나는 것과는 조금 다르다. Qualcomm, Mediatek 과 같은 AP 업체, AI PC 칩셋을 대응 준비 중인 Qualcomm, Intel, Nvidia로 수혜가 집중될 것으로 보인다. 소켓, 기판, IP 쪽 낙수 효과 정도를 기대해 볼 수 있다.

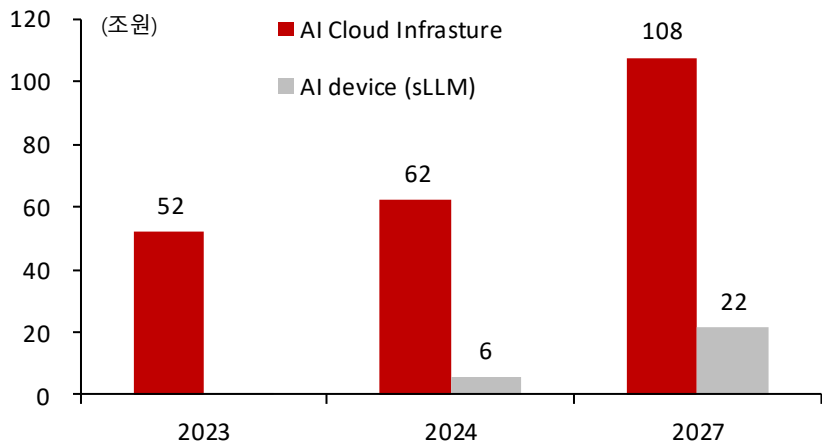
국내 관련 투자 기회는 Cloud AI 쪽이 더 클 것으로 생각한다. 규모 면에서도 크고 선택지도 많으며 국내 업체가 강점이 있는 파생 영역 등이 있기 때문이다. Cloud AI 관련 소켓, 기판, HBM 업체에 대한 투자 전략을 추천한다.

생성형 AI 시장 규모 전망



자료: 블룸버그 인텔리전스

Cloud AI, On-device AI 시장 규모 추정 (하드웨어 기준)



자료: SK증권

**Compliance Notice**

작성자(관리자)는 본 조사분석자료에 게재된 내용들이 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭없이 신의성실하게 작성되었음을 확인합니다.  
 본 보고서에 언급된 종목의 경우 당사 조사분석담당자는 본인의 담당종목을 보유하고 있지 않습니다.  
 본 보고서는 기관투자가 또는 제 3자에게 사전 제공된 사실이 없습니다.  
 당사는 자료공표일 현재 해당기업과 관련하여 특별한 이해 관계가 없습니다.  
 종목별 투자의견은 다음과 같습니다.  
 투자판단 3 단계(6개월기준) 15%이상 -> 매수 / -15%~15% -> 중립 / -15%미만 -> 매도

**SK 증권 유니버스 투자등급 비율 (2024년 01월 15일 기준)**

매수	93.89%	중립	6.11%	매도	0.00%
----	--------	----	-------	----	-------