ETRI Insight

ICT정책 이슈&트렌드 2024-12

소버린 AI 동향 및 R&D에 대한 시사점

김성민

ETRI 한국전자통신연구원
Electronics and Telecommunications Research Institute

본 보고서는 ETRI ICT전략연구소 기본사업인
"국가 지능화 기술정책 및 표준화 연구"를 통해 작성된 결과물입니다.

본 보고서의 내용은 연구자의 견해이며 ETRI의 공식 의견이 아님을 알려드립니다.
보고서의 시장 자료는 출판사의 사용 허가를 받아 사용되었습니다. 재사용을 원하시면 출판사에 문의하시기 바랍니다.

목 차 CONTENTS

Executive Summary

I. 개요 ····· 1

1. 연구 개요 ····· 1
2. 소버린 AI 개요 ····· 2

II. AI 시장 현황 및 특징 분석 ····· 7

1. AI 시장 전망 ····· 7
2. AI 투자 점유율 ····· 8
3. LLM 시장 동향 및 특성 ····· 9
4. AI 이용자 점유율 ····· 11

III. 소버린 AI 관련 국내외 동향 ····· 12

1. 현재 AI 서비스의 핵심 기술 - LLM ····· 12
2. 전략 자산화에 대비해야 할 핵심 분야 - 국방 AI ····· 18
3. AI 개발 역량 확보를 위한 핵심자원 - 컴퓨팅 인프라 ····· 23

IV. 소버린 AI 확보를 위한 시사점 ····· 28

참고문헌 ····· 30

Executive Summary

1 소버린 AI 개념 및 범위

○ **(개념)** 국가 AI R&D 관점에서 "소버린 AI"를 다음과 같이 정의함

- 국가가 국민들을 보호하고, 국민들의 경제·사회 활동을 위해 필요한 AI를 스스로 개발하고 운영할 수 있으며, (전략 자산화에 대비하여) 자국의 미래 AI 및 AGI 개발 역량을 확보하는 것

○ **(중요성)** AI 기술주권, AI Arm Race 대비, 미래 AI 기술패권 확보

- (AI 기술 주권) 외부에 의존하지 않고 자국 데에터를 활용하여 공공의 이익과 목적에 부합하는 AI 모델을 스스로 개발하고 활용할 수 있는 능력 필요
- (AI Arm Race 대비) 자국의 국방 및 치안 등 안보에 직결되는 분야에 대해 외국 및 민간 AI에 의존하지 않는 자주적이고 독립적인 AI 기술 확보 필요
- (미래 AI 기술 패권) 미래 AI는 기술 유무와 발전 정도가 핵무기 보유국과 비보유국간의 힘의 불균형만큼 국가 간 역량에 결정적 차이를 낼 것으로 전망하며 전략기술화하고 있음

○ **소버린 AI의 3단계 정의**

[소버린 AI의 3단계]

구분	데이터	AI 모델	해외/민간 사업자 관여도
1단계: 데이터 소버린	자국 법률 준수 및 자국 내에 보유	자국 언어, 문화, 가치관 반영 스스로 LLM 적용/운영 (오픈소스 모델 활용 가능)	국내 사업자 주도 해외 사업자 기술 활용
2단계 : 모델 소버린	자국 기업 내에 데이터 보유, 자국 정책 및 규제 적용	자국 기업이 스스로 모델 개발/운영	국내 사업자 주도 국내 자체 기술 활용
3단계 : AI 기술 소버린 공공 소버린	자국 기업/기관이 데이터 생성, 유통, 활용을 체계으로 관리	자국 기업/기관이 공공분야 AI 및 AGI 기술진화 주도	국가(공공) 및 국내 사업자 주도 국내 자체 기술 고도화

○ **검토 범위 및 검토 내용**

① 현재 AI 기술 경쟁력의 핵심, 『초거대 언어모델(LLM)』

② 국가가 외국 및 민간 AI에 의존하지 않고 가장 안전하게 개발하고 활용할 수 있어야 하며, 전략 자산화에 대비해야 할 핵심 분야, 『국방 AI』

③ 스스로 AI를 개발하고 운영하기 위한 핵심 자원, 『컴퓨팅 인프라』

2 AI 시장 현황 및 특징 분석

○ 소버린 AI가 성공하기 위해서는 빅테크 기업의 AI 대비 차별적 경쟁력을 확보하고, 시장에 진입 및 안착할 수 있어야 하므로, AI 시장의 현황 및 특성을 분석함

○ (시장규모 전망) 글로벌 AI 시장은 2023년 3,780억~4,836억 달러 규모로 추정되며, 연평균 20%로 성장하여 2027년에는 8천억~1조 달러 규모가 될 것으로 전망됨

○ (투자 점유율) 2021년~2023년간 상위 6개국의 AI 투자 금액을 분석해 보면, 미국이 이들 투자의 60%를 차지하고 있으며, 특히 미국 민간 부문의 투자가 압도적 비중을 차지하고 있음

○ (GPU구매 점유율) 미국 빅테크 기업들은 AI 개발의 핵심 자원인 H100 구매에서도 압도적인 비중을 차지하며, 2023년 엔비디아의 상위 12개 고객사 판매량 중 81%를 미국 기업들이 차지함

○ (초거대 언어모델수) EPOCH AI에 따르면, 2024년 4월 기준 초거대 언어모델(LLM) 수는 총 81개로 조사되었으며, 이는 2023년 4월 42개 대비 2배 가량 증가한 숫자임

○ (LLM 이용자 수) LLM 사용 현황을 보면, 세계적으로 확인된 초거대 AI 모델수만 81개에 달하나, ChatGPT 사용 비중이 82.5%, 상위 5개사의 비중이 91.2%를 차지하며 Top 5 사업자로 쏠림 현상이 뚜렷함

3 소버린 AI 관련 국내외 동향

□ **현재 AI 서비스 핵심 기술 - LLM : 국가별 '차별화'와 함께 '확장적' 소버린 추진**

○ 빅테크 의존도를 줄이고 독립적인 AI 및 클라우드 인프라 구축을 목표로 하되, 부족한 기술 및 자원을 동맹국 및 파트너 기업으로부터 조달하고, 해외 주변국으로의 시장 확장을 고려하여 협력 기반 소버린을 추진하고 있음

○ (특징 1) 소버린 AI는 빅테크 기업 AI에 대한 '차별화' 전략의 일환으로, 언어/문화/지역적 특성으로 차별화하는 한편, 기업용 IT 솔루션으로 '맞춤화'하는 전략을 병행하는 것으로 볼 수 있음

- (현지화·지역 특화) 지역별 언어와 문화, 법제도적 요구사항에 맞는 데이터 학습 및 서비스 제공을 추진하며, 이를 위해 현지 기업과 파트너쉽 추진

- (B2B 서비스 제공) 산업용 또는 기업용으로 클라우드 서비스 및 기업용 솔루션과 결합하여 제공되면서 수익성 확보 추진

○ (특징 2) 수익성과 경제성을 확보하기 위해, 기술적으로는 '멀티모달 및 경량화'로 진화시키는 한편, 지역 또는 적용 서비스의 확장을 추진하고 있음

- (멀티모달 및 경량화 추진) 언어, 이미지, 영상 등 멀티모달 데이터를 처리할 수 있는 AI를 개발하는 한편, 경량화된 모델을 통해 비용 절감 및 다양한 하드웨어 환경 적용 확장

- (다국어 지원) 대부분의 모델이 자국어 및 주변국 언어에 대한 강점을 확보하고 자국뿐만 아니라 주변국으로 시장 확장 시도

□ **전략 자산화에 대비해야 할 핵심 분야 - 국방 AI : 폐쇄적 경쟁 심화 전망**

○ 국방에 AI가 도입되면서 군사 자산관리, 의사결정 지원 등이 효율화되고 있으며, 전장상황에서 AI의 적용은 군사력 향상에 영향을 줄 것임

○ 람다 모델을 중국에서 군사용으로 활용한 정황이 포착되고, 이에 대한 대응으로 미국 국방 관련 기관 및 관련 국가에 그간 금지되었던 람다 모델이 군사용으로 허용되면서 AI의 국방 분야 활용은 더욱 가속화될 것으로 보임

○ 오픈소스 AI의 위험성에 대한 인식이 높아지게 되면서 자국 및 동맹국 이외의 국가에게는 기술 공개 및 공유가 금지될 것으로 전망됨

○ 국방분야 경쟁력 확보에 활용 가능한 국방 특화 소버린 AI 자체 모델 확보가 시급함

□ **AI 개발 역량 확보를 위한 핵심 자원 - 컴퓨팅 인프라 : 다양한 확보 전략**

○ **컴퓨팅 자원의 신속한 확보를 위해 다양한 전략 구사** : ▷직접구매, ▷GPUaaS, ▷엔비디아 협력, ▷클라우드 협력(지분투자) 등 다양한 방식 중 전략적 선택 필요

4 소버린 AI 확보를 위한 제언

□ **핵심 이슈**

○ 소버린 AI는, ▷소수의 빅테크 기업들이 주도하는 LLM 시장에 진입하고 안착하기 위한 기업 전략이자, ▷각국 정부의 AI 주권 확보를 위한 정부정책의 이중적 성격을 띠고 있음

○ 소버린 AI가 정부 정책적 성격이 있더라도 결국 AI기업이 LLM 개발 및 운영의 주체가 되어야 하므로, 소버린 AI가 성공하기 위해서는 LLM 시장의 지배적 사업자인 빅테크 기업의 AI 대비 차별적 경쟁력을 확보할 수 있어야 함

○ 국가 AI R&D 관점에서 소버린 AI의 핵심 이슈는 아래와 같이 정리할 수 있음

① LLM 모델의 경쟁력 확보를 위한 진화 : 멀티모달과 경량화로 고도화

- (멀티모달과 경량화) 수익성과 경제성을 확보하기 위해, 기술적으로는 '멀티모달 및 경량화'로 진화시키는 한편, 지역 또는 적용 서비스의 확장을 추진하고 있음

② AI 시장 선점을 위한 '속도' : 핵심 역량의 신속한 확보를 위해 폐쇄적 협력이 확대될 전망

- (투자와 협력) 자사가 보유하지 못한 자원을 신속하게 확보하기 위해 협력 확대

- (폐쇄적 협력) 국방 AI는 긴밀한 동맹국간 또는 협력 기업간에만 폐쇄적으로 협력할 전망

③ 컴퓨팅 확보 : ▷직접구매, ▷GPUaaS, ▷엔비디아 협력, ▷클라우드 협력(지분투자) 등

다양한 방식 중 전략적 선택 필요

□ **국가 및 출연연 R&D에 대한 시사점**

○ (국내 AI 기업) 글로벌 사업자 대비 차별화가 가능한 전문AI 및 특화 AI 타겟시장 발굴 필요

- 신속하게 경쟁력 있는 AI 기술을 확보하기 위해 출연연-대학-기업들간 적극적인 협력이 필요하며, 타겟시장 선정 시 해외 시장을 동시에 고려해야 함

○ (정부) 속도감 있는 컴퓨팅 자원 확보와 국가 AI 역량 집결을 위한 중심체로서 역할을 해야 하며, 해외 시장 개척을 위해 타겟 국가 정부와 협력의 중심이 되어야 함

○ (출연연) 해외 기업 및 민간 기업에 의존하지 않고 개발 및 운영 가능한 멀티모달 AI 모델과 국방 AI 모델 개발이 필요함

- 컴퓨팅 자원의 조속한 확보 및 출연연이 개발한 기술의 조기 상용화(서비스 제공)을 위해 민간 AI 기업과 적극적인 협력체계 마련이 필요

Ⅰ 개요

1. 연구 개요

□ 연구 배경

o AI가 기술, 산업, 일자리, 안보 등 전방위적으로 영향력을 확대하는 한편, 미국 소수 빅테크로의 집중 현상이 심화되면서 전 세계 각국은 AI 주권을 지키려는 'AI 민족주의'를 강화하고 있음[1)]

* AI 민족주의 : 이코노미스트지는 AI가 정치, 경제, 군사, 사회적으로 중요해짐에 따라 AI 기술을 보유하고 통제하는 것이 자국의 이익과 안보에 중요해지고 있기 때문에 국가가 주도하여 독자적 AI 기술개발 역량을 확보하고 전략적 동맹을 형성하며 경쟁국에 기술접근을 제안하는 AI 민족주의 시대가 열렸다고 논의함(AI Nationalism)[2)]

* AI 제국주의 : AI 선도국들이 개발도상국에 AI 기술을 보급하면서 전력 공급을 확보하고 영향력을 확대하는 현상에 대해 AI 기술력으로 디지털 영토를 장악하는 새로운 형태의 제국주의라는 논평도 있음[3)]

o 세계 각국과 국내에서 '소버린 AI'에 대한 관심과 논의가 활발해지고 있으나, 다양한 주체들이 다양한 관점에서 논의하면서 그 개념 및 목적에 대한 혼란도 가중되고 있음

o 본 고에서는 '국가 AI 이니셔티브 확보를 위한 R&D 전략' 관점에서 '소버린 AI'를 어떻게 바라보아야 할지 '개념과 범위'를 정리하고 앞으로 '소버린 AI' 기술 경쟁력 확보를 위한 전략 방향을 도출하고자 함

□ 연구 목적

o 국가 AI 이니셔티브 확보를 위한 **『국가 AI R&D 관점의 소버린 AI』** 정의

o 소버린 AI 확보를 위한 전략 방안 제안

□ 연구 내용

o 소버린 AI 개요 : 다양한 관점 검토 및 **『국가 AI R&D 관점의 '소버린 AI'』** 정의

o 소버린 AI 관련 국내외 동향 : ①초거대 언어모델(LLM), ②국방AI, ③컴퓨팅 인프라

o 소버린 AI 확보를 위한 시사점 도출

---

1) ZDNET Korea (2024.7.17.), "美에 종속되기 싫다…네이버도 꽂힌 '소버린 AI'에 전 세계 들썩".
2) The Economist (2024.1.1.), "Welcome to the era of AI nationalism".
3) 매일경제 (2024.11.27.), "왜 소버린AI 인가", 매경춘추(국가인공지능위원회 법·제도분과위원장 강민구 변호사 논평).

2. 소버린 AI 개요

□ 소버린 AI의 등장 배경

o **(LLM 쏠림)** 세계적으로 LLM 수는 증가하고 있으나 ChatGPT가 82.5%, 상위 7개사가 95%를 차지하고 있으며, 미국 빅테크 AI로의 쏠림은 한동안 지속될 전망

* 선별된 40개 GenAI의 2024년 3월 월별 방문수 트래픽 기준. ChatGPT가 82.5%, Gemini 4.7%, Poe 1.5%, Perplexity 1.4%, Claude 1.1%, DeepAI 1.1%, Copolit 0.9% (World Bank Group, 2024.8)

o **(문화적 종속 및 가치관 편향 우려)** 생성형 AI들이 미국 중심 데이터로 학습해 미국 가치관에 편향되어 있어 문화와 가치관이 종속되고, 국가 정체성 약화로 이어질 우려 높음[4)]

o **(데이터 보호)** 국민의 데이터를 안전하게 보호하여 민감한 데이터가 유출되거나 외부에 의해 악용될 위험을 방지하고 국가의 데이터 주권 확보에 대한 필요성 증가

o **(정부기능에 AI 활용 증가)** 각국 정부들이 AI 기술을 자국의 정부 활동에 깊숙이 활용하고, 국가 안보 및 정치 활동에 활용하면서 통제 가능한 AI에 대한 필요 증대

- 정부의 크고 작은 의사결정을 지원하는 도구로 활용이 높아지면서 잠재적 편향성과 이해관계의 충돌 문제를 해결하기 위해 설명 가능하고 투명한 AI에 대한 필요성 증대

□ 소버린 AI 관련 다양한 쟁점[5)]

o AI주권과 소버린 AI에 대한 관심이 높아지는 한편, AI 글로벌 오픈소스 생태계에서 자유롭고 빠르게 성장해온 **AI의 기술적 특성** 및 **협소한 내수시장**만으로는 성장이 어려운 **국가적 특성** 등을 고려할 때, 과연 소버린 AI가 바람직한 방향인가, 어디까지 '소버린'을 확보해야 하는가 등 다양한 논의가 있음

o **(한국적 AI vs. 포용적 AI)** 한국의 언어, 문화, 역사, 가치관 등이 잘 반영된 AI가 필요하다는 견해와, 기술은 중립적이어야 하며 다양한 사람들이 함께 사용하고 혜택을 누릴 수 있는 포용적 AI(Inclusive AI)가 더 중요하다는 견해가 있음

o **(LLM vs. Full stack 인프라)** 소버린 AI의 대상에 대해 LLM에 대한 관심이 가장 높으나, 독자적 AI 확보를 위해서는 GPU, AI 반도체, 데이터, 전력 등 모든 인프라가 필요함

o **(자립(민족주의) vs. 협력(글로벌리즘))** 소버린 AI는 기술적 자립을 우선시하는 개념이나, LLM 파운데이션 모델에 국한한 개념이 아닌 AI 개발을 위한 모든 기술 계층을 고려할 때 현실적으로 자립 가능성과 해외 협력 전략(특히 GPU)을 함께 고려해야 함

---

4) 한국소프트웨어저작권협회 (2024.7.25.), "소버린 AI의 국제적 동향", SW·저작권 동향리포트(제2024-14호).
5) ZDNET Korea (2024.11.25.), '소버린 AI' 어떻게 봐야할까..."담론 기반 현실적 전략 필요". 지디넷코리아와 포티투마루가 공동으로 '소버린 AI, 반드시 필요한가'를 주제로 개최한 좌담회 내용 및 관련 자료들을 기반으로 정리함.

o **(Nation State vs. Network State)** '국가(Nation State)'가 AI 주권의 주체가 되어야 한다는 견해와, 주권은 기업과 시민사회 등 다양한 주체로 확장되었으며 국가는 네트워크 허브로서 다양한 행위자를 연결하고 조율하는 네트워크 스테이트(Network State)가 되어야 한다는 견해가 공존

o **(자국 AI 기업 지원 vs. 대기업 지원에 대한 거부감)** AI 시장은 민간 기업이 중심이 될 수밖에 없고 파운데이션 모델 확보를 위해서는 막대한 자원이 필요하므로 자국 AI 기업을 적극 지원해야 한다는 견해와, AI 기업이라고 해서 대기업에게 직접 지원을 주는 것에 대한 거부감이 공존

o **(오픈소스 기반 신속한 개발 vs. 독자 기술 확보)** 고성능 최신 AI 모델들이 오픈소스로 공개되면서 이를 기반으로한 최고 성능의 모델을 신속하게 개발하는 것이 중요하다는 견해와, 오픈소스 모델의 숨겨진 위험성 및 상용화의 한계, 오픈소스 정책변화 등의 위험을 대비하기 위해 독자 기술의 확보가 필요하다는 견해가 공존함[6)]

* 오픈소스는 불특정 다수의 개발자들이 참여하여 만들어가므로 프로그램 개발 속도 및 업데이트 속도가 빠르고 호환성이 높은 반면, 고도화가 어렵고 지적재산권 및 보안 문제 등의 가능성 높음
* 구글 '젬마' 등은 소스코드와 데이터는 공개하지 않고 모델만 공개하는 '오픈모델' 정책으로 개발자들이 맞춤형으로 가공하고 조정할 수 있으나 재배포와 변형은 금지하고 있음

[참고] 메타의 '라마'와 LG '엑사원 3.0'의 오픈소스 정책

메타의 '라마'[7)]

o 2024년 8월까지 3억 5천 만건의 다운로드를 기록하며 오픈소스 AI를 주도
o Llama Stack이라는 개발도구 패키지(데이터 전처리, 하이퍼파라미터 튜닝, 모델 평가 등 각종 라이브러리와 프레임워크를 포함)를 모델과 함께 공개
o 라마(Llama) 3.2 모델은 'Llama 3.2 커뮤니티 라이선스(Llama 3.2 Community License)' 하에 배포. 메타의 지적재산권 하에 라마 3.2 모델을 사용, 복제, 배포, 수정 및 파생 작업을 허용
o 단, 재배포 시 해당 라이선스를 함께 제공하고, 'Built with Llama'라는 문구를 명시해야 함.
o 또한, 월간 활성 사용자 수가 7억 명을 초과하는 경우 메타로부터 별도의 라이선스를 받아야 함.
o 사용 정책에서는 폭력 또는 테러 등의 불법적, 위법적 활동에 해단 사용을 금지하고 있음.

LG 엑사원 3.0

o 한국어와 영어를 학습하고 이해할 수 있으며 Real-world Use cases에서 1위, 코딩, 수학 등 13개 벤치마크에서 1위를 할 정도로 전문산업현장에서 활용성이 높게 개발된 모델로 오픈소스로 공개됨.
o 엑사원 3.0은 연구 목적으로 사용이 허가된 모델로, 상업적 사용을 위해서는 별도의 라이선스 계약이 필요하며, 모델과 파생물을 재배포할 수 있으며, 이때 라이선스 사본을 포함해야 함.
o 비상업적 맥락에서 경쟁 대회에 모델을 제공할 수 있지만, 상업적 라이선스가 없는 한 상업적 목적으로는 사용할 수 없음.

6) Tech World (2024.9.15.), "오픈소스와 AI, 무엇이 문제인가" 및 ETRI 내부 토론회 등에서 논의됨.
7) https://github.com/meta-llama/llama-models/blob/main/models/llama3_2/LICENSE.

□ 소버린 AI 에 대한 다양한 정의

o 소버린 AI에 대해 다양한 기관에서 다양한 관점에 따라 정의하고 있음

- **(네이버 – 자국 자산+문화/가치관)** 각 국가가 자체 데이터와 인프라를 활용하여 그 국가나 지역의 제도, 문화, 역사, 가치관을 정확하게 이해하는 AI를 개발하고 운영하는 것[8)]
- **(엔비디아 – 자체 인프라+자체 AI)** 한 국가가 자체 인프라, 데이터, 인력 및 비즈니스 네트워크를 사용하여 AI를 생산하는 국가의 역량을 의미[9)]
- **(가트너 – 국가 주권 안보)** 국가가 상업 시장에 대한 의존도를 낮추고 AI의 개발과 사용을 자체적으로 운영하려는 노력. 국가의 정치적·문화적 차이를 반영하여 자주적 목표를 달성하는 데 중점. 가치창출과 자국 상황에 적합한 위험 감소 전략 개발을 포함[10)]

❸ 국가 AI R&D 관점에서 『소버린 AI』

□ AI 이니셔티브 확보를 위한 국가 AI R&D 관점에서 『소버린 AI』

o 소버린 AI에 대한 다양한 정의와 쟁점들이 있으나 본 고는 'AI 이니셔티브 확보를 위한 국가 AI R&D 전략방향 수립'을 목적으로 하므로, 다음과 같이 소버린 AI를 정의하고자 함

국가 AI R&D 관점에서 "소버린 AI"

o **국가가 국민들을 보호하고, 국민들의 주체적인 경제·사회 활동을 위해 필요한 AI를 스스로 개발하고 운영할 수 있으며,** 전략 자신화에 대비하여 **자국의 미래 AI 및 AGI 개발 역량을 확보하는 것**

- (AI 기술 주권) 국가가 상업 시장에 대한 의존도를 낮추고 AI의 개발과 사용을 자체적으로 운영하며 자주적 목표를 달성하기 위한 자원과 기술력의 확보
- (미래 AI 기술 패권) AI를 지속적으로 진화 발전시키며 자체적인 미래 AGI 기술개발 역량 및 필요한 인적· 물적 자원의 확보

출처 : 저자 작성

□ 국가 AI R&D 관점에서 '소버린 AI'의 중요성

o **(AI 기술 주권)** 외부에 의존하지 않고 자국 데이터를 활용하여 공공의 이익과 목적에 부합하는 AI 모델을 스스로 개발하고 활용하고 관리할 수 있는 능력 필요

- AI 모델 개발을 위한 데이터, 컴퓨팅 자원, 모델 개발 능력 등 AI 개발을 위한 모든 계층의 기술 및 자원 확보가 필요하며, 이를 위한 경쟁이 심화되고 있음

8) CLOVA (2024.8.19.), "소버린 AI : AI 시대 네이버의 새로운 도전과 과제".
9) NVIDIA Korea (2024.3.4.), "소버린 AI(Sovereign AI)란?" (https://blogs.nvidia.co.kr/blog/what-is-sovereign-ai).
10) Gartner (2024.6.), Hype Cycle for Artificial Intelligence 2024.

o **(AI Arm Race 대비)** 자국의 국방 및 치안 등 안보에 직결되는 분야에 대해 외국 및 민간 AI에 의존하지 않는 자주적이고 독립적인 AI 기술 확보 필요

- 최근 중국이 메타의 오픈소스 모델 '라마'를 군사용 AI 개발에 활용했다는 정황이 포착되면서 메타와 오픈AI가 자사의 AI 모델을 국방용으로 제공하기로 정책을 변경
- 오픈소스 AI가 국가 안보에 위험이 될 수 있다는 경계가 높아지는 한편으로, 기밀정보 동맹체와 같은 고수위 군사동맹간에만 군사용 AI를 공유할 가능성도 높아지고 있음

o **(미래 AI 기술 패권)** 미래 AI는 기술 유무와 발전 정도가 핵무기 보유국과 비보유국 간의 힘의 불균형만큼 국가 간 역량에 결정적 차이를 낼 것으로 전망하며 국가 전략기술화하고 있음

- 미국정부가 AI에 관한 '국가안보각서'에서 AI 기술을 핵무기 같은 국가의 전략 자산으로 지정하고 개발 지원 및 통제를 본격화하는 한편[11], 중국과의 첨단 기술경쟁에서 우위를 점하기 위해 AGI 개발을 지원하는 'AI 맨하튼 프로젝트'를 제안('24.11.19)[12]함

□ 소버린 AI의 3단계

o 소버린 AI를 ▷**데이터**, ▷**AI 모델**, ▷**개발 및 운영 사업자에 대한 해외 의존도(또는 자주성 및 독립성)**에 따라 아래와 같이 3단계로 구분할 것을 제안함

- **(1단계 : 데이터 소버린)** 데이터 보호·관리에 대해 자국의 법률·규제에 따라 통제 가능하며, AI 모델에 자국의 언어, 문화, 가치관을 반영하여 LLM 모델을 개발/운영하는 단계
- **(2단계 : 모델 소버린)** 해외 사업자의 오픈소스 모델 정책이 불리하게 변경될 경우, 자국 기업이 스스로 데이터를 유통/관리하며 스스로 AI 모델을 개발하고 운영할 수 있는 단계
- **(3단계 : AI 기술/공공 소버린)** 국가 안보 및 공공분야에 대해 국가(공공)가 주도적으로 AI를 개발-관리-운영할 수 있으며, AGI의 전략 자산화에 대응할 수 있는 단계

[소버린 AI의 3단계]

	데이터	AI 모델	해외/민간 사업자 관여도
1단계: 데이터 소버린	자국 법률 준수 및 자국 내에 보유	자국 언어, 문화, 가치관 반영 스스로 LLM적용/운영 (오픈소스 모델 활용 가능)	국내 사업자 주도 해외 사업자 기술 활용
2단계 : 모델 소버린	자국 기업 내에 데이터 보유, 자국 정책 및 규제 적용	자국 기업이 스스로 모델 개발/운영	국내 사업자 주도 국내 자체 기술 활용
3단계 : AI 기술 소버린 공공 소버린	자국 기업/기관이 데이터 생성, 유통, 활용을 체계으로 관리	자국 기업/기관이 공공분야 AI 및 AGI 기술진화 주도	국가(공공) 및 국내 사업자 주도 국내 자체 기술 고도화

출처 : ETRI 내부 토론회 내용 및 소버린 AI 관련 문헌을 기반으로 저자 작성

11) 조선일보 (2024.10.31.), "바이든, 미국 최초 AI 안보각서 서명".
12) AI 타임즈 (2024.11.20.), "미국 의회, 'AI 맨하튼 프로젝트' 제안…'중국보다 먼저 AGI 개발해야'"

□ 검토 범위 및 검토 내용

① 현재 AI 기술 경쟁력의 핵심 요소 - 『초거대 언어모델(LLM)』

- 현재 LLM 시장의 현황을 분석하고, 소버린 AI를 위한 LLM 개발과 시장에서 사용되고 성장하기 위해 필요한 기술 및 시장 요건들을 모색해보고자 함

② 국가가 외국 및 민간 AI에 의존하지 않고 가장 안전하게 개발 및 활용해야 하며, 전략 자산화에 대비해야 할 핵심 분야 - 『국방AI』

- 주요국 정부의 동향 분석을 통해 우리 정부가 국가와 국민을 위해 AI를 어떻게 개발하고 운영해야할지 효율적인 방안을 모색해보고자 함

③ 스스로 AI를 개발하고 운영하기 위한 핵심 자원 - 『컴퓨팅 인프라』

- 컴퓨팅 자원을 확보하기 위한 기업 및 각국 정부의 노력을 유형별로 분석하여 우리에게 가장 효율적인 방안을 모색해보고자 함

[참고] 소버린 데이터와 소버린 클라우드

1. 소버린 데이터
- o (개념) 데이터에 대한 국가적 통제권과 소유권을 강조하는 개념으로 데이터가 특정 국가의 법률과 규정에 따라 저장, 처리, 관리되어야 한다는 원칙을 의미[13)]
- o (주요 특징) ▷데이터의 물리적 위치가 해당 국가 내에 있어야 함, ▷데이터에 대한 접근과 사용이 국가의 법률에 의해 통제됨, ▷개인정보 보호와 국가 안보를 위한 데이터 관리 강조

2. 소버린 클라우드
- o (개념) 이는 특정 국가나 지역의 법률과 규정을 준수하면서 클라우드 서비스를 제공하는 인프라
- o (주요 특징) ▷클라우드 인프라가 해당 국가나 지역 내에 위치, ▷현지 법률 및 규제 준수 ▷데이터 접근성과 보안성 강화, ▷국가나 지역의 특정 요구사항에 맞춤화 가능

3. 유사점과 차이점
- o 소버린 데이터와 소버린 클라우드 모두 '데이터 주권'을 확보하는 것이 주된 목적임
- o 소버린 데이터는 법률, 정책, 규제 등을 통해 구현하는 데에, 소버린 클라우드는 물리적 인프라 구축 및 서비스 제공을 통해 구현하는 차이가 있음.
- o 단, 두 개념 모두 물리적 인프라를 자체적으로 구축하는 것을 전제하지 않음 : AWS(2024.7), MS 애저(2023.12), 구글 클라우드(2024.8), 등이 EU 소버린 클라우드 서비스를 시작하는 등 글로벌 클라우드 기업들이 고객 국가의 정책에 맞춰 소버린 클라우드/데이터 서비스를 제공하고 있음

13) https://7hinking.tistory.com/314.

Ⅱ AI 시장 현황 및 특징 분석

* 소버린 AI가 성공하기 위해서는 빅테크 기업의 AI 대비 차별적 경쟁력을 확보하고, 시장에 진입 및 안착할 수 있어야 하므로, AI 시장의 현황 및 특성을 분석함

1. AI 시장 전망

o 글로벌 AI 시장은 2023년 3,780억~4,836억 달러 규모로 추정되며, 연평균 20%로 성장하여 2027년에는 8천억~1조 달러 규모에 이를 것으로 전망됨

o AI 시장을 구성하는 AI 서비스와 AI SW에 대해 가트너는 AI 서비스가 AI 시장의 63%를 차지하는 중심으로, IDC는 AI SW가 전체의 93%를 차지하는 중심으로 전망함

* AI 서비스 시장 : 최종 사용자(기업, 정부, 기관 등)의 요구에 맞춰 AI 모델, 플랫폼, 데이터 서비스 등을 개발하고 도입 및 운영하는 데에 지출하는 금액(Spending)으로 추정
* AI SW 시장 : 사용자가 직접 AI 기능을 구축하거나 클라우드를 기반으로 사용하는 AI 제품과 플랫폼 시장

[글로벌 AI 시장 규모 전망]

(단위 : 백만 달러)

전망 기관	세부	2022	2023	2024	2025	2026	2027	CAGR (23~27)
Gartner	AI SW+서비스	124,316	377,167	443,319	536,713	657,540	803,671	20.8%
	AI SW	124,316	146,437	173,199	206,207	247,459	297,914	19.4%
	AI 서비스	-	230,730	270,120	330,506	410,081	505,757	21.7%
IDC	AI SW+서비스	401,780	483,622	589,076	711,291	852,290	1,019,350	20.5%
	AI SW	372,773	448,762	546,828	659,998	790,025	943,711	20.4%
	AI 서비스	29,007	34,860	42,248	51,293	62,265	75,639	21.4%

출처 : Gartner AI SW - Gartner(2023.11), Forecast Analysis: Artificial Intelligence Software, 2023-2027, Worldwide
Gartner AI 서비스 - Gartner(2024.8), Forecast Analysis: Artificial Intelligence Services, Worldwide
: IDC SW - IDC(2023.9), Forecast Analysis: Artificial Intelligence Software, 2023-2027, Worldwide
IDC 서비스 - IDC(2024.8), Forecast Analysis: Artificial Intelligence Services, Worldwide

주) 도표의 저작권은 Gartner社와 IDC社에 있으며, Gartner社와 IDC社의 동의하에 사용되었습니다. 추후 이용 시 Gartner社와 IDC社에 문의하시기 바랍니다.

o AI SW 시장에 대해 생성형 AI의 성장이 연평균 72%로 매우 높으나, 전통적인 머신러닝에 기반한 Predictive AI 시장이 2028년 전체 AI의 70%를 차지할 것으로 전망됨(옴디아)

[AI SW 유형별 시장 전망]

(단위 : 백만 달러)

AI type	2022	2023	2024	2025	2026	2027	2028	CAGR (22-28)
Generative	2,274	6,868	15,379	24,265	34,830	46,145	58,470	71.8%
Predictive	51,703	68,844	85,784	99,905	112,972	124,293	134,529	17.3%
총합계	53,977	75,712	101,163	124,170	147,801	170,437	192,999	23.7%

출처 : Omdia, Artificial Intelligence Software Market Forecasts - 1H24 Data, 2024.5.

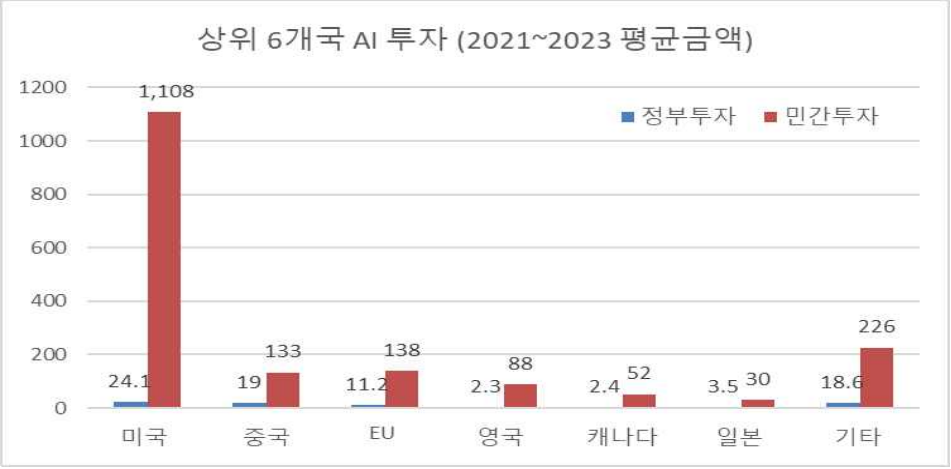
주) 도표의 저작권은 Omdia에 있으며, Omdia의 동의하에 사용되었습니다. 추후 이용 시 Omdia에 문의하시기 바랍니다.

2. AI 투자 점유율

□ 글로벌 정부· 민간 분야 AI 투자 금액

o 2023년 글로벌 AI 투자(정부+민간) 규모는 총 1,419억 달러로 추정되며, 이는 2015년 329억 달러 대비 4.3배 성장한 수준임

o 2021년~2023년간 상위 6개국의 AI 투자 금액을 분석해 보면, 미국이 이들 투자의 60%를 차지하고 있으며, 특히 미국 민간 부문의 투자가 압도적 비중을 차지하고 있음

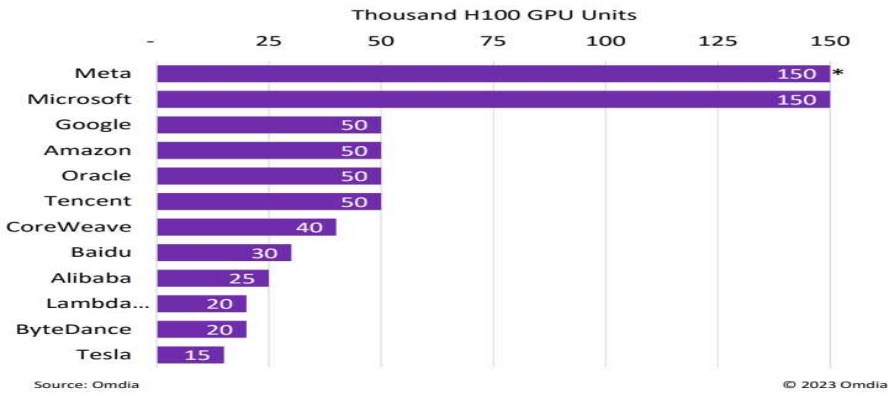


[상위 6개국 정부+민간 AI 투자 금액(2021~2023 평균)]

출처 :한국지능정보사회진흥원(2024.6.28.), "글로벌 정부·민간 분야 AI 투자 동향 분석" 자료 재구성

□ GPU 구매 현황

o 미국 빅테크 기업들은 AI 개발의 핵심 자원인 H100 구매에서도 압도적인 비중을 차지하며, 2023년 엔비디아의 상위 12개 고객사 판매량 중 81%를 미국 기업들이 차지함



[2023년 엔비디아의 주요 고객별 H100 판매량]

출처 : Omdia (2023), Cloud and Data Center Market Snapshop=November 2023.

주) 도표의 저작권은 Omdia에 있으며, Omdia의 동의하에 사용되었습니다. 추후 이용 시 Omdia에 문의하시기 바랍니다.

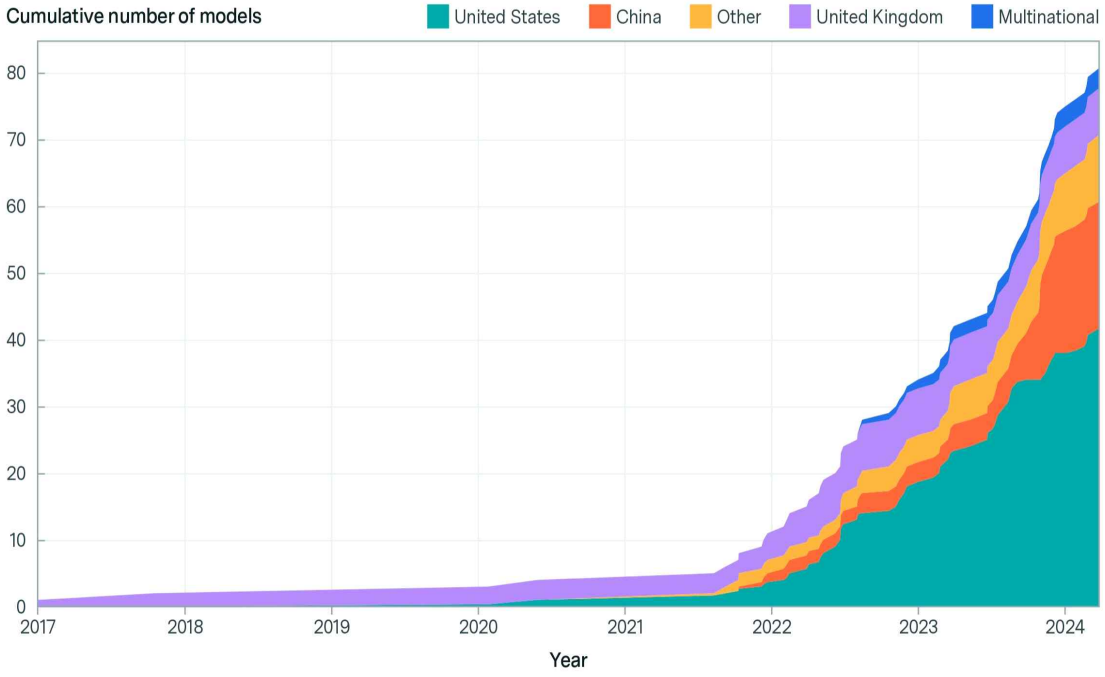
3. LLM 시장 동향 및 특징 : AI 모델 점유율

□ LLM 개발 및 공급 현황 및 특징[14)] : 모델 수 증가 추세, 미국 빅테크 중심

o **(초거대 언어모델 수 : 증가)** EPOCH AI에 따르면, 2024년 4월 기준 초거대 언어모델(LLM) 수는 총 81개로 조사되었으며, 이는 2023년 4월 42개 대비 2배 가량 증가한 숫자임

o **(국가별 보유 : 미국 집중)** 국가별 대규모 모델을 보면, 81개 대규모 AI 모델 중 미국에서 개발된 모델수는 43개, 중국 19개, 영국 6개, 이외의 국가에서 개발된 모델은 10개, 여러 국가 연구자와 조직이 참여하는 협업으로 개발한 모델은 3개로 조사됨

- 단, SPRI(2024.8월) 자료에서는 EPOCH AI 조사에서 국내 초거대 AI 모델들이 누락되었음을 지적하고, 네이버, 삼성, LG, KT, 엔씨소프트, KT, 코난테크놀로지 등 기업들이 개발한 모델이 2021년 3개, 2023년 8개로 누적 11개가 있다고 밝힘[15)]



[국가별 초거대 AI 모델 보유 현황]

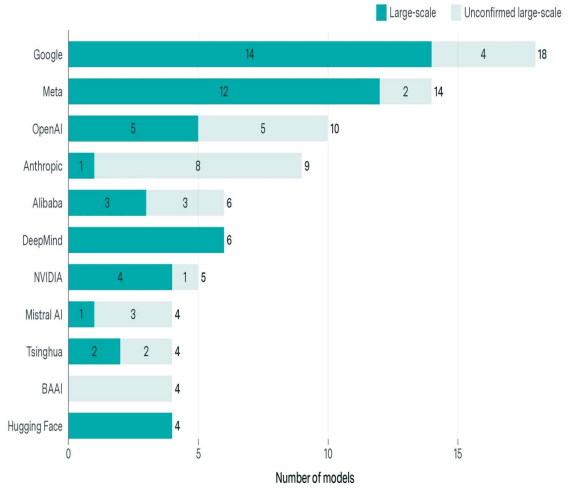
출처 : EPOCH AI (2024.11. 접속)

14) 미국 민간 연구기관인 'EPOCH AI'가 2024년 7월 업데이트한 초거대 AI 모델 현황 데이터를 기반으로 분석함. EPOCH AI는 스텐포드대학교 HAI 연구의 'AI Index'에 소스데이터 제공. 초거대AI의 기준은 미국 AI 개발 행정명령과 EU AI법의 요구사항에 따라 훈련 컴퓨팅 임계값 10^{23} FLOP (GPT-3 수준) 이상 규모의 모델로 정의함 (https://epoch.ai/blog/tracking-large-scale-ai-models#most-large-scale-models-are-developed-by-us-companies).

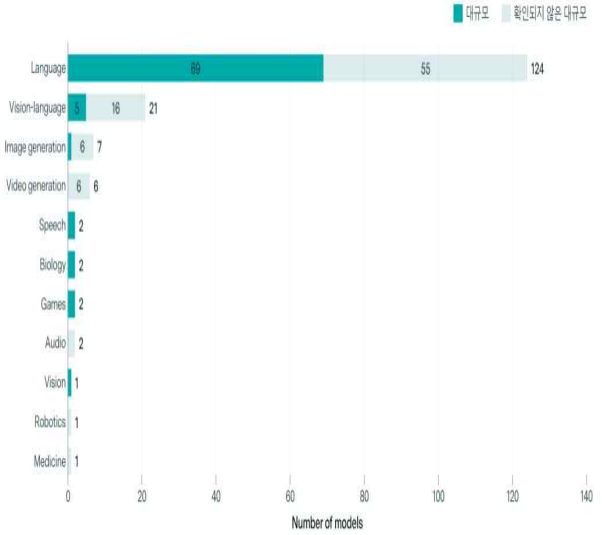
15) 봉강호 (2024.9.), 글로벌 초거대 AI 모델 현황 분석(2020~2023), SPRI 이슈리포트 IS-178. SPRI.

o **(기관별 보유수 : 빅테크 집중)** 대규모 AI 모델의 선두 개발자들은 구글, 메타, 딥마인드, 오픈AI, 앤트로픽 등 빅테크 기업들이며, 산업계에서 개발한 모델이 총 71개, 산학 협력 6개, 정부 기관 개발 모델 2개임

o **(오픈소스 모델)** 81개 대규모 모델 중 38개가 다운로드 가능한 모델로 오픈소스 중 가장 큰 모델은 Falcon-180B 모델(10^{24} FLOP)이었음



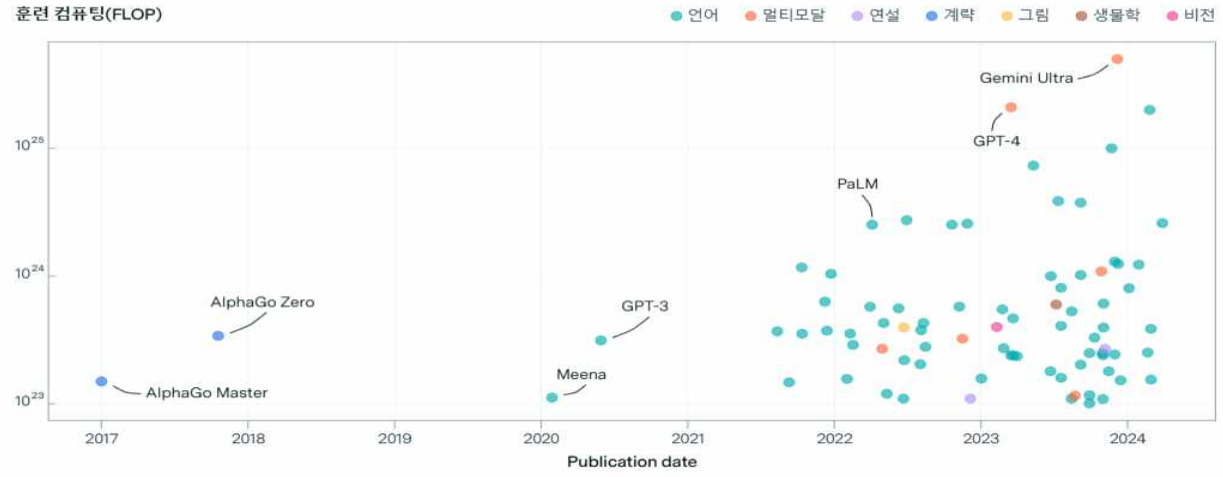
[기관/기업별 초거대 AI 모델 수]



[대규모 모델 종류별 모델수]

출처 : EPOCH AI (2024.11. 접속)

o **(모델의 종류 : 언어모델 중심에서 다양화 시작)** LLM 모델 중 대부분 (81개 중 69개, 85%)가 언어모델이고, 이 중 5개(6%)는 시각-언어 모델이며, 이미지 생성, 비디오 생성, 스피치, 바이오, 게임, 오디오, 로보틱스 분야도 1~2개씩 생기고 있음



[도메인 및 연도별 대규모 모델]

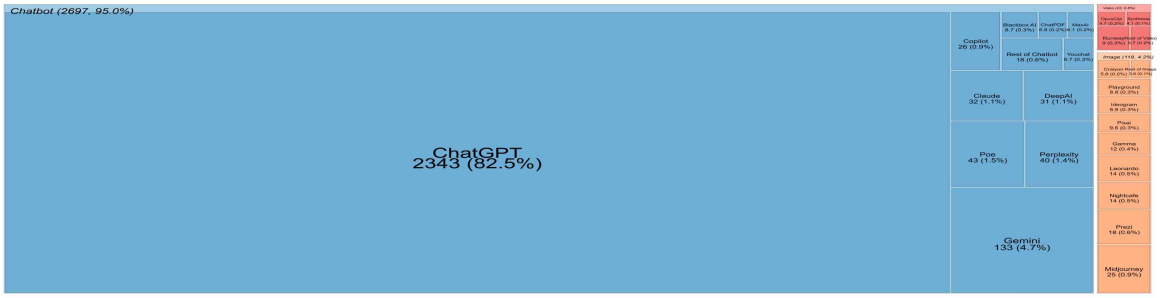
출처 : EPOCH AI (2024.11. 접속)

4. AI 이용자 점유율

□ LLM 수요측 특징 : Top 5 사업자 점유율 91%

o LLM 사용 현황을 보면, 세계적으로 확인된 초거대 AI 모델수만 81개에 달하나, ChatGPT 사용 비중이 82.5%, 상위 5개사의 비중이 91.2%를 차지하며 Top 5 사업자로 쏠림 현상이 뚜렷함

- 생성형 AI 서비스 중 선별된 40개 GenAI의 2024년 3월 월별 방문수 트래픽 기준으로 ChatGPT가 82.5%, Gemini 4.7%, Poe 1.5%, Perplexity 1.4%, Claude 1.1%, DeepAI 1.1%, Copolit 0.9% (World Bank Group, 2024.8)[16]



[생성형 AI 이용 현황]

출처 : World Bank Group, 2024.8

o 기업용 생성형 AI의 경우 1등 사업자인 오픈AI의 점유율이 2023년 50%에서 2024년 34%로 하락한 반면, 엔트로픽은 동기간 12%에서 24%로 두 개 상승하였으나[17] Top 5 사업자의 점유율은 91%로 쏠림 현상 뚜렷함

[기업용 생성형 AI 시장 점유율 변화 (2023 vs. 2024)]

출처 : Menlovc (2024.11), 2024: The State of Generative AI in the Enterprise.

16) Yan Liu&He Wang (2024.8.), "Who on Earth Is Using Generative AI?", Policy Research Working paper 10870, World Bank Group.
17) Menlovc (2024.11.), 2024: The State of Generative AI in the Enterprise.

Ⅲ 소버린 AI 동향 관련 국내외 동향

1. 현재 AI의 핵심 기술 - 『LLM』

❶ 해외 소버린 LLM 주요 동향

o **(프랑스-미스트랄AI)** 자체 생성형 AI 모델 '르 챗'을 개발, 스페인어, 프랑스어, 독일어 등 EU 회원국 언어로 검색과 분석이 가능해 유럽 사용자들에게 더욱 적합한 분석 결과 제공[18)]

- 미스트랄AI는 2023년 4월 구글 딥마인드와 메타 출신 연구원들이 설립한 AI 스타트업으로 설립 4주만에 LLM 출시 계획을 발표
- 현재 기업가치 20억 유로(2조 9,500억 원)으로 평가받는 유니콘 기업으로 엔비디아, 삼성전자, 세일즈포스 등이 투자
- MS와 파트너쉽을 통해 클라우드 플랫폼 애저를 활용하는 한편, 이를 통한 비즈니스 확장의 기회를 모색하고 있음[19)]

[프랑스 - 미스트랄AI의 Mixtral 시리즈 특징]

구분	특징
모델명	Mixtral 7B, Mixtral 8×7B, Mixtral 8×22B, NeMo, Small/Medium/Large/Large 2
오픈소스 여부	오픈소스 모델로 제공
파라미터 크기	Mixtral 7B, Mixtral 8×7B, Mixtral 8×22B, NeMo (120억)
지원 언어	영어, 프랑스어, 독일어, 스페인어, 이탈리아어 등
주요 특징	LLM 글로벌 평가 2위 기록, GPT-4 터보 대비 20% 저렴한 비용, 다양한 언어 지원

출처 : 한국소프트웨어저작권협회(2027.7.25.), "소버린 AI의 국제적 동향"

o **(핀란드-사일로AI)** 유럽 최대 민간 인공지능연구소 사일로 AI(Silo AI)는 북유럽 언어 기반 LLM인 'Poro(포로)'와 '바이킹(Viking)'을 개발하여 공개, 미국의 반도체 기업 AMD가 AI 솔루션의 신속한 구축 및 확산을 위해 사일로 AI를 인수함('24.7월)[20)]

[핀란드 - 사일로AI의 LLM 모델 포로(Poro)와 바이킹(Viking) 특징]

구분	특징	특징
모델명	Poro	바이킹
오픈소스 여부	오픈소스 모델 제공	오픈소스 모델 제공
파라미터 크기	포로 34B(342억)	바이킹 7B, 바이킹 13B, 바이킹 33B
지원 언어	영어, 핀란드어 등 24개 유럽 언어	핀란드어, 아이슬란드어, 노르웨이어, 스웨덴어 등 북유럽 언어 특화
주요 특징	영어, 핀란드어 등 유럽 언어와 파이썬, 자바 등 프로그래밍 언어 21조 개 토큰으로 학습	2조 개의 토큰 데이터셋으로 훈련, 최대 4천 토큰의 커넥스트창 지원

출처 : AI 타임즈 (2023.11.15.), "EU도 '소버린 AI'.. 유럽언어 전문 LLM 오픈소스 공개"

18) 한국소프트웨어저작권협회 (2027.7.25.), "소버린 AI의 국제적 동향", SW·저작권 동향리포트(제2024-14호).
19) 연합뉴스 (2024.2.27.), "MS, 10개월 된 '유럽판' 오픈 AI에 투자...파트너쉽 체결".
20) 인공지능신문 (2024.7.11.), "AMD, 유럽 최대 민간 인공지능연구소 '사일로AI', 899억 원에 인수...엔터프라이즈 AI 솔루션, 전 세계 확장 위해".

o **(중국-문샷AI)** 스타트업 '문샷 AI'는 중국어 문장 처리에 특화된 챗봇 '키미'를 개발. 문샷AI는 알리바바가 36%의 지분을 보유하고 있으며 중국 정부의 전폭적인 지지를 받고 있음

- 문샷AI의 챗봇 키미(Kimi)는 한번에 한자 200만개를 처리할 수 있고 10분 이내에 신속한 학습이 가능하며 금융/법률/과학 분야의 장문 문서를 분석하고 요약하는 능력을 갖춤

[중국 - 문샷 AI의 LLM 모델 Kimi 특징]

구분	특징
모델명	Kimi (키미)
지원 언어	중국어
주요 특징	한자 200만 개를 한 번에 처리 가능, 10분 이내 신속한 학습 전문가 수준의 작문, 금융, 법률, 과학 등 분야의 장문 문서 분석 및 요약 가능
서비스 형태	기본 챗봇 'Kimi', 장문 글쓰기, 쇼핑, 소셜 엔터테인먼트 기능이 추가된 'Kimi+'
주요 특징	21조 개 토큰으로 학습 - 블룸(BLOOM) 아키텍처 사용 - 북유럽 언어에 특화

출처 : 중앙일보(2024.6.11), "대기업 제쳤다, 화제성 1위 중국판 챗GPT 만든 AI 천재" 등 언론자료 정리

o **(일본-후지쯔/NTT)** 후지쯔(Fujitsu)는 심층학습 프레임워크를 개발하여 트랜스포머 성능을 개선하는 연구를 통해 산학연이 공동으로 일본어에 최적화된 13B 모델을 개발하고 공개하였으며[21], NTT는 일본어 처리 능력을 갖춘 '츠즈미' 제공 시작함

- 후지츠(Fujitsu)는 Fugaku-LLM 동경공업대학교 총괄, 도후쿠대학, 후지쯔, 이화학연구소, 나고야대학교, 사이버에이전트, Kotoba technologies에서 공동 개발
- NTT는 40여년간 축적해온 자연어처리 기술에 기반하여 우수한 일본어 처리능력과 멀티모달 대응이 가능한 소형 경량된 LLM 'tsuzumi'을 발표함(2023.11)[22],

[일본 후지쯔(Fugaku)와 NTT(tsuzumi)의 LLM 모델 특징]

구분	Fujitsu 등 산-학-연 공동 개발	NTT
모델명	Fugaku-LLM	tsuzumi (경량판/초경량판)
오픈소스 여부	오픈소스 모델로 제공 (연구 및 상업목적 이용 가능)	상업용 서비스로 제공 (오픈소스 아님)
파라미터 크기	13B	7B(70억), 0.6B(6억)
지원 언어	일본어, 영어 4,000억 토큰 학습 - Japanese MT-Bench에서 일본산 오픈 모델 중 최고 성능	일본어, 영어 (향후 중국어, 한국어, 프랑스어, 독일어 등 확대 예정)
주요 특징	일본 산학연이 공동 개발,	우수한 일본어 처리능력/ 소형 경량화 문서그래픽, 청각 등을 통합하는 멀티모달 대응
컴퓨팅 인프라	이화학연구소의 슈퍼컴퓨터 Fugaku를 이용하여 학습 - 후지쯔 국산 CPU 활용	GPU 1~2개에서 고속 추론 가능

출처 : 후지쯔 홈페이지 언론보도, NTT 언론보도 자료 정리

21) 후지쯔 홈페이지 언론보도 (2024.5.10.), "슈퍼컴퓨터 '후다케'에서 배운 대규모 언어모델 '후가쿠-LLM'공개".
22) NTT R&D (2024.8.26.), "NTT의 LLM 'tsuzumi' : 그래픽 문서 이해 가능".

o **(일본-사카나AI/파나소닉)**

- AI 스타트업 사카나 AI는 다양한 오픈소스 모델을 진화적으로 병합하여 최적의 모델을 자동으로 생성하는 기술을 개발하여, 일본어와 수학에 특화된 언어모델(7B) 및 비전-언어(VLM) 모델, 이미지 생성 모델, AI 사이언티스트 등을 출시함[23)]
- 파나소닉이 AI스타트업 스톡마크(Stockmark)가 개발한 인스트럭트 모델에 자사의 기업 데이터를 학습시켜 일본어에 특화된 Pansonic-LLM-100b 개발 계획을 발표(7.2)하였는데 업무용 LLM으로는 일본 최대 규모가 될 전망[24)]

[일본 사카나AI와 파나소닉 LLM 모델 특징]

구분	AI 스타트업 사카나 AI	파나소닉-스톡마크
모델명	EvoLLM-JP, EvoVLM-JP, EvoSDXL-JP	Panasonic-LLM-100b
오픈소스 여부	오픈소스 공개	Stockmark-LLM-100b는 오픈소스
파라미터 크기	7B	100B
지원 언어	일본어	일본어
주요 특징	여러 LLM의 장점을 결합한 진화적 모델	일본 경제산업성 등의 지원을 받아 stockmark가 개발한 Stockmark-LLM-100b 인스트럭트 모델에 파나소닉HD 기업데이터 학습. 멀티모달 기반 모델 통합
컴퓨팅 인프라	엔비디아(최대 투자자) 플랫폼 활용	정보 없음

출처 : 후지쯔 홈페이지 언론보도, NTT 언론보도 자료 정리

o **(인도-크루트림/지오인포컴)** AI 스타트업 크루트림은 인도어를 학습한 LLM '크루트림' 공개, 힌디어뿐 아니라 타밀어, 텔루구어 등 10가지 이상의 현지어 지원

- 크루크림(Krutrim)은 2023년 4월 설립된 기업으로 힌디어, 칸나다어 등 다양한 인도 지역 언어를 이해하고 생성할 수 있는 모델을 개발 중, 2024.1월 기업가치 10억 달러를 인정받은 인도 최초의 AI 유니콘 기업이 됨. 음성 기반 AI 비서 및 AI 컴퓨팅에 최적화된 칩개발도 추진[25)]
- 인도의 대기업 릴라이언스 인더스트리(Reliance Industry)는 엔비디아와 협력하여 인도에 AI 슈퍼컴퓨터를 설치하고 힌디어 LLM을 개발하고, 릴라이언스의 통신사인 지오 인포컴(Reliance Jio Infocomm)에서 다양한 AI 서비스를 개발할 계획[26)]

  * 릴라이언스 인더스트리(무케시 D.암바니 회장)는 시가총액 기준 인도 최대의 기업으로 석유, 에너지, 유통, 통신 등을 주요 사업으로 하며 최근 통신사를 중심으로 AI로 확장

23) AI타임스 (2024.3.24), "'수백세대 걸친 모델 교배로 성능 고도화'... 사카나, 병합 진화 모델 공개".
24) 파나소닉 홈페이지 언론보도 (2024.7.2.), "파나소닉 HD와 스톡마크, 국내 최대 규모(1,000억 파라미터)의 독자 일본어 LLM「Panasonic-LLM-100b」개발로 협업".
25) 국민일보 (2024.2.2.), "인도 최초 LLM AI 스타트업 '크루트림' 유니콘 등극".
26) https://www.kiep.go.kr/aif/newsBriefDetail.es?brdctsNo=353616&mid=a30100000000&systemcode=02.

[인도 - 크루트림의 LLM 모델 특징]

구분	크루트림(AI스타트업)	지오인포컴-봄베이(인도공과대학교)
모델명	Krutrim	BharatGPT
오픈소스 여부	오픈소스로 제공	비공개 (상업적 활용 중심)
파라미터 크기	2028년까지 10조 개 이상 목표	-
지원 언어	20개 언어 이해, 힌디어/타밀어/텔루구어 등 10개 언어로 텍스트 생성	힌디어, 타밀어 등 다국어 지원
주요 특징	2조 개 이상의 토큰으로 훈련 - 음성기반 대화형 AI 챗봇 - 힌글리쉬 지원	음성-텍스트 챗봇으로 인도인의 문해력 문제를 도울 수 있음을 강조함
컴퓨팅 인프라	Krutrim Cloud 인프라 사용	모기업 릴라이언스 인더스트리가 엔비디아와 협력하여 설치한 AI 슈퍼컴 사용

출처 : 국민일보(2024.02.02.), "인도 최초 LLM AI 스타트업 '크루트림' 유니콘 등극" 등 언론보도 자료 정리
Time(2024.3.14.), "India Is Emerging as a Key Player in the Global AI Race" 등 언론보도 자료 정리

❷ 국내 소버린 LLM 개발 현황

o **(네이버)** 2021년 초 초거대 AI '하이퍼클로바' 공개, 지속적 성능 개선으로 후속 모델 '하이퍼클로바X' 출시, 자체 개발한 LLM을 바탕으로 한국어뿐 아니라 한국의 사회·문화 맥락에 대한 이해도가 높다는 평가

- 하이퍼클로바X(2023.8.24) 이후, 시각정보 처리 능력을 업데이트하여 대형멀티모달모델(LMM)으로 업그레이드하고, 이미지 인식-처리가 가능한 대화형 AI 에이전트 기능도 제공(2024.8.22)

  * 네이버는 최근 사우디아람코와 제휴해 중동 맞춤형 AI와 클라우드 구축에 참여하는 한편, 필리핀 인터넷 서비스 업체 '컨버지'와 함께 필리핀 클라우드 및 스마트 시티 시장에 진출하는 등 국제 협력 확대를 통해 해외 시장 진출 추진[27)]

[한국 - 네이버의 LLM 모델 특징]

구분	특징
모델명	하이퍼클로바X, 하이퍼클로바X대시(HCX-DASH)
오픈소스 여부	오픈소스 아님
파라미터 크기	82B (820억개)
지원 언어	한국어, 영어, 일본어, 아랍어, 힌디어, 베트남어 등
주요 특징	• 한국어, 일반 상식, 수학, 코딩 부문에서 우수한 성능 - 높은 번역 정확도 • LMM 이미지 및 음성처리가 가능한 LMM으로 고도화하고, • 멀티모달 대화형 AI 에이전트 기능을 '클로바X'에서 제공 • 레드티밍(Red-teaming)을 통한 안전성 검증, • 네이버 클라우드 업무협업 플랫폼 네이버웍스 협업 툴에 AI 적용하여 서비스
컴퓨팅 인프라	네이버 클라우드 기반

출처 : 지디넷코리아(2024.7.21.), AI타임즈(2024.8.22.), "네이버, '하이퍼클로바X' LMM으로 고도화...AI 음성비서로 챗봇 업그레이드" 등 언론보도 자료를 기반으로 정리

27) IT비즈뉴스 (2024.5.12.), "네이버클라우드, 필리핀 클라우드·스마트시티 시장 진출한다".

o **(LG)** LG AI 연구원이 개발한 엑사원 3.0은 한글과 영어를 지원하는 성능과 경제성을 모두 추구한 모델로, 코딩, 수학, 실제 사용성 등 13개 벤치마크에서 최상위 성능 기록

- 엑사원 2.0은 엑사원 유니버스(전문가용 대화형AI), 엑사원 디스커버리(신소재,신물질, 신약개발), 엑사원 아틀리에(창의적 이미지 생성)의 다양한 전문 서비스로 제공

[한국 - LG의 LLM 모델 특징]

구분	특징
모델명	엑사원 3.0
오픈소스 여부	경량 모델(7.8B 파라미터)을 연구 목적으로 오픈소스로 공개
파라미터 크기	7.8B (78억 개)
지원 언어	한국어, 영어
주요 특징	• 성능과 경제성 향상: 이전 모델 대비 추론 처리 시간 56% 단축, 메모리 사용량 35% 감소, 구동 비용 72% 절감 • 전문 분야 데이터 학습: 특허, 소프트웨어 코드, 수학, 화학 등 6,000만 건 이상의 전문 데이터 학습 • 다양한 모델 크기 제공: 초경량 모델부터 고성능 모델까지 용도에 따라 설계
컴퓨팅 인프라	LG AI 연구원 자체 인프라 활용 추정

출처 : LG AI 연구원 블로그 및 LG 보도자료 기반 정리

o **(KT)** MS와 AI와 클라우드 관련 파트너쉽(2029년까지 5개년)을 맺고 '협력적 소버린 모델'을 통해 데이터 주권과 AI 주권 확보 추진(2024.10.1.)

- 디지털 자주권 확보를 목표로 자체 초거대 AI '믿음'을 개발하는 한편, 멀티 LLM 전략으로 글로벌 LLM과 특화된 산업별 AI를 병행하여 활용하고 있음

  * 태국 IT기업 자스민 테크놀로지 솔루션(Jasmine Technology Solutions Public Company Limited, 이하 JTS)과 생성형 AI 플랫폼 개발과 도입을 위한 파트너십을 체결하고 동남아시아 진출 추진

❸ 소버린 LLM 동향 분석의 시사점

□ 소버린 AI 기업의 유형별 특징

o **(자국 AI 스타트업의 급성장)** '프랑스의 오픈AI', '인도의 오픈AI', '중국의 오픈AI' 등 각국의 오픈AI를 표방하는 AI 전문 스타트업들이 천재적인 CEO와, 대기업과 정부의 적극적인 투자를 받으며 단기간에 급성장하는 경우가 많음

- 구글 딥마인드, 메타AI 등 AI 전문 빅테크 출신의 CEO인 경우가 많으며, 우수한 성능의 모델을 MS 및 자국 클라우드 기반으로 이를 학습시키고, 서비스화 하고 있음

- '천재적인 AI 인재의 우수한 기술력'+ '스타트업의 신속하고 유연한 기업 환경' + '정부와 대기업의 적극적인 자금 지원' + '클라우드 기반의 풍부한 컴퓨팅 자원을 신속하게 확보' ==> AI 유니콘 등극, 경쟁력 있는 자국 LLM 확보

o **(AI 기업으로 변신하는 ICT 대기업)** 기존의 ICT 기업, 통신서비스 기업들이 LLM 모델을 개발하여 자사의 서비스에 적용하며, AI 기업으로 변신하고 있음

- 일본의 후지쯔, NTT, 한국의 네이버, 카카오, 인도의 지오인포컴 등의 기업들은 자사의 서비스에 인공지능을 적용하여 고객들에게 차별화된 서비스 제공 추진

[소버린 AI 기업의 유형별 특징 정리]

유연하고 빠르게 성장하여 유니콘이 된 **AI 스타트업**	• 사례 : 프랑스-미스트랄AI, 중국-문샷AI, 일본-사카나AI, 인도-크루트림 • 핵심역량 : **빅테크 출신 천재형 CEO**와 우수한 연구진, 유연성 • 자원 : 정부와 빅테크의 **대규모 투자**, ▌컴퓨팅: **빅테크 클라우드 이용** • 비즈니스 : 아직 수익원 불분명, 유료화 또는 기업 인수
AI 기업으로 변신하는 **ICT 대기업**	• 사례 : 후지쯔(ICT), NTT(통신), 지오인포컴(인도 통신사), 네이버 등 • 핵심역량 : **기존 ICT 서비스/고객, 클라우드, 데이터** • 비즈니스 : AI를 적용한 ICT 서비스 제공, 해외 진출

출처 : 저자 작성

□ 소버린 AI의 특징 : '차별화'와 함께 '확장적' 소버린 추진

o 빅테크 의존도를 줄이고 독립적인 AI 및 클라우드 인프라 구축을 목표로 하되, 부족한 기술 및 자원을 동맹국 및 파트너 기업으로부터 조달하고, 해외 주변국으로의 시장 확장을 고려하여 협력 기반 소버린을 추진하고 있음

o **(종합 1)** 소버린 AI는 빅테크 기업 AI에 대한 '차별화' 전략의 일환으로, 언어/문화/지역적 특성으로 차별화하는 한편, 기업용 IT 솔루션으로 '맞춤화'하는 전략을 병행하는 것으로 볼 수 있음

- (현지화·지역 특화) 지역별 언어와 문화, 법제도적 요구사항에 맞는 데이터 학습 및 서비스 제공을 추진하며, 이를 위해 현지 기업과 파트너쉽 추진
- (B2B 서비스 제공) 산업용 또는 기업용으로 클라우드 서비스 및 기업용 솔루션과 결합하여 제공되면서 수익성 확보 추진

o **(종합 2)** 수익성과 경제성을 확보하기 위해, 기술적으로는 '멀티모달 및 경량화'로 진화시키는 한편, 지역 또는 적용 서비스의 확장을 추진하고 있음

- (멀티모달 및 경량화 추진) 언어, 이미지, 영상 등 멀티모달 데이터를 처리할 수 있는 AI를 개발하는 한편, 경량화된 모델을 통해 비용 절감 및 다양한 하드웨어 환경 적용 확장
- (다국어 지원) 대부분의 모델이 자국어 및 주변국 언어에 대한 강점을 확보하고 자국뿐만 아니라 주변국으로 시장 확장 시도

2. 전략 자산화에 대비해야 할 핵심 분야 - 『국방 AI』

❶ 주요국 국방 AI 동향 및 특징

□ **미국**

o **(중앙정보국 : CIA)** ChatGPT 스타일의 의사결정지원 AI 도구를 개발하여 정보 수집 및 분석 자동화, 외국어 문서의 실시간 번역 및 분석, 대규모 데이터에서 의미 있는 인텔리젼스 추출, 패턴인식을 통한 위협 탐지 등에 활용[28)]

o **(국방부 : DoD)** 미국방부의 최고 디지털 및 인공지능 사무소 CDAO(Chief Digital and Artificial Intelligence Office)가 테스크포스 리마(Lima)를 설립(2023.8.10.)하고, 국방부 전반에 걸쳐 생성형 AI와 LLM을 활용하여 전투 시뮬레이션을 생성하고 전투 계획을 개선하는 등에 활용 추진[29)]

o **(국방전문기업 : 팔란티어 테크놀로지)** CIA, FGI 등의 데이터를 위탁받아 AI를 개발하는 기업으로, 미국 국방부의 최고 수준 보안 인증인 Impact Level 6(IL6)을 획득하고, 국방부용 API 플랫폼을 개발하고 LLM 기반 솔루션을 제공[30)]

- 국방부에 아마존 및 앤트로픽과 협력하여 클로드 3 및 3.5 모델을 AWS에서 제공, 다양한 AI 기능을 통합하여 국방 및 정보기관의 특수 업무 및 의사결정 지원을 위해 제공
- 팔란티어는 페이팔과 오픈AI의 공동창업자 피터 틸이 설립한 기업으로, 세계 각국의 정보기관들을 대상으로, 도시나 국가 단위의 범죄나 테러를 막기 위한 팔란티어 고담, 경영진의 위기 감지를 위한 팔란티어 파운드리 등을 제공하고 있음
- 미국 국방부뿐만 아니라 한국군 AI 시장 진출도 추진하면서 국내 국방 제안 발표회에서 한국군을 위한 다양한 AI 기반 다영역 작전·항공·정보 작전 솔루션을 제안하고 있으며[31)], LIG넥스원과 '미래 무기체계 빅데이터 플랫폼 활용을 위한 업무협약'을 체결하고 정찰용 무인수상정, 초소형 영상 레이더 위성 등을 위해 AI 알고리즘을 제공할 계획

o **(AI 기업 : 메타)** 그간 메타는 라마 모델을 공개하면서도 군사, 전쟁, 핵 관련 산업, 스파이 활동 등에는 사용할 수 없다는 제한 규정을 두었으나, 미국 국방 기관 및 방산 관련 민간 업체들에게 라마 모델을 제공한다고 발표함(2024.11.04.)[32)]

- 록히드마틴과 레이도스 등 방산업체와 MS, AWS, 오라클 등 국방기관 SW 제공업체 포함하며, 기밀정보 동맹체인 캐나다, 호주, 뉴질랜드, 영국의 .관련 기관 및 계약업체에 제공

---

28) 한국경제 (2023.8.15.), https://www.hankyung.com/article/2023081521041.
29) 미국 국방부 홈페이지 (2023.8.10.), "DOD Announces Establishment of Generative AI Task Force".
30) Converge (2024.11.26.), "Anthropic, Palantir, and AWS Tune Claude AI Models for U.S. Defense Operations".
31) 비즈한국 (2024.4.3.), "글로벌 AI 기업 '팔란티어', 한국 방산시장 진출 타진".
32) Policy Economy (2024.11.5.), "군사적 이용 안 된다더니, 미 국방 기관에 AI 문 활짝 연 메타".

- 중국 개발자들이 라마2를 기반으로 군사용 AI 모델 ChatBIT를 개발하여 인민해방군의 전략 기획과 지휘 결정에 활용하려는 움직임이 포착되었다는 소식이 결정된 후 취해진 조치로 이에 대응하기 위한 조치로 풀이되고 있음
- 이로 인해 오픈소스 AI가 국가 안보에 위험이 될 수 있다는 경계가 높아지고 있음

o **(AI 기업 : 오픈AI)** 군사적 사용을 전면 금지하였던 오픈AI가 정책을 변경하여 미국 방위기술 기업인 안두릴(Anduril)과 전장분야 AI 개발에 협력하기로 계약함(2024.12.4.)[33]

□ 중국

o 중국은 AI를 통해 미국과의 군사력 격차를 줄이고 비대칭 전력을 확보하려는 전략을 추진하고 있으며, 이를 위해 민간 기술을 군사 분야에 이전하려하고 국방부에서 군사용 AI 개발 중

o **(군사용 AI 모델 개발)** 최근 중국 인민해방군(PLA) 전략지원부대 정보공정대학 연구진은 챗GPT와 유사한 거대 언어 모델(LLM)을 활용하여 군사용 AI를 개발

- 군사용 AI 사례로 바이두의 '어니'와 아이플라이텍의 '스파크' 등으로 군사용 시뮬레이션 수행한 연구 결과가 논문으로 발표됨

o **(오픈소스 모델 사용)** 중국 인민해방군(PLA)과 관련된 최고 수준의 연구기관들이 메타의 라마 모델이 군사적 활용을 제한했음에도 불구하고 군사적 용도에 맞게 개발한 것으로 보도됨[34]

- 중국의 군사과학원(AMS)을 포함한 3곳의 연구원 6명이 2024년 6월에 발표한 AI 논문을 통해 군사 분야의 대화 및 질문 응답 작업에 최적화된 라마 기반 챗봇 '챗비트(ChatBIT)'를 개발했다고 공개함
- 라마 13B에 사용자 지정 매개변수를 추가하고 군사 중심의 AI도구를 구축하였으며, 10만 건의 군사 대화 기록으로 미세조정하여 챗봇이 군사 관련 질문에 답할 수 있도록 함

□ 일본

o 일본 방위성은 저출산·고령화의 진행 속에서도 자위대 체제를 유지하기 위해 AI로 「대원의 부담 경감과 인력·업무의 효율화」로 설정하고, 7개 분야에서 중점 활용 추진 방침을 제시(2024.7월)

- ① 목표의 탐지·식별, ② SMS 등의 방대한 정보의 수집·분석, ③ 지휘관의 판단 지원, ④ 장비품의 수요 예측 등 후방지원의 효율화, ⑤ 무인기 등의 제어, ⑥ 사이버 보안 능력의 향상, ⑦ 사무작업의 효율화

33) MIT Technology Review (2024.12.4.), "OpenAI 's new defense contract completes its military pivot".
34) Asia Times (2024.11.2.), "China's People's Liberation Army weaponizing Metaj's AI".

❷ 국내 국방 AI 동향

o **(국방부-국방전략기술선정)** '2023~2037 국방과학기술혁신 기본계획'을 확정(23.4)하고, AI·첨단과학기술 기반 구축 및 국가적 차원의 국방연구개발 역량 확보를 위한 10대 분야 30개 국방전략기술을 선정함 : 인공지능 분야 ①지능형 전장인식/판단, ②지능형 통합 지휘결심, ③스마트 전력지원, ④국방 AI 플랫폼 기술이 포함됨[35)]

* 미국도 국방과학기술 14개 핵심 분야를 선정하였는데, 자율 인공지능, 휴먼머신 인터페이스 및 알고리즘·데이터·컴퓨팅에 기반한 AI와 유무인 협업 플랫폼을 포함한 사항은 유사함

- ChatGPT 등 LLM에 국방 데이터를 융합해 국방 GPT를 개발하여 군수, 전술, 무기체계 개발, 교범 등을 학습한 국방 특화 LLM을 개발하여 각 군 지휘통제 시스템 혁신 및 차세대 전술 무전기 개발에 활용 가능[36)]
- 한국의 전차, 함정 등 각종 무기체계에서 생성되는 데이터는 수집 기술과 저장공간 부재로 폐기되는 경우가 많고 품질 수준도 낮아 양질의 데이터 확보가 어려운 상황
- 직접 통제가 가능한 초거대 국방 AI를 군이 자체 개발해야 할 필요성도 주장되고 있음

o **(국방부-생성형AI 서비스)** 국방 생성형AI 서비스를 국방부 직원 대상으로 제공(2024.7.3.), 소형언어모델 기반 40만 건 군자료 학습, 군사용어 특화 번역, 군 내부규정 질의응답, 문서 요약 등 서비스 제공

o **(국방과학연구소 ADD)** 국방부는 대전 국방과학연구소(ADD) 내에 '국방AI 센터'를 설립하고(2024.4), AI 기반의 유무인 통합체계 개발과 전장상황 인식향상 등의 개발을 주도하며, 민간 AI 기술의 군사적 적용을 추진

* 국방AI센터는 기술전문성을 가진 민간연구원과 군사전문성을 갖춘 현역 군인을 합해 약 110여 명으로 구성됐으며, △인공지능 기반 유무인복합체계·전장상황인식 등 인공지능 관련 핵심기술 개발 △군 인공지능 소요기획 지원 및 기술 기획 △민간 인공지능 기술의 군 적용을 위한 산·학·연 협업 강화 등의 역할을 수행할 예정[37)]

- 드론 군집 운용을 위한 강화학습 기반의 군집 제어 알고리즘을 개발하고 있으며, 수백 대의 드론이 대형 유지, 장애물 회피, 공격 대응 등의 임무 수행이 가능하도록 시뮬레이션 환경에서 학습 진행[38)]

o **(ETRI 국방안전융합연구본부)** AI 군참모 시스템 개발. 영상, 음성, 센서 데이터 등을 활용해 전장 상황을 실시간으로 분석하고, 설명가능한 AI 기술을 적용하여 AI의 판단 근거를 제시할 수 있도록 개발

o **(방위사업청)** AI 빅데이터 기반 무기체계 및 부품 가격이 적절한지 판별하기 위한 국방통합원가시스템을 구축(2024.7)하고, AI 문자인식 기술을 활용해 무기체계 부품의 가격

35) 대한민국 정책브리핑 (2023.4.19.), "미래전장을 유도할 과학기술 강군 건설... 국방과학기술혁신 발표".
36) 한국경제(2023.8.15.), "미래전쟁은 데이터 싸움...국방GPT 도입해야".
37) 보안뉴스(2024.4.2.), "국방부, 국방AI센터 출범으로 '과학기술 강군' 위한 의지 확인".
38) 헬로디디(2024.5.8.), "군집AI, AI군참모...AI 활용이 미래전 양상 바꿔".

추이 분석 등에 활용[39)]

o **(한국과학기술정보연구원(KISTI))** 과학LLM 고니(8B, 13B)를 육군에 적용('24.8)[40)]

- KISTI는 2023년 4월 육군 군수사령부와 업무협약을 체결하고 국방 분야의 디지털 전환을 위한 협력을 강화하고 있음

o **(코난테크놀로지)** 사이버&네트워크 연구원과 협력하여 '군인사업무분야 초거대 언어모델 적용 방안 연구'를 시작하고 인사정책 수립, 육군 인사 사령부 업무, 기록물 관리 등에 LLM을 활용하는 방안 연구(2024.8.5.)

[국내외 국방 AI 기업 주요 내용 요약]

국가	프로젝트/기업	주요 내용	특징
미국	팔란티어 테크놀로지스	AIP(Artificial Intelligence Platform for Defense): LLM 기반 군사용 AI 플랫폼. 전장 데이터 실시간 분석 및 지휘관 의사결정 지원	모든 네트워크에서 LLM 및 AI 배치 가능 정보접근 권한 세부 통제 Azure, OpenAI와 통합으로 클라우드기반 AI서비스 강화
	미 국방부(CDAO)	태스크포스 리마(Task Force Lima): 생성형 AI 도구 분석 및 통합 주도. AI바이어스바운티프로그램:LLM의 편향성과 위험식별 및 통제	LLM의 위험 요소 식별 국방부내 생성형AI 도구 적용
	스케일 AI	헤르메스(Hermes): 군사 작전 계획에 특화된 LLM 전장활동과 투자 설계 지원	훈련된 사용자 없이 사용 시 환각 및 편향 위험 군사 계획 수립 지원
중국	바이두 어니, 아이플라이텍 스파크	민간 LLM을 활용하여 군사 시뮬레이션과 작전 수립에 적용	기존 군사 AI와 LLM 연결 전투시뮬레이션 성공
일본	방위성 (AI 활용 기본 방침)	7대 중점 분야 설정: 목표 탐지·식별, 지휘관 판단 지원, 무인기 제어, 사이버 보안, 사무작업 효율화 등	AI 활용으로 자위대 인력 효율화 목표
한국	국방생성형 AI	군사용 LLM 기반 AI 40만 건의 군 자료 학습 군사용어 번역, 질의응답, 문서요약 등 제공	국방 내부망에서만 작동 보안성 강화 및 실시간 학습 제한
	코난테크놀로지	군 인사업무를 위한 LLM 연구	인사 정책 및 기록물 관리 지원
	ETRI (AI 군참모)	정찰, 상황 인지, 방책 수립 등 지원	XAI 기술 중점 멀티모달 데이터 활용

출처 : 저자 작성

39) 연합뉴스 (2024.7.24.), "무기체계 원가 검증 쉬워진다...AI 기반 국방통합원가시스템구축".
40) 머니투데이 (2024.2.20.), "KISTI, 거대언어모델 '고니'로 AI 육군 양성 지원한다".

❸ 국방 AI R&D에 대한 시사점

o 국방에 AI가 도입되면서 군사 자산관리, 의사결정 지원 등이 효율화되고 있으며, 전장상 황에서 AI의 적용은 군사력 향상에 영향을 줄 것임

o 람다 모델을 중국에서 군사용으로 활용한 정황이 포착되고, 이에 대한 대응으로 미국 국방 관련 기관 및 관련 국가에 그간 금지되었던 람다 모델이 군사용으로 허용되면서 AI의 국방분야 활용은 더욱 가속화 될 것으로 보임

o 오픈소스 AI의 위험성에 대한 인식이 높아지게 되면서 자국 및 동맹국 이외의 국가에게는 기술 공개 및 공유가 금지될 것으로 전망됨

o 국방분야 경쟁력 확보에 활용 가능한 국방 특화 소버린 AI 자체 모델 확보가 시급함

3. AI 개발 역량 확보를 위한 핵심 자원 - 『컴퓨팅 인프라』

❶ 컴퓨팅 인프라 관련 동향

□ 주요 빅테크 기업의 컴퓨팅 인프라 투자 동향

o 클라우드 빅3 (MS, 아마존, 구글)은 자사의 AI 개발 및 클라우드 고객사에게 제공할 클라우드 서비스를 위해 각각 향후 1천억 달러~1천5백억 달러 규모의 컴퓨팅 인프라 투자를 계획하고 있음

- (Microsoft & OpenAI) 2024년 3월 발표한 '스타게이트 프로젝트'를 통해 2028년까지 1,000억 달러를 투자하여 세계 최대 AI 전용 데이터센터와 슈퍼컴퓨터를 구축할 계획
- (Amazon) 향후 15년 동안 데이터센터 구축에 1,500억 달러(약 208조 원)를 투자하겠다고 발표
- (Google) 2021년에 95억 달러 규모의 데이터센터 투자 계획을 발표했으며, 이미 2017~2021년 동안 370억 달러를 미국 26개 주의 데이터센터 구축에 투자

o 메타, 테슬라, 소프트뱅크 등은 자사 AI 개발 및 서비스 운영을 위해 2024년 수백억 달러 투자

- (Meta) 라마 모델 훈련을 위해, 2024년 클러스터당 24,576개의 H100 GPU를 포함한 세계 최대 규모의 AI 컴퓨팅 인프라를 공개하고, 연내 35만 개의 H100을 추가하는 등 AI 프로젝트에 400억 달러를 투자할 예정
- (테슬라) AI 훈련을 위한 컴퓨팅 인프라 구축에 2024년에만 약 100억 달러(약 14조 원)을 투자하여, 슈퍼클러스터 '콜로서스', 텍사스 슈퍼컴 '코르텍스', 뉴욕 버팔로 슈퍼컴 '도조' 등에 H100을 추가로 구축
- (소프트뱅크) 2025년까지 생성형 AI 개발을 위한 컴퓨팅 인프라 강화에 1,500억 엔(약 9억 6,000만 달러)을 투자할 계획

□ GPUaaS 동향

o (개요) GPUaaS(GPU as a Service)는 클라우드 환경에서 GPU를 서비스로 제공하는 방식으로, AI 개발 도구 및 다양한 SaaS 서비스들과 함께 GPU를 필요한 만큼 활용할 수 있도록 제공하여 GPU(GPUaaS) 시장이 크게 성장하고 있음

- GPUaaS는 고가의 온프레미스 하드웨어를 구매하고 유지 관리할 필요 없이 복잡한 시뮬레이션, 데이터분석, AI 트레이닝 등 리소스 집약적인 프로젝트에 유연하게 GPU를 사용할 수 있음
- 생성형 AI 등장 이후 연산성능 요구가 기하급수적으로 증가함에 따라 강력한 GPU 필요. GPU와 같은 부품과 생성형 AI 플랫폼까지 구독 서비스로 출시

o (해외 사업자 현황) 기존 클라우드 서비스 기업 및 신규 기업들이 진입하고 있으며 미국 기업들이 주도적임

- Vultr(미국), Linode(미국), AWS(미국), Google(미국), Microsoft(미국), NVIDIA(미국), IBM(미국), Lambda Labs(미국), Hewlett Packard Enterprise Development LP(미국), CoreWeave(미국), OVH SAS(프랑스), Alibaba 클라우드(중국), 삼성(한국), 후지쯔(일본) 등

[해외 GPUaaS 주요 제공 기업의 특징]

기업	서비스 명칭	주요 특징
NVIDIA	NVIDIA GPU Cloud (NGC)	딥러닝 및 AI 작업에 최적화된 소프트웨어 스텍 제공 클라우드 및 온프레미스 GPU 리소스 통합 가능
AWS	Amazon Elastic GPU	EC2 인스턴스에 GPU 성능 추가하는 옵션제공 다양한 워크로드에 적합, 유연한 스케일링 가능
Google Cloud	Google Compute Engine (GCE)	NVIDIA Tesla GPU 지원 AI 및 ML 작업에 최적화 TPU 옵션 제공, 고성능 클러스터링 제공
Microsoft Azure	Azure N-Series	NVIDIA Tesla GPU 기반 AI 및 데이터분석 워크로드에 적합, 고급 네트워킹 옵션 제공
IBM Cloud	IBM Cloud GPU, 왓슨X.ai	NVIDA Tesla 및 V100 GPU 제공, IBM Watson과 통합 가능 AI 및 HPC 작업에 최적화 기업 서비스 개발 돕는 생성형 AI 플랫폼 '왓슨X.ai'
Oracle Cloud	Oracl Cloud Infrastructure(OCI)	NVIDIA GPU 기반, 유연한 가격 정책 제공 AI, 데이터 분석, 고성능 컴퓨터에 최적
Alibaba Cloud	Alibaba Elastic GPU Service	NVIDIA GPU 지원 대규모 데이터 처리 및 AI 작업에 적합 중국 내 다양한 데이터센터 위치
Lambda Laps	GPUaaS	AI에 특화된 GPUaaS wprhd

출처: 언론자료, 각사 홈페이지 등에서 작성

o (국내 사업자 현황) 국내에서도 기존 클라우드 서비스 기업, 통신사, SI 기업 및 신규 기업들이 진입하고 있음

[국내 GPUaaS 주요 제공 기업의 특징]

기업	서비스 명칭	GPU 모델	주요 특징	타겟시장
삼성SDS	삼성 클라우드 플랫폼(SCP) GPU 서비스	NVIDIA A100, V100, H100	AI/ML, 데이터분석에 최적화된 고성능 GPU 제공 한국내 데이터센터 운영, 프로젝트 용도 및 규모에 따른 유연한 구독	엔터프라이즈, 연구기관
KT 클라우드	AI SERV에 슬라이딩기술	-	GPU 0.2장 단위, 이용 시간단 과금 등 고객맞춤형 구독	스타트업
LG CNS	DAP GenAI 플랫폼	-	기업용 생성형 AI 서비스 개발 플랫폼 자체 서버나 클라우드에 DAP GenAI 설치후 구독	기업
SK텔레콤	SKT GPU클라우드 서비스	NVIDIA A100, T4, H100	AI/ML, 영상처리 등 다양한 워크로드 최적화 미국 '람다'와 파트너십으로 AI 클라우드 서비스 2024.12. 출시 예정	AI스타트업, 중소기업
카카오 클라우드	Beyond Compute Service	NVIDIA A100, T4	AI 특화(대규모 데이터세트에서 AI 학습), 고성능 그래픽 애플리케이션, 게임 등에 사용 BCS중 GPU 인스턴스 기반으로 운영	스타트업 (ex.스캐터랩, AI스타트업)
NHN 클라우드	-	NVIDIA H100	광주, 판교 및 평촌센터에 총99.5페타플롭스 연산능력의 AI GPU 설비 구축	공공기관
네이버클라우드	GPU Server	Tesla T40, V100, T4	하이퍼클로바X 연계 클라우드 서비스	
Gabia Cloud	-	NVIDIA H100, A100	공공기관 클라우드 서비스 중심	공공기관, 기업
Elice	Elice Cloud	NVIDIA A100, H100	GPU 서버팜 클라우드 인프라로 모듈형 데이터센터를 기반으로 구축 AI 연구 및 학습 관련 서비스 중심	KAIST 등 대학

출처: 언론자료, 각사 홈페이지 등에서 작성

□ 엔비디아의 AI 팩토리 구축 사업

o 엔비디아는 컴퓨터 제조업체들과 협력하여 AI 팩토리 및 데이터센터를 구축하는 사업을 전개하며 생태계 확장을 추진

- 엔비디아의 최신 GPU뿐만 아니라 소프트웨어 스텍, AI 워크로드, 네트워킹 인프라, 냉각 시스템에 이르는 인프라의 구성요소를 제공하며 오라클 등의 컴퓨팅 기업과 협력[41)]
- 베트남 정부[42)], 일본의 소프트뱅크, GMO 인터넷 그룹, Hihgreso, KDDI[43)], 대만의 폭스콘 등과 협력하여 엔비디아 AI 팩토리 구축[44)]

41) NVIDIA Newsroom (2024.6.2.), "Computer Industry Joins NVIDIA to Build AI Factories and Data Centers for the Next Industrial Revolution".
42) Reuters (2024.12.5.), "Nvidia to build AI research, data centres in Vietnam with govt".
43) ComputerWeekly.com (2024.11.15.), "Nvidia powers AI development in Japan and Indonesia".
44) NVIDIA Newsroom (2024.6.2.), "엔비디아, 'AI 공장' 설립 위해 폭스콘과 파트너십 구축".

❷ 컴퓨팅 인프라 구축 방식 비교

□ 직접 GPU 구매 및 구축

o (장점) ▷하드웨어 및 데이터에 대한 보안 및 주권 유지에 유리하고 ▷장기적으로 비용 절감 효과가 가장 큼

o (단점) ▷초기 투자 비용이 높고, ▷하드웨어 유지보수 및 관리 부담 크며, ▷기술 발전에 따른 장비 노후화의 위험을 부담해야 함

o (예시) 대한민국 정부, 클라우드 사업자, 메타 등

□ GPUaaS 이용

o (장점) ▷초기 투자 부담이 적음, ▷유연한 확장성 제공, ▷최신 기술에 대한 신속한 접근 가능

o (단점) ▷장기적으로 비용이 증가할 수 있음, ▷데이터 보안 및 프라이버시 우려. ▷서비스 제공자에 대한 의존성 증가.

o (예시) 국방부는 아마존, 마이크로소프트, 구글, 오라클과 약 12조 원 규모의 클라우드 서비스 계약을 체결

□ **엔비디아와 협력 :** 엔비디아의 AI factory 투자 및 엔비디아와 국가간의 협력 모델

o (장점) ▷최신 GPU 기술과 소프트웨어 스택 제공, ▷AI 워크로드에 최적화된 솔루션, ▷엔비디아의 글로벌 기술 지원 활용 가능

o (단점) ▷특정 기업 기술에 대한 종속 우려, ▷엔비디아의 가격 정책에 따른 비용 증가 위험. ▷클라우드 기반 솔루션 이용 시 데이터 주권 문제 발생 가능

o (예시) 대만의 폭스콘, 미국의 Dell

□ 클라우드 사업자의 AI 기업 지분투자 등 직접협력

o (방식) 클라우드 사업자가 AI 기업의 잠재력에 대해 지분투자를 단행하고, 이와 함께 AI 기업에게 컴퓨팅자원(GPU 클라우드, 데이터 센터 이용 등)을 제공하며, AI 기업은 클라우드 사업자의 인프라를 활용하여 (컴퓨팅 인프라 구축 비용/시간 없이) 모델 개발 및 훈련을 수행

o (클라우드 사업자) AI 기업을 통해 시장 점유율을 확대하고 AI 혁신이 가능해지나, 실패할 경우의 투자 리스크를 감당해야 하며 해당 AI 기업의 기술에 종속될 가능성 있음

- (장점) ▷대상 기업이 클라우드 인프라를 지속적으로 사용하게 되어 장기적인 고객 확보

가능, ▷(AI 혁신 접근성 강화) 유망 AI 기술에 초기 접근 가능하며 협력 기업의 성공을 통해 간접적으로 기술 트렌드 선점

- (단점) ▷투자한 AI 기업이 실패할 경우, 지분 손실 가능, ▷ 투자 기업이 특정 기술에 집중하면 클라우드 사업자도 기술적 방향성을 제한받을 수 있음

o (AI 사업자) 초기 투자부담을 줄이고 단기간에 최첨단 클라우드 인프라를 활용할 수 있으나 클라우드 기업에 대한 의존도가 높아지고 전략적 독립성이 저하될 수 있음

- (장점) ▷초기 AI 기업들이 컴퓨팅 비용 부담을 줄이고 연구개발에 집중 가능, ▷최첨단 GPU와 클라우드 인프라를 손쉽게 확보, ▷클라우드 사업자의 기술적 지원 및 협업을 통해 성장을 가속화
- (단점) ▷클라우드 사업자의 인프라와 생태계에 지나치게 의존하며 다른 클라우드 서비스로의 전환이나 협력이 어려워질 수 있음. ▷ 투자자가 경영 의사결정에 영향을 미칠 가능성 있음

[클라우드 사업자의 AI 기업 투자시 장단점]

구분	클라우드 사업자	AI 사업자
장점	✓ 시장 점유율 확대 ✓ AI 혁신 접근성 강화 ✓ 수익 창출	✓ 자금 확보 원활 ✓ 컴퓨팅 자원 접근성 향상 ✓ 파트너십 통한 기술 지원 및 시장 확장
단점	✓ 투자 리스크 ✓ 기술 종속 가능성	✓ 클라우드 의존 심화 ✓ 전략적 독립성 저하

출처 : 저자 작성

(예시 1) Microsoft와 OpenAI[45)]

o (지분투자) Microsoft는 OpenAI에 100억 달러 규모의 투자를 단행

o (컴퓨팅 자원 제공) Azure 클라우드 인프라를 OpenAI가 독점적으로 사용

o (협력 결과) OpenAI는 ChatGPT와 같은 혁신적인 모델을 개발했으며, Microsoft는 Azure 사용량 증가와 AI 기술 통합으로 이익을 얻음

(예시 2) 구글(Google)과 앤트로픽(Anthropic)[46)]

o (지분투자) Google이 앤트로픽(Generative AI 기업)에 3억 달러를 투자

o (컴퓨팅 자원 제공) 앤트로픽은 Google Cloud에서 자사 AI 모델을 훈련

o (협력 결과) Google은 자사 클라우드 점유율 확대 및 AI 기술 경쟁력 강화

(예시 3) Amazon과 Hugging Face

o (협력 및 투자) Amazon Web Services(AWS)는 Hugging Face와 파트너십을 맺고 지분투자

o (협력 결과) Hugging Face는 AWS 기반에서 AI 툴과 모델을 제공하며, Amazon은 AI 연구 및 생태계를 강화

45) 연합뉴스 (2023.1.24.), “MS, '챗GPT' 개발 오픈AI에 새로운 투자 발표…12조 원 규모”.
46) IT Times (2023.2.5.), “구글, 오픈AI•챗GPT 경쟁사 '앤트로픽'에 3억 달러 투자”.

Ⅳ 소버린 AI 확보를 위한 시사점

□ 소버린 AI 기업의 특징 요약

o **(자국 AI 스타트업의 급성장)** 프랑스, 인도, 중국 등 각국의 '오픈AI'를 표방하는 AI 전문 스타트업들이 단기간에 급성장하는 경우가 많음

- AI 스타트업의 급성장을 가능하게 한 요인은 ▷우수한 기술력과 AI 인재 확보, ▷정부와 대기업의 적극적인 자금 지원, ▷컴퓨팅 자원의 신속한 확보 등으로 판단됨

o **(AI 기업으로 변신하는 ICT 대기업)** 기존의 ICT 기업, 통신서비스 기업들이 LLM 모델을 개발하여 자사의 서비스에 적용하며, AI 기업으로 변신하고 있음

- 일본의 후지쯔, NTT, 한국의 네이버, 카카오, 인도의 지오인포컴 등의 기업들은 자사 서비스에 인공지능 기술을 적용하여 고객들에게 차별화된 서비스 제공 추진

o **(국방 AI 제공 기업)** 국방용 AI 제공을 위해 최고 수준의 보안 인증을 획득한 방산 기업들이 국방기관 및 AI 전문기업과 협력하는 경우가 많음

- 중국이 오픈소스AI를 국방에 활용하자, 메타와 오픈AI가 국방분야에 진출하게 됨
- 미국이 기밀정보 동맹체 이외에는 국방용 AI 기술을 제공하지 않을 것으로 알려지면서, 국내 기술력으로 강건한 국방AI 확보의 중요성은 더욱 높아질 전망

□ 핵심 이슈

o 소버린 AI는 소수의 빅테크 기업들이 주도하는 LLM 시장에 진입하고 시장에 안착하기 위한 기업들의 시장 전략이자, 각국 정부가 자국의 AI 주권을 확보하기 위한 정책적 목적의 이중적 성격을 보유하고 있음

o LLM의 특성상 결국 AI기업이 LLM 개발 및 운영의 주체가 되어야 하므로, 소버린 AI가 성공하기 위해서는 빅테크 기업의 AI 대비 차별적 경쟁력을 확보할 수 있어야 함

o **국가 AI R&D 관점에서 소버린 AI의 핵심 이슈는 아래와 같이 정리할 수 있음**

① LLM 모델의 진화 경쟁 - 멀티모달 AI로 고도화 경쟁

- (멀티모달과 경량화) 수익성과 경제성을 확보하기 위해, 기술적으로는 '멀티모달 및 경량화'로 진화시키는 한편, 지역 또는 특정 도메인 전용 서비스로 확장 필요

② 국방분야 경쟁력 확보에 활용 가능한 국방 특화 소버린 AI 자체 모델 확보 시급

- 오픈소스 AI의 위험성에 대한 인식이 높아지게 되면서 자국 및 동맹국 이외의 국가 기관에게는 기술 공개 및 공유가 금지될 것으로 전망으로 자체 국방 AI 모델 확보 시급

③ **AI 시장 선점를 위한 '속도전' : 컴퓨팅 자원 및 핵심 역량 확보를 위한 거래와 협력**

- (투자와 협력) 부족한 자원과 역량을 확보하기 위해 다양한 방식으로 서로 거래하고 협력하여 AI 시장에 진입하고 정착해야 함
- (컴퓨팅 확보) ▷직접구매, ▷GPUaaS, ▷엔비디아 협력, ▷클라우드 협력(지분투자) 중 전략적 선택 필요

□ 정부 출연연 AI R&D에 대한 기회와 위협 요인

o **(기회 요인)** AI의 진화 속도가 빨라지고 활용 범위가 확장되면서 AI 기술주권 확보, 즉 소버린 AI에 대한 관심이 높아지고 있으며, 정부에서도 투자를 늘리고 있음

- 국내 AI 전문기업들이 활용할 수 있는 고성능 멀티모달AI를 조기에 개발하고 국내 기업에게 공개할 경우 활용도가 높을 것으로 기대됨
- 특히, 민간에게 맡기기 어려운 국방 AI의 경우, AI 기술력과 국방분야 도메인 전문성을 모두 보유한 정부 출연연들이 개발하기 적합한 분야로 판단됨

o (위협 요인) 빅테크 기업 대비 국내 기업 및 정부의 예산은 매우 한정적이며, GPU 구매력도 매우 열악한 상황임

- 현재 글로벌 AI 생태계에서는 필요한 자원을 신속하게 확보하기 위해 경쟁과 협력이 활발하게 일어나고 있는데, 정부 출연연들은 경직된 정부 R&D 체계에 따라야 하므로 예산, 컴퓨팅 자원, 인력 확보 등에 제약이 많음

□ R&D에 대한 시사점

o (국내 AI 기업) 글로벌 사업자 대비 차별화가 가능한 전문AI 및 특화 AI 타겟시장 발굴이 필요

- 신속하게 경쟁력 있는 AI 기술을 확보하기 위해 출연연-대학-기업들간 적극적인 협력이 필요하며, 타겟시장 선정 시 내수시장과 해외 시장을 동시에 고려해야 함

o (정부) 속도감 있는 컴퓨팅 자원 확보와 국가 AI 역량 집결을 위한 중심체로서 역할을 해야하며, 해외 시장 개척을 위해 타겟 국가 정부와 협력의 중심이 되어야 함

o (정부 출연연) 해외 기업 및 민간 기업에 의존하지 않고 개발 및 운영 가능한 멀티모달 AI 모델과 국방 AI 모델 개발이 필요함

- 컴퓨팅 자원의 조속한 확보 및 출연연 개발 기술의 조기 상용화(서비스 제공)을 위해 민간AI 기업과 적극적인 협력체계 마련이 필요
- 국방 AI 개발을 위해 국방부 및 방산 전문기업과 협업을 통해 안전하고 경쟁력 있는 국방AI 기술의 조기 확보 추진 필요

참고문헌

국내 자료

국민일보 (2024.2.2.), “인도 최초 LLM AI 스타트업 ‘크루트림’ 유니콘 등극”.

대한민국 정책브리핑 (2023.4.19.), 미래전장을 유도할 과학기술 강군 건설...국방과학기술혁신 발표.

매일경제 (2024.11.27.), “왜 소버린AI 인가”, 매경춘추.

머니투데이 (2024.2.20.), “KISTI, 거대언어모델 ‘고니’로 AI 육군 양성 지원한다”.

보안뉴스 (2024.4.2.), “국방부, 국방AI센터 출범으로 ‘과학기술 강군’ 위한 의지 확인”.

봉강호 (2024.9.), 글로벌 초거대 AI 모델 현황 분석(2020~2023), SPRI 이슈리포트 IS-178. SPRI.

비즈한국 (2024.4.3.), “글로벌 AI 기업 ‘팔란티어’, 한국 방산시장 진출 타진”.

연합뉴스 (2024.7.24.), “무기체계 원가 검증 쉬워진다...AI 기반 국방통합원가시스템구축”.

연합뉴스 (2023.1.24.), “MS, ‘챗GPT’ 개발 오픈AI에 새로운 투자 발표…12조 원 규모”.

연합뉴스 (2024.2.27.), “MS, 10개월 된 ‘유럽판’ 오픈 AI에 투자...파트너쉽 체결”.

인공지능신문 (2024.7.11.), “AMD, 유럽 최대 민간 인공지능연구소 ‘사일로AI’, 899억 원에 인수... 엔터프라이즈 AI 솔루션, 전 세계 확장 위해”.

조선일보 (2024.10.31.), “바이든, 미국 최초 AI 안보각서 서명”.

중앙일보 (2024.6.11.), “대기업 제쳤다, 화제성 1위 중국판 챗GPT 만든 AI 천재”.

파나소닉 언론보도 (2024.7.2.) “파나소닉 HD와 스톡마크, 국내 최대 규모(1,000억 파라미터)의 독자 일본어 LLM「Panasonic-LLM-100b」개발로 협업”.

한국경제 (2023.8.15.), “미래 전쟁은 데이터 싸움...국방 GPT 도입해야”.

한국지능정보사회진흥원 (2024.6.28.), “글로벌 정부·민간 분야 AI 투자 동향 분석” 자료 재구성.

한국소프트웨어저작권협회 (2024.7.25.), 소버린 AI의 국제적 동향, SW·저작권 동향리포트(제2024-14호)

헬로디디 (2024.5.8.), “군집AI, AI군참모...AI 활용이 미래전 양상 바꿔”.

후지쯔 언론보도 (2024.5.10.), “슈퍼컴퓨터 ‘후다케’에서 배운 대규모 언어모델 ‘후가쿠-LLM’ 공개”.

AI 타임즈 (2023.11.15.), “EU도 ‘소버린 AI’.. 유럽언어 전문 LLM 오픈소스 공개”.

AI 타임즈 (2024.11.20.), “미국 의회, ‘AI 맨하튼 프로젝트’ 제안...‘중국보다 먼저 AGI 개발해야’”.

AI 타임스 (2024.3.24.), “수백세대 걸친 모델 교배로 성능 고도화... 사카나, 병합 진화 모델 공개”.

CLOVA (2024.8.19.), “소버린 AI : AI 시대 네이버의 새로운 도전과 과제”.

IT비즈뉴스 (2024.5.12.), “네이버클라우드, 필리핀 클라우드·스마트시티 시장 진출한다”.

IT Times (2023.2.5.), “구글, 오픈AI • 챗GPT 경쟁사 ‘앤트로픽’에 3억 달러 투자”.

Policy Economy (2024.11.5.), “군사적 이용 안 된다더니, 미 국방 기관에 AI 문 활짝 연 메타”.

Tech World (2024.9.15.), “오픈소스와 AI, 무엇이 문제인가”.

ZDNET Korea (2024.7.17.), “美에 종속되기 싫다…네이버도 꽂힌 ‘소버린 AI’에 전 세계 들썩”.

ZDNET Korea (2024.11.25.) “‘소버린 AI’ 어떻게 봐야할까...‘담론 기반 현실적 전략 필요’”.

◈ 국외 자료

AsiaTimes (2024.11.2.), “China’s People’s Liberation Army weaponizing Metaj’s AI”.

ComputerWeekly.com (2024.11.15.), “Nvidia powers AI development in Japan and Indonesia”.

Converge (2024.11.26.), “Anthropic, Palantir, and AWS Tune Claude AI Models for U.S. Defense Operations”.

Gartner (2024.6.), Hype Cycle for Artificial Intelligence 2024.

Gartner (2024.8.), Forecast Analysis: Artificial Intelligence Services, Worldwide.

Gartner (2023.11.), Forecast Analysis: Artificial Intelligence Software, 2023-2027, Worldwide.

The Economist (2024.1.1.), “Welcome to the era of AI nationalism”.

IDC (2023.9.), Forecast Analysis: Artificial Intelligence Software, 2023-2027, Worldwide.

IDC (2024.8.), Forecast Analysis: Artificial Intelligence Services, Worldwide.

Menlovc (2024.11.), “2024: The State of Generative AI in the Enterprise”.

MIT Technology Review (2024.12.4.), OpenAI 's new defense contract completes its military pivot.

NVIDIA Korea (2024.3.4.), “소버린 AI(Sovereign AI)란?”.

NVIDIA Newsroom (2024.6.2.), “엔비디아, ‘AI 공장’ 설립 위해 폭스콘과 파트너십 구축”.

NVIDIA Newsroom (2024.6.2.), “Computer Industry Joins NVIDIA to Build AI Factories and Data Centers for the Next Industrial Revolution”.

NTT R&D (2024.8.26.), “NTT의 LLM ‘tsuzumi’: 그래픽 문서 이해 가능”.

Omdia Research (2023), Cloud and Data Center Market Snapshop - November 2023.

Reuters (2024.12.5.), “Nvidia to build AI research, data centres in Vietnam with govt”,

U.S. Department of Defense (2023.8.10.), “DOD Announces Establishment of Generative AI Task Force”.

Yan Liu&He Wang (2024.8.), “Who on Earth Is Using Generative AI?”, Policy Research Working paper 10870, World Bank Group.

◈ 웹사이트

https://epoch.ai/

https://github.com/meta-llama/llama-models/blob/main/models/llama3_2/LICENSE

https://www.lgresearch.ai/

https://www.nvidia.com/ko-kr/

저자 소개

김성민 ETRI ICT전략연구소 기술정책연구본부 기술경제연구실 책임연구원
e-mail: songmin516@etri.re.kr Tel. 042-860-6172

소버린 AI 동향 및 R&D에 대한 시사점

발행인 한 성 수
발행처 한국전자통신연구원 기술정책연구본부
발행일 2024년 12월 31일

www.etri.re.kr

Electronics and Telecommunications Research Institute
34129 대전광역시 유성구 가정로 218
TEL.(042) 860-6114 FAX.(042) 860-6504