

DeepSeek 충격의 의미와 개인, 기업 및 국가의 활용전략

(2025.02.12)

메가존클라우드(주)

박 문 기

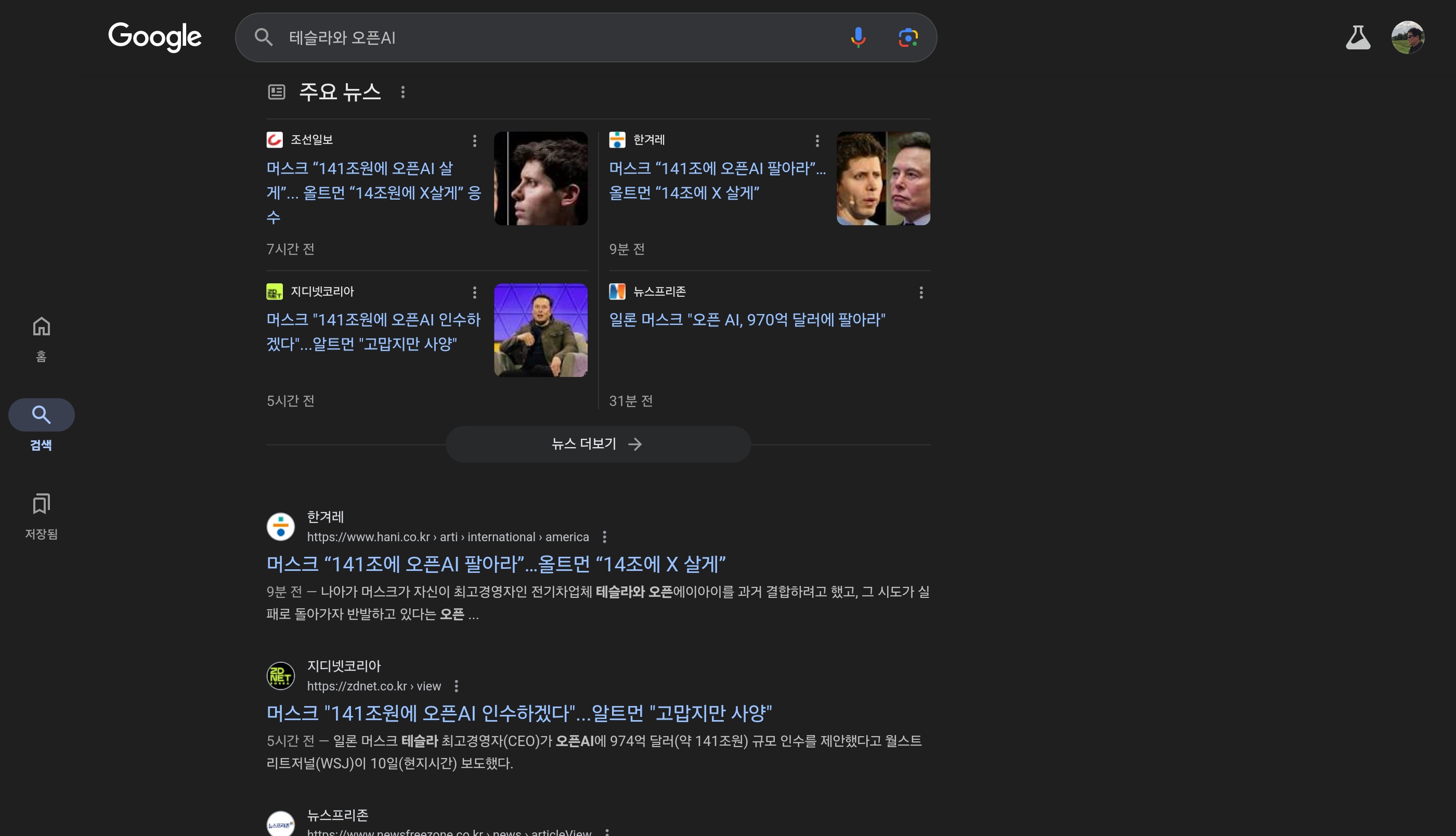
m k b a h k@megazone닷com

+팔이-일공-오삼육칠-오육일사

I. 서문

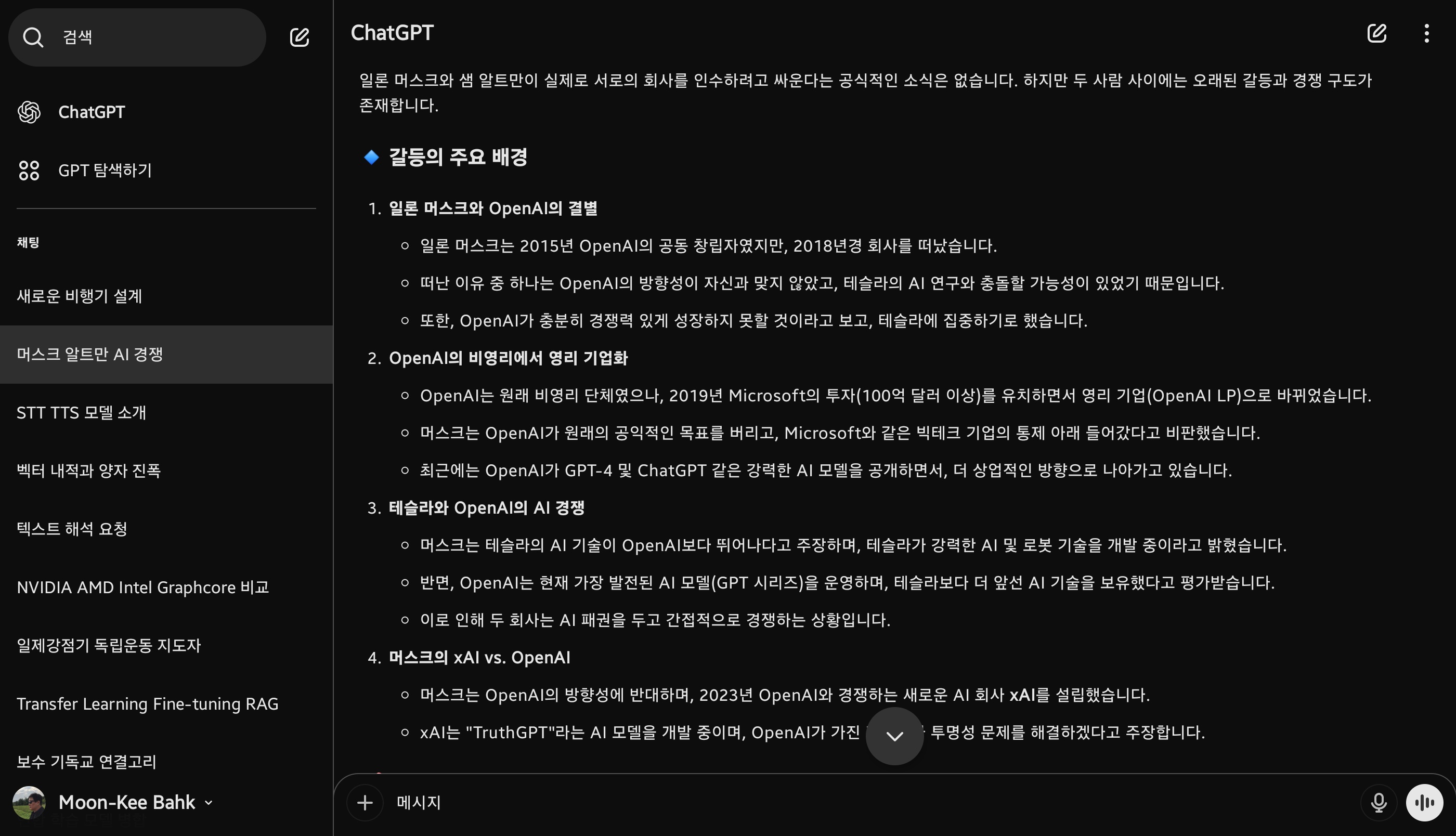
LLM(Large Language Model)이란?

일론 머스크와 샘 알트만의 싸움에서 시작합니다. 왤까요?



ChatGPT에게 물어봤습니다. 답을 알려줍니다.





답은 테슬라도 "TrueGPT"를 만들고 있다는 것입니다. TrueGPT==GPT, ChatGPT=Grok과 같습니다.

그런데, 왜요? 테슬라는 Auto-Pilot를 만들고 있는데 왜 GPT를 만드는 것일까요?

LLM, 초-거대 언어모델은 AGI(AI General Intelligence)의 최종후보로 인간으로 비유하면 "뇌(Brain)", 자동차로 말하면 "엔진"에 해당합니다. 우리의 뇌는 분석하고 판단하고 추론합니다. 뇌가 없는 눈과 뇌가 없는 귀가 무슨 의미가 있을까요?

Auto-Pilot은 응용이고 뇌의 역할을 하는 LLM(초-거대 언어모델)이 필수제로 필요하기 때문입니다. OpenAI로 부터 무료로 ChatGPT서비스를 받지 못한 일론 머스크가 대대적으로 GPU에 중복투자해 LLM을 만드는 xAI를 설립할 수 밖에 없는 짜증내는 상황이었던 것입니다.

ChatGPT가 나온 이후에 대부분의 AI관련 산업의 이것의 개발에 집중하고 있었습니다. 물론 ChatGPT를 만들기 위해 사용했던 엄청난



물량의 DataSet와 GPU 숫자때문에 많은 AI Engineer, 기업과 국가들이 좌절을 하기도 했습니다.

어째든 중요한 것은 이것이 인간의 "뇌"에 해당하는 핵심이기 때문입니다. 그래서 이것을 모든 AI들의 Foundation Model, Basic Model, Basis Mode이라고 불리웁니다. 즉 모든 다른 AI 서비스들은 기본적으로 이것이 꼭 필요한 존재입니다. 이전에 주변부에 해당하는 눈, 코, 귀, 언어번역들 같은 ASI(AI Sepecial Intelligence)를 만들고 있던 사람들이 이것을 인식하고 일단 주변부 만드는 작업들을 내려놓고 LLM(초-거대 언어모델)을 만들어야 한다고 누구나 부르짓고 집중했던 것입니다.

최종 AI의 서비스 관점에서 보면,

AGI ASI Service Application

LLM --> RAG#1 -> Agent#1(눈, 코, 입,...)

RAG#2 Agent#2(자율주행차, 로봇...)

.... ...

RAG#n Agent#n(기업용 ChatBot,...)

Fine Tuning

Transfer Learning

Basic, Basis 특화#1 제품 및 서비스(Edge Device)

Foundation

GPT(Transformer) ChatGPT Operator

범용LLM 특화LLM 특화된 제품 및 서비스

범용 뇌 특화된 전문가뇌 특화된 솔루션 및 컨설팅

학교 기업핵심기술 제품 또는 상품



인간의 뇌에 해당하는 가장 중요한 AI Foundation Model이고, 이것을 기반으로 이미지처리, 번역, 추론등의 다양한 Multi Modal로 발전할 수 있는 것입니다.

그런데, 이 중요한 AI Foundation Model이 미국의 OpenAI ChatGPT, Google Gemini, Meta(Facebook) Llama(오픈소스) 등을 중심으로 발전해 오고 있었습니다.

그런데, 중국의 한 기업이 만든 DeepSeek가 오픈소스를 공개되면서 그 성능이 ChatGPT에 근접하고, ChatGPT에 비해 1/10 GPU성능만을 사용하도록 초초초-최적화된 것으로 밝혀 지면서 AI시장 전체에 근본적인 변화와 충격을 일으키고 있습니다. 왜 일까요?

AI의 시장은 크게

AlphaGO(2016년)---ChatGPT(2022년)---DeepSeek(2025년)

의 시대로 나뉘어 질 수 있다.

(AlphaGO)는 지난 20여년간 2번의 혹도한 겨울을 벗어나는 AI의 가능성에 포문을 열었습니다.

(ChatGPT)는 모든 AI 모델들의 뇌, Foundation Model로 AGI의 중요한 후보로 모든 사람들은 흥분시켰지만 동시에 이것을 가지지 못한 AI Engineer들과 기업들, 국가들은 좌절했었고 혹시 영원히 이것들을 API 서비스 접속을 통해 돈을 지불해야 하는가에 좌절한 시기이기도 했습니다. 우리의 뇌를 영원히 미국에 의존하는 꼴 이었던 것입니다.

DeepSeek는 이제 드디어 AI의 응용서비스 시대로 전환, AI Agent시대로의 전환과 그동안 ChatGPT를 딸아가지 못한 사람과 기업, 국



가들에게 우리가 AI의 뇌를 가질 수 있다는 또 다른 희망의 불씨도 제공한 것입니다.

보다 구체적으로 이 세가지를 하나씩 기술적으로 다시 설명해 보겠습니다.

(AlphaGO)는 AI의 가능성에 포문을 열었지만, 바둑만 둘 수 있는 ASI(특화AI)로써 사람들은 아직 멀었네 했습니다. 아주 방심해 있었고, AI Engineer들도 특화된 AI(뇌가 아닌, 핵심이 아닌 주변부, 코어가 아닌 눈, 코, 잎, 귀, 손, 발, 번역, 이미지인식 등의 역할을 하는 하위모듈단위 특화 AI)들 개발에 열을 올리고 있었습니다.

(ChatGPT)는 AGI(모든 것을 할 수 있는 뇌, 핵심, 코어, Basic, Basis, Foundation Model)로써의 가능성, 인간을 능가하는 엄청난 지적, 추론능력 때문에 모든 사람들에게 충격을 주면서, 이것을 만들기 위한 엄청난 돈, 엄청난 DataSet, 엄청남 GPU사용량을 가지지 못한 AI Engineer들, 기업, 국가들은 또한 "엄청난 좌절"을 맛보았습니다. 이것을 가진 미국, OpenAI, Google은 인터넷의 다음세대 AI 시대도 여전히 우리가 이끈다는 자긍심이 하늘 끝까지 올라간 상태였던 것이구요.

ChatGPT로 인한 좌절과 고난의 행군시대에 미국을 제외한 다른 나라들은 이러한 상황에서도 우리가 그래도 할 수 있는 것을 무었인가? 라는 고민을 하면서, 꾸역꾸역 아래와 같은 것들을 해오고 있었습니다.

(AI Engineer와 기업들은)

오픈소스로 공개되었던 LLM모델 이었던 Meta(Facebook) Liama의 학습이 완료된 모델을 가지고 와서 자-기업이나 특정기업을 위해



서 Trasfer Learning(Fine Tunning)을 했습니다. 이것에 학습기간이 오랜 걸려서 지금 현 시점 데이타반영이 늦은 것을 보완하기 위해 현재 실-시간 데이타를 검색하여 결과에 반영하는 RAG(Retrieval-Augmented Generative, 검색 후 증강, 생성)기능을 통해 AI Agent서비스를 제공하고 있었다. 물론 고 퀄리티 AI-Agent 서비스를 위해서는 ChatGPT에 API연동을 하여 RAG를 만들고 비용을 지불해 오고 있었습니다.

(국가들은)

AI에 대한 산업화를 위해 무엇을 해야하는 지 몰랐다고 봐야 합니다. AI기업들이 비싼 GPU를 많이 많이 사야한다고 징징거리기 때문에 일단 특별한 정책이나 법적인 지원이 없이 국가의 예산으로 GPU를 대량 구매해서 AI데이타센터를 구축하고, 소수의 수량을 GPU를 특정 기업, 연구단체, 개인들에게 년간 무상으로 제공하고 있었습니다. 국가의 예산인 만큼 한 개인, 기업, 단체에게 모든 GPU를 몰아주는 대신 보다 많은 수의 국민들이 GPU를 나누어 사용하는 것이 국민편익이 증가된다는 안이한 판단을 하고 있었구요. 기업들이 한국형LLM을 만들어야 한다는 달콤한 말에 속아서 그저 그렇게 시간을 보내고 있었다. 특히 윤석열 정부 들어와서 대규모의 R&D예산을 경감하면서 이것들 마져 위기에 오자 공무원들 다들 더 무기력하게 대처하고 있었던 것입니다.

2000년 초 인터넷이 나왔을 때 한국은 초-고속 국가망을 구축하고 수 많은 닷컴기업들이 내 놓으면서 마치 우리가 세계의 가장 선두에 있는 것처럼 느끼면서 NAVER와 Daum->Kakao같은 같은 걸출한 기업들을 배출하기도 했습니다. 수 많은 벤처들이 등장했었던 그때를 기억하시는 지요.

그런데 지금 AI시대는 AI관련 벤처기업들이 과연 있는가 할 정도로 너무나 매우 조용한 시대를 보내고 있는 것 같았습니다. 심지어 윤석열



정부들어와서 그동안 한번도 깍긴적이 없는 R&D예산까지 연구카르텔 운운하면서 깍김을 당하는 수모를 겪으면서 심각한 분노와 좌절을 동시에 품어오고 있었습니다.

그런데, 이런 시기에 갑자기 중국의 한 AI기업이 학습이 완료된 LLM, 초-거대 언어모델을 오프소스로 내 놓았습니다. 미국의 AI기업들에게는 새로운 엄청난 종류의 다른 좌절적 충격을 주었고, 반대로 미국의 따라가기가 힘들어 했던 그외 나라들을 DeepSeek를 추앙하며 DeepSeek를 기반으로 드디어 우리도 미국과 ChatGPT를 딸라잡을 수 있다는 LLM을 만들 수 있다라는 "엄청난 기회"를 잡았다고 환호하고 있습니다.

전문가들은 ChatGPT와 DeepSeek간의 기술력 차이가 약 7개월 정도라고 평가합니다. 즉, 우리나라도 7개월이면 미국과 같은 ChatGPT를 만들어 딸아갈 수 있는 AI의 뇌, 즉 AGI 모델을 가질 수 있다는 희망을 갖게 된 것입니다. 얼마난 흥분되는 일입니까?

이전에 Meta(facebook)에서도 오픈소스로 공개한 Liama가 있었음에도 불구하고 왜 중국의 DeepSeek에 대해 전-세계의 국가와 기업, AI Engineer들이 환호와 좌절의 이유는 무었이고, 반대로 내부의 안이함을 까발림을 당하면서 미국 중심의 시각을 가진 사람들과 미국의 기업들, 일부 AI Engineer들은 DeepSeek가 DataSet은 공개하지 않았다느니, 중국 것이라 보안문제가 있다등 DeepSeek를 애써 폄홰하면서 깍아내려야 하는 근본적인 이유, 그들의 멘탈상태, 부러움, 창피함, 안이함, 그들이 가진 쪽팔림... DeepSick하게된 이유가 무었일까?

DeepSeek를 도데체 무엇을 공개한, 오픈소스화 한 것일까요? 그리고 이것은 영향력은 무었일까요?



여기서 잠깐, 참고로 오픈소스체계들을 소프트웨어나 서비스체계개발에 있어서 1등에 비해 거대한 자본과 자원투입이 불가능한 2등들이 집단지성과 연합학습(Federated Learning)을 활용해 1위를 근접하거나 오히려 1위가 되는 2등의 매우 현명한 전략이다라는 점을 먼저 인지하셔야 합니다. Windows vs. Linux의 대결과 같은 것입니다.

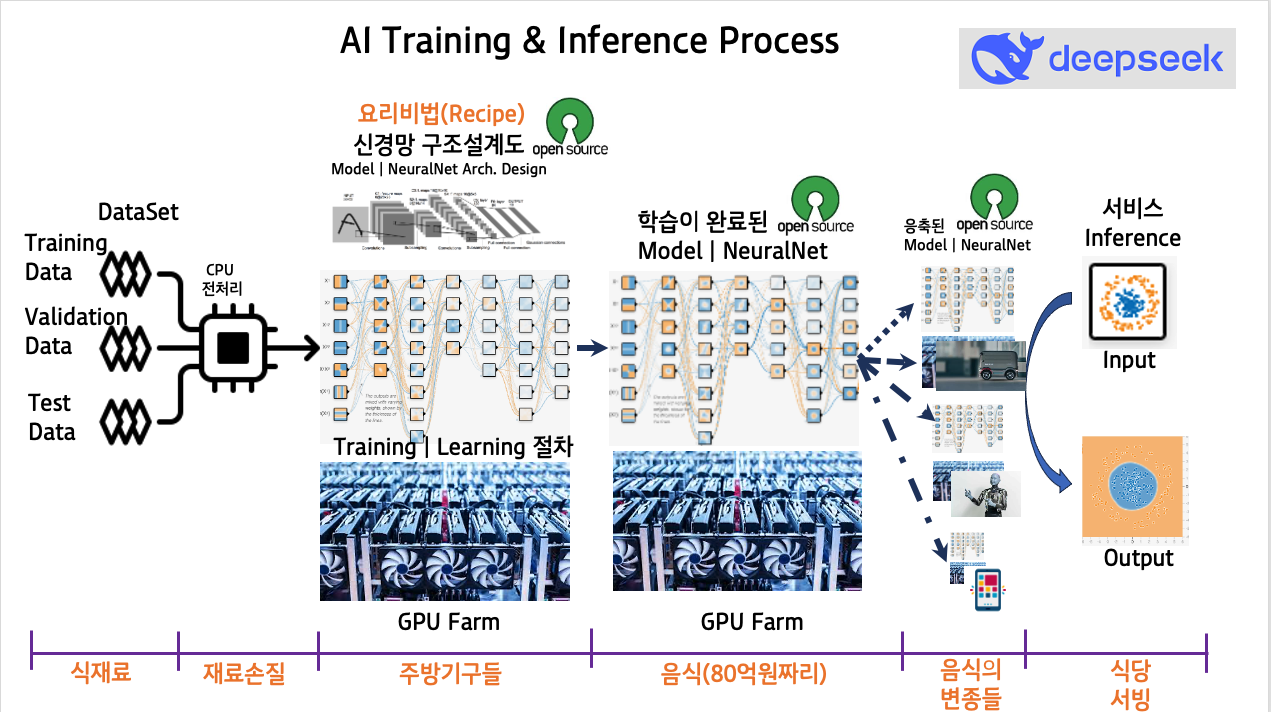
II. 본문

1. DeepSeek의 학습이 완료된 모델 공개의 의미

DeepSeek는 무엇을 공개(오픈소프화)한 것인가요? 무엇이 다른 오픈소스 AI모델들, 예를들면 Meta(facebook)의 LIama와 다른 것일까요?

이것을 이해하기 위해서는 먼저 AI 모델의 개발과 서비화하기 위한 전과정을 조금 이해할 필요가 있고, 이것들에서 DeepSeek가 어느 부분의 무엇들을 공개한 것인가에 대해 알아야 합니다. 그렇게 기술적이지 않으니 한번 이해해 보세요.





AI모델을 만드는 과정을 음식조리과정으로 비교해 예를든다면

1) DataSet: 음식 재료들

2) 모델구조 또는 네트워크 구조, 설계도: 레시피(Recipe)

3) GPU 팜: 조리기구들

4) 학습이 완료된 모델: 조리완료된 음식

5) AI-Agent: 음식판매점

과 비슷합니다.

여기서 DeepSeek가 공개한 것은 음식의 Recipe뿐만 아니라 80억짜리 음식자체를 인류를 위해서 공개한 것입니다. 80억짜리를 지구상의 모든 개인, 기업, 국가들에게 공개한 것입니다. 중국의 다른 기업들이 도덕적으로 어떨지 몰라도 적어도 80억짜리 음식을 모든 인류에게 공개한 것은 당영히 추앙받아야 마땅합니다. 심지어 저는 이것에 노벨평화상을 부여해야 한다고 주장합니다.

그런데 80억 짜리의 음식과 Recipe의 가치를 인식하고 어떻게 사용할 것인가는 전적으로 개인, 기업, 국가의 능력과 태도에 달려있는 것



입니다.

한마디로 중국의 DeepSeek가 80억짜리 물건을 지구상의 모든 인간들이 사용할 수 있도록 준 것이라 했습니다. 물론 이것은 Digial제품이기 때문에 얼마든지 복제해 사용해도 됨으로 추가적인 요리재료(DataSet)은 필요가 없습니다. 맛있게 먹기만 하면되고 상점을 만들어 이것을 토대로 만들어진 상품을 팔기까지 해도 돕니다.

2024년에 인공지능분야가 노벨상의 많이 탓습니다. AlphaFold가 노벨생물학상을, 인공지능의 아버님에 해당하시는 2번의 혹독한 연구겨울을 보내면서 개고생하신 제프리 힌트(Geoffrey Hinton)교수와 존 홉필드(John Hopfield)가 수상했는데, 2025년에는 DeepSeek가 노벨평화상을 받아야 한다고 전 개인적으로 한번더 주장합니다.

2. DeepSeek의 공개와 연합학습의 파급효과

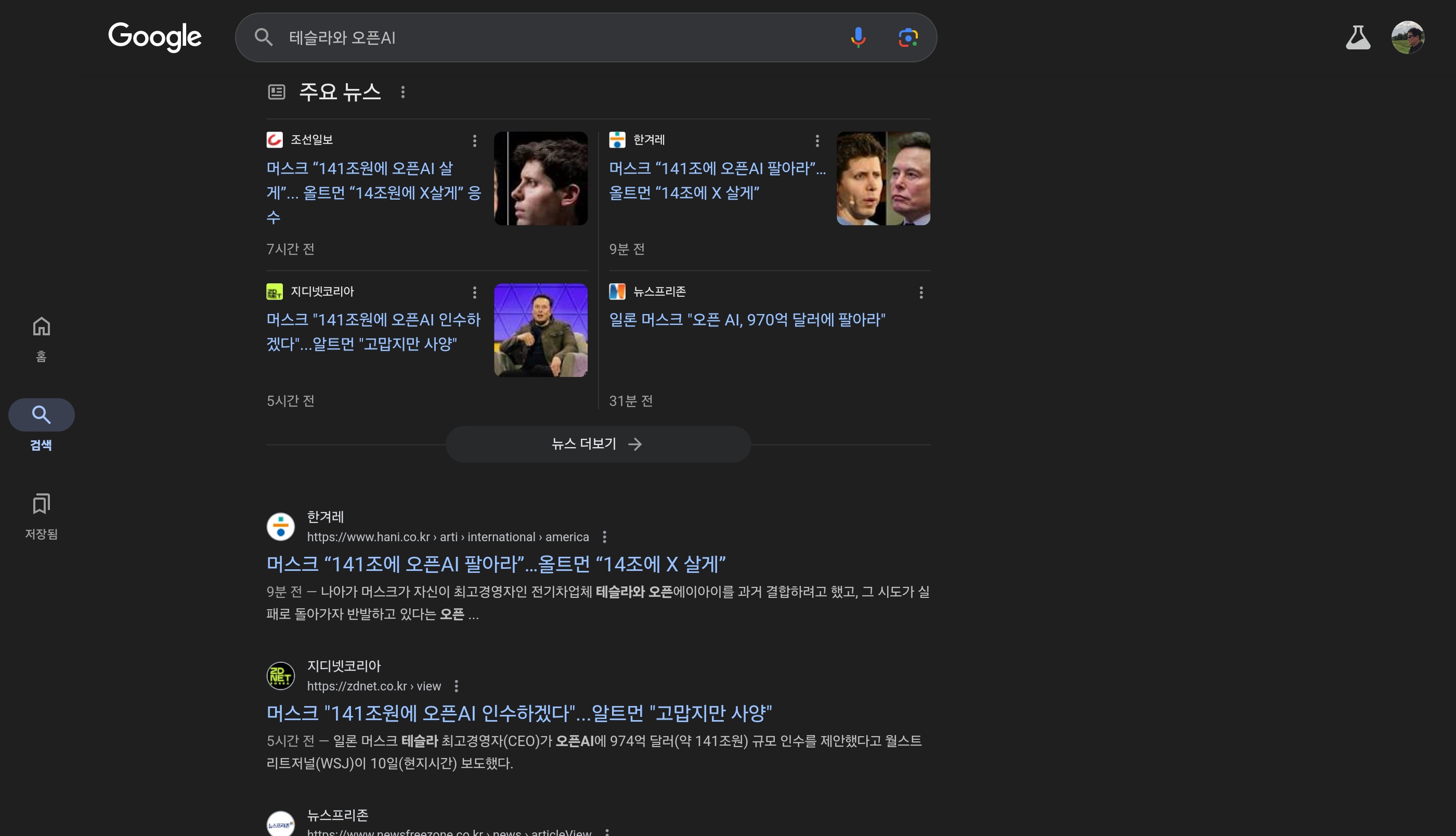
중국과 DeepSeek는 도데체 무었을 노린 것일까요?

답은 2등이 집단지성으로 연합학습을 통해서 1등이 되고자 하는 것입니다.

이것을 이해하려면 AI의 학습방법인 전이학습(Transfer Learning), 연합학습(Federated Learning), 분산학습(Distributed Learning)을 조금만 이해하시면 된다. 여기서 오픈소스화를 통해 집단지성을 이용해 2등이 1등되는 방법에 대한 노-하우를 아시게 될 것입니다.

(테슬라의 전략) 전기차 H/W버리고 Auto-Poilot판매하는 것





즉, 스페이스x에서 우주용 로켓은 아주 잘 만들지만, 테슬라가 자율주행 전기자동차 하드웨어를 만드는 기술은 기존의 다른 자동차회사보다 잘 만드는 것 같지는 않습니다. 자동차 하드웨어는 경쟁도 아주 심하구요. 그들도 이것을 잘 알고 있습니다.

그래서 테슬라가 AI에 투자를 엄청나게 하는 것은 긍극적으로는 자기는 전기자동차 하드웨어를 만들지 않고, 운영체졔인 Auto-Pilot를 타-자동차회사들에게 판매하고 싶은 것이 최종의 목적이라고 판단됩니다. 마치 컴퓨터의 운영체제처럼용. 그래서 몇년 후에는 전기자동차에 자율주행용 Auto-Pilot 1.0, 2.0, 3.0,...등등 어떤 것이 들었는냐에 따라 자동차의 가격이 달라질 것입니다. 그런데, 이런 Auto-Pilot도 Foundation Model인 LLM이 꼭 필요했던 것입니다.

초기에 일론머스크도 OpenAI에 투자했고, 자신의 사회적 영향력으로 OpenAI에 초기투자를 많이 확보해준 것에 대한 공헌이 지대했을 것으로 판단됩니다. OpenAI의 초기에는 만들어진 ChatGPT를 공개하기로 했고, 이것은 ChatGPT+Auto-Pilot으로 테슬라는 무료로 사용하고 싶었던 것입니다. 그런데 너무 많이 GPU 리소스 투자가 필요했고, 마이크로소프트의 도움을 받을 수 없게 되는 상황이 되고, 유료화를 진행하면서 테슬라 입장에서는 나중에 ChatGPT에 접근이 필요할 때 마다 API Call당 비용을 지불해야 하는 상황을 당하게 되어 이들

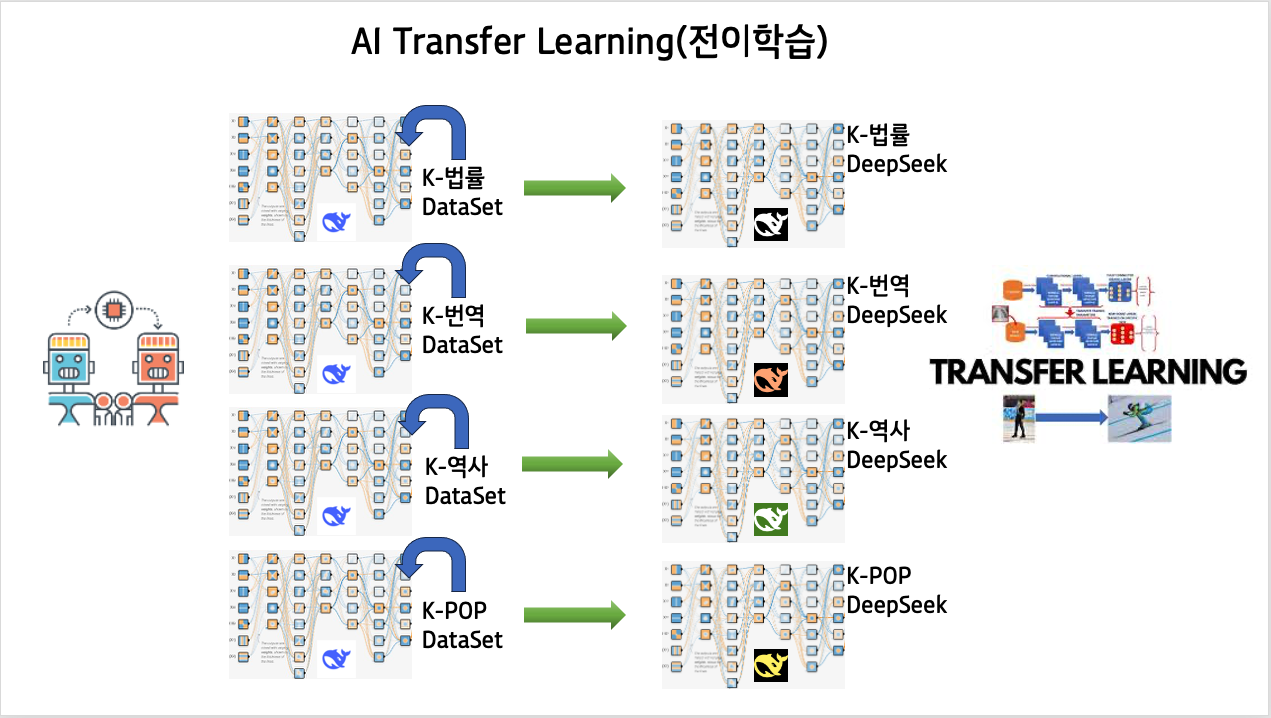


간에 불화가 생겼다고 했습니다.

그래서, 일론 머스크는 xAI를 만들어서 "TrueGPT와 Grok"라는 LLM모델과 서비스를 만들고 있던 것이었고, 이곳에 엄청난 중복 투자가 일어나고 있어서, 일론머스크의 입장에서는 엄청난 부담이 되었을 것을 예측할 수 있습니다. 그래서 서로 으르렁되고 있었던 것입니다.

다시 한번 반복하면. 왜?. LLM은 단순한 초-거대언어모델이 아닌 인간의 "뇌"에 해당하고, 모든 AI-Agent서비스에 기저로 사용되는 AI Foundation Model이기 때문입니다.

(2등 현대자동차의 Auto-pilot 공개해서 1등이 되는 전략: 전이학습과 연합학습)

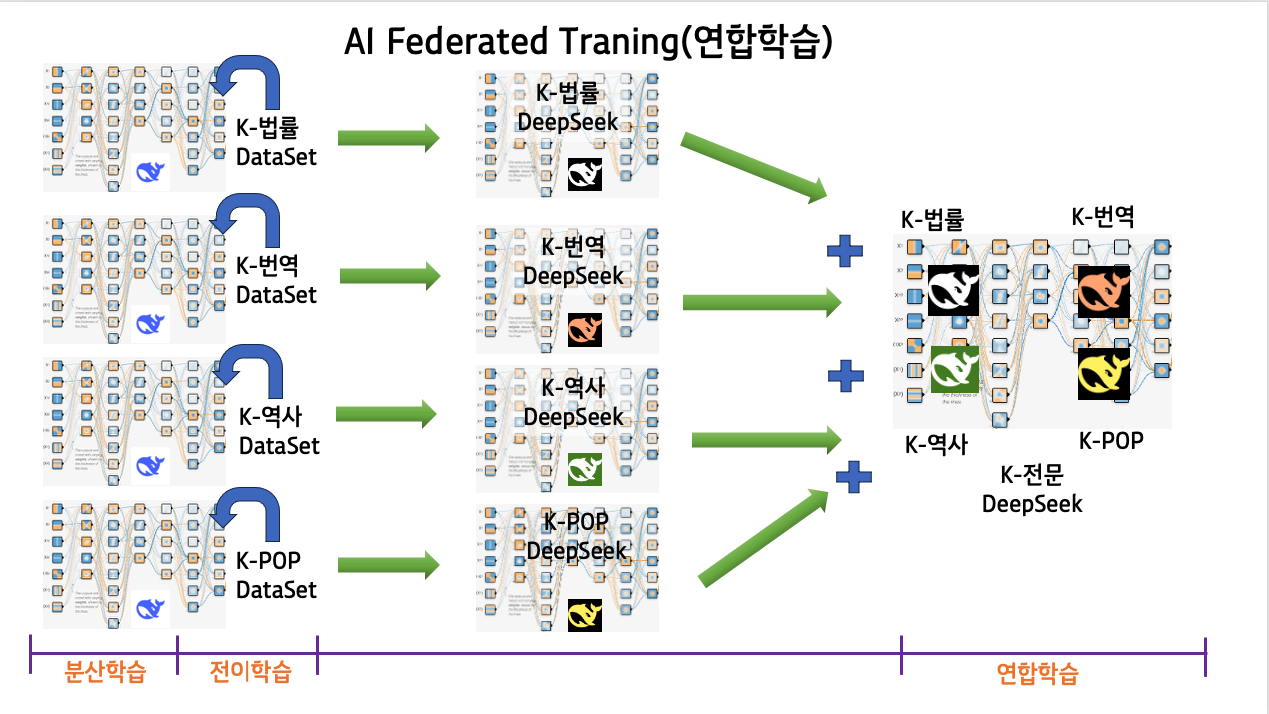


전이학습(Transfer Learning)이란 특정업무를 해결하기 위해서 학습된 AI모델에 추가적인 기능과 또는 다른 업무에도 같이 사용할 수 있도합니다. 기존 학습이 완료된 모델에 새로운 일부/특정 AI DataSet를 추가로 학습시켜 기본+추가의 기능을 갖춘 AI모델을 생성하는 과정입니다.



이것은 기존 기본 DataSet에 추가 특정DataSet를 데이타 수준에서 병합시켜 처음부터 다시 학습하면 GPU들 관련 리소스가 처음보다 더 많이, 시간이 더 들어가기는 것을 방지하는 효율적인 방법입니다.

기존 학습이 완료된 신경망에 추가로 다른 DataSet를 학습시키는 것입니다. 추가적인 시간 리소스 사용으로 충분한 것입니다. 위의 그림에서 스피드스케이팅을 하는 사람에게 추가적으로 스키를 배우게해서 스피트스케이팅과 스키를 동시에 잘하는 사람을 만드는 것과 유사합니다.



연합학습(Federated Learning)은 기본적으로 학습이 완료된 표준적인 모델에 각각 특성이 다른 DataSet를 학습시킨 여러개의 모델을 다시 하나로 통합하면 다양한 특성을 가진 기능을 가진 큰, 더 똑똑한 모델을 만드는 방법인 것입니다.



원래 연합학습에 참여하는 개별 특성화된 전이학습을 또는 여러개가 분산해서 한다고 하여 분산학습 개념도 같이 포함됩니다.

원래 연합학습은 비밀이나 개인정보를 외부에 유출할 수 었는 의료데이타를 이용한 AI모델들을 분산|전이학습 한 후에 하나로 통합하는 것을 위해서 개발된 기술입니다.

예를들면, 어떤 의료용 AI를 만든다고 가정할 때, 일단 국가나 관련 협회에서 기본-표준화된 모델을 만듭니다. 이 모델을 각각의 병원이나 보내서 그 기관의 내부의 자료나 DataSet를 이용해서 추가 학습을 시킵니다. 여기서 반드시 개인정보들은 미리 DataSet에 당연히 제거해야 하겠지요. 그러면 이 모델들에게는 각 병원의 정보|특성이 추가되어 있는 전이학습 모델이 될 것입니다. 다시 병원들에 의해서 전이학습이 완료된 모델들을 국가나 협회로 보내면 이 학습한 각각의 모델들의 융합(연합)하여 하나의 큰 모델을 만드는 방식입니다. 필요하다면 이 과정에 참여한 각 병원들에게 국가나 협회를 통합모델을 공유해 주면 필요한 병원에서는 통합모델을 사용할 수 있는 것입니다. 이 방법은 학습된 모델을 보내는 것이지, 절대 개인정보를 외부에 보내는 것은 아니기 때문에 구현 가능한 AI학습체계가 된 것입니다.

우리는 DeepSeek을 기반으로 집단지성을 발휘해 연합학습함으로써 2등이 1등이 되는 강력한 가능성을 가지게 되는 것입니다.

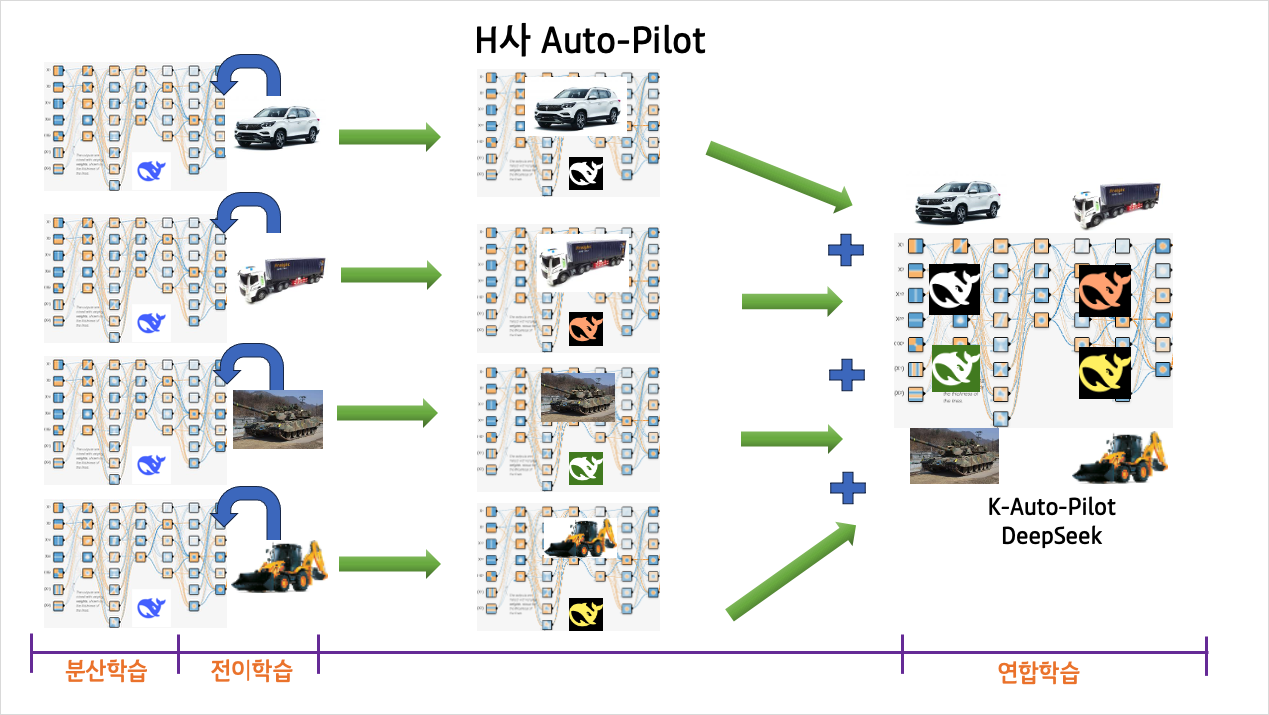
기본 DeepSeek 모델이 공개되어 있기 때문에 각각의 회사나 전-세계 나라들은 각기 특화된 추가 언어 및 특성 DataSet을 이용하여 전이학습(Transfer Learning)을 분산해서 하고, 전이학습이 완료된 DeepSeek특화모델들을 공유합니다. 이것들을 다 합치면 ChatGPT보다 더 뛰어난 DeepSeek 연합모델이 되는 것입니다. 이것을 중국이 노리는 것이고, 이 효과를 전-세계국가들이 누릴 수 있는 것입니다.이것의 파급효과때문에 DeepSeek게 AI시장의 헤게모니 싸움에서



이긴 것이라 하는 것입니다. 이것이 DeepSeek가 AI시장에 던진 "충격"인 것입니다.

정리하면, DeepSeek가 초기 학습한 모델을 공개했고, 이 모델을 기반으로 집단지성(분산학습, 전이학습, 연합학습)을 지성을 발휘해 완전히 똑똑한 모델을 만들 수 있는 것입니다. 이제 이해가 되십니까?

자 현대자동차를 예로 들어 추가로 설명해 보겠습니다.



현대자동차는 Auto-Pilot를 만들고 있기는 하지만 결코 1위인 테슬라를 따라 잡을 수 없을 것입니다. 그럼 이런 2위가 1위가 될 수 있는 방법이 DeepSeek와 같이 현대자동차가 만든 Auto-Pilot모델을 오픈소스화 하는 것입니다.

제가 현대자동차라면 이러한 약간의 "조건이 붙은 전략-조건부 공개"으로 자신이 현재까지 완료된 Auto-Pilot를 공개할 것 같습니다. 조건부 전략이라는 것은 내가 Auto-Pilot를 공개할 것이니 너도 나름 특화된 장비에 대한 이것을 이용해 자율주행을 학습시킨 후 반드



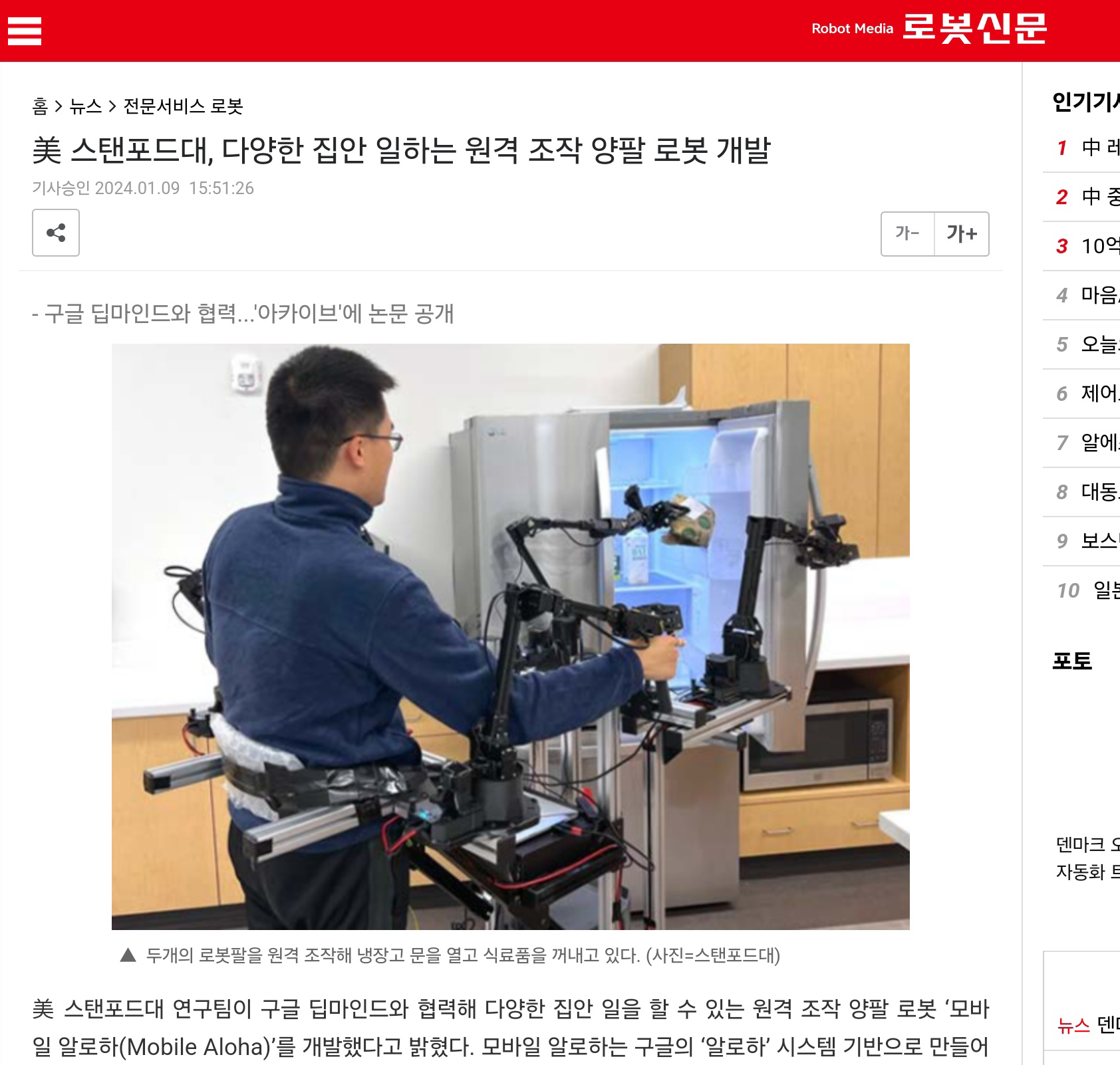
시 다시 오픈소스로 공개해 현대자동차와 다른 회사들이 병합해 사용할 수 있어야 한다라는 것, 조건입니다. 각자의 회사들이 각자 특화된 Auto-Pilot를 전이학습한 후 현대자동차는 이것들 모아서 통합하는 연합학습을 진행하게 되면 다양한 특화된 분야에서도 적응이 가능한 더 똑똑한 Auto-Pilot이 만들어 지는 것입니다.

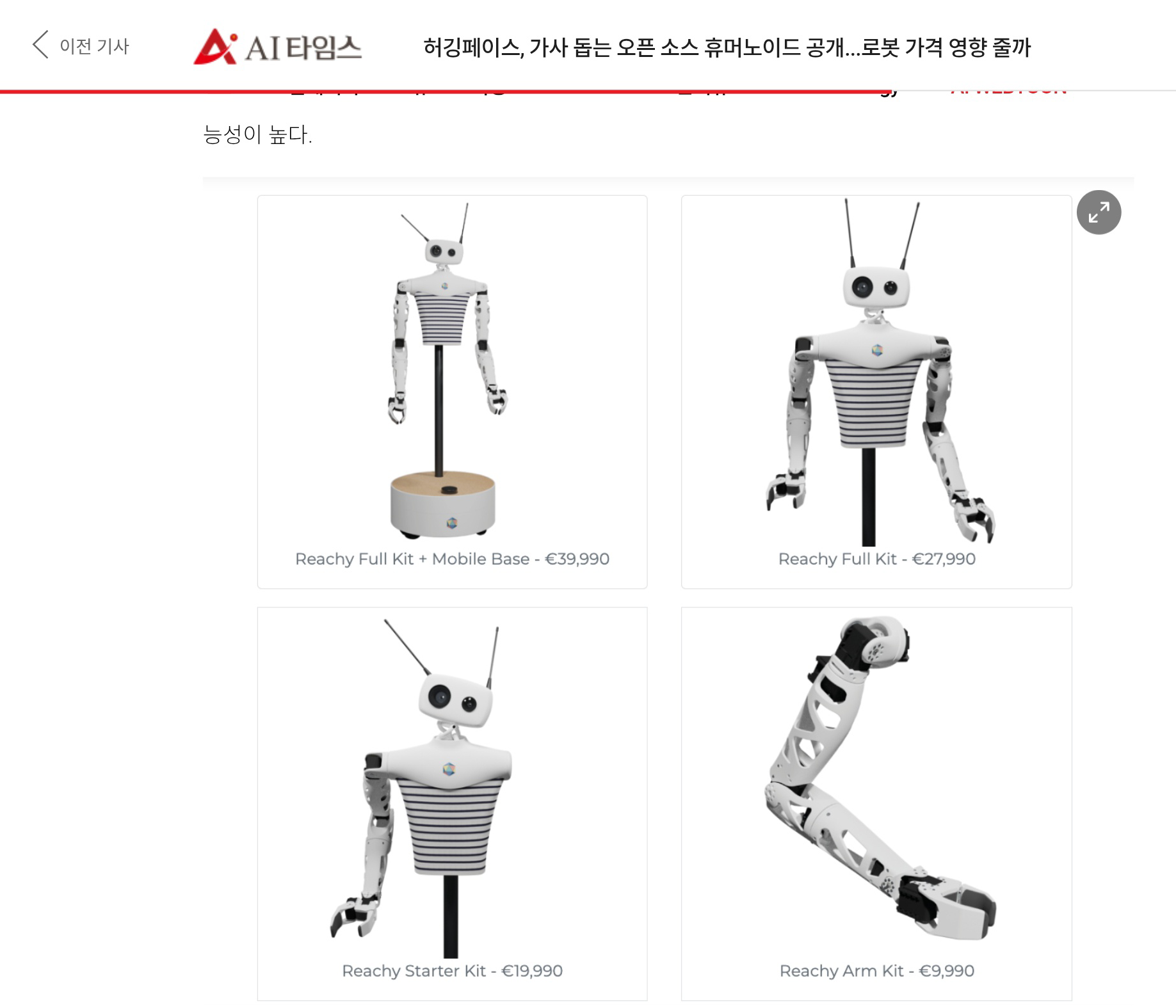
이것이 오픈소스의 전략들이고 가장 성공적인 사례가 Linux였던 것입니다.

또, 다른 사례도 말씀드리겠습니다.

(성공적인 연합학습 사례)

스텐포드-구글 딥마인드 알로하 오픈소스 로봇프로젝트)





(개요)



로봇을 수 많은 다양한 상황에서 훈련을 시키는 것이 어렵다.

특히 가정에서의 다양한 일들, 그중에서 가장 어려운 일은 설거지하는 것이다.

그래서,

* 3D 프린터로 로봇을 만들 수 있는 설계도를 오픈소스로 공개함
* 자체 제작이 어려우면 3,500백만원에 구매도 가능
* 로봇훈련에 필요한 AI 강화학습 모델도 오픈소스로 공개
* 각자 강화학습한 모델을 하나로 통합하면 다양한 상황에서 학습된 똑똑한 로봇이 됨.
* AI 연합학습의 대표적 활용 사례

이런한 LLM의 뇌의 역할과 연합학습한 로봇사례의 사전지식으로 DeepSeek를 분석해 볼 수 있는 능력이 생겼다고 판단됩니다.

3. DeepSeek의 공개의 의미를 바로 알기

1) Windows vs. Linux 의 대결: Windows는 Microsoft라는 기업이 자신의 대규모 인력과 자본을 투자해서 만든 독점적 권리의 제품이다. 정치적으로 말하면 독재국가와 비슷하다. Linux는 리눅스토팔즈가 만들어서 오픈소스로 공개한 버젼이다. End User용 멋진 인터페이스를 가진 Windows에 비해 거칠지만 성능을 요구하는 서버시장용은 Windows Server에 비해 Linux 들이 단연히 많이 사용합니다. 그러면서 Linux는 무료입니다. 물론 죽어도 돈내고 기술지원받고 싶으면 Ubuntu나 Redhat을 사용하면 됩니다. Ubuntu나 Redhat도 Linux가 오픈소스이었기 때문에 제품으로 판매하지는 못하고 기술지원 서비스를 판매하는 것입니다. 이 비지니스 모델도 DeekSeek를 기반으로 서비스를 만들 기업들에게 적용이 가능합니다. 그러나 이것들도 기술지원 받지 않고 그냥사용해도 라이센스상 전혀 문제는 없습니다. 오픈소스 Linux가 있다고 해서 Window가 없어지는 것도 아니



다. 이 세상에는 왕정/독재체제와 민주공화정체제와 그 외의 다양한 국가체계들이 동시에 존재하는 것처럼 말이다. DeepSeek가 아무리 좋은 의도로 만들어져 있다고 했던, 나는 죽어도 돈주고 ChatGPT를 사용하겠다는 사람은 그냥 사용하면 되는 것입니다.

2) 모든 사람들에게 80억 가치가 있는 LLM(초거대 언어모델, 뇌, Brain)을 공짜로 제공하고, 수정도 가능하고, MIT License에 의해 상업적으로 활용해도 전혀 문제가 없습니다.

3) AI Agent들의 핵심 뇌, 엔진을 무료로 제공한 것이다.

4) 이것은 AI핵심 기술 LLM개발에서 벗어나 AI-Agent의 서비스 시대로의 전환을 의미합니다. 이제 본격적으로 AI Agent Store가 생기기 시작할 것이다. Apple App. Store와 같은 것처럼 말입니다.

5) 개인과 기업들은 DeepSeek Foundatation 모델에 자신들만의 특화 콘텐츠를 학습시켜 더 똑똑한 DeepSeek의 수많은 특화지식들을 갖춘 변종들을 만들어 낼 것이다.

6) 이 변종들을 기업이나 국가들은 지속적으로 이 특화모델들을 병합(Merge)하면서 모든 분야에 더더욱 똑똑한 통합 DeepSeek을 만들어 갈 것이다. 이것이 2등이 오픈소스를 통해 작은 집단지성들이 모여서 큰 1등을 추월하는 방법이라도 했습니다.

7) 미국과 OpenAI ChatGPT의 중심인 AI 독재체제로 부터 AI 민주체제로 변화하는 것이다. 위에서 말한 것과 같이 이 2가지 체제는 동시에 존재할 것이다. 서로서로 긍정적인 자극을 주면서 말이다.

2000년 IT기술과 인터넷이 붐업할 초창기에 우리나라는 국가 표준 운영체제를 만들겠다는 포부로 Hana Linux, Buyeo Linux,



Harmonia Linux라는 Linux 베포판들을 만들려고 시도했으나 Microsoft Windows의 견제와 보다 편리하고 멋지 운영체제를 사용하려는 사람들의 개인취양과 욕망이 굳이 돈을 지불하면서 까지 Windows를 사용하겠다는 의사를 꺽을 수는 없었습니다. 지금에 와서 보면 얼마나 많은 국가적인 비용을 낭비한 것일까? 그리고 과거, 현재, 미래에도...얼마나 많은 비용 낭비할 것일까?

Android 운영체제는 또 어떠한가요? 초기에 Android를 만든 스타텁이 삼성전자를 찾아와 이것을 팔겠다고 해지만 삼성은 거절했고 Google이 인수했고 라이센스 문제와 기술력으로 인해 삼성전자는 이것에 대한 비용을 어마어마하게 지불하고 있는지 다 아시지 않습니까?. 이와 같은 실수를 우리는 다시는 하지 말아야 합니다. 단지 중국이 만들었다고 폄홰하거나 정치적인 이유로 또는 감정적으로 싫어서 우리에게 부여된 엄청난 기회를 놓치면 않되는 것입니다.

8) DeepSeek는 초등학교->중학교->고등학교->대학교을 교육받고 대학교를 졸업한 인제를 전 지구상에 무료로 사용하시도록 풀어놓은 것이다. 우리는 더 이상 초등학교부터 대학교까지 많은 시간을 들여 인제를 양성할 필요가 없습니다. 대학교 졸업한 인제는 기업들이 자기 기업에 특화된 OJT교육만 수행해 바로 써 먹을 수 있는 것이다. 이것은 중국이 DeepSeek를 통해 AI분야에서 인류공영에 이바지 한 바입니다. 이제 우리는 ChatGPT, Gemini등 유료의 제품과 DeepSeek등의 무료버젼의 박사급 인제들을 비교해 가면서 필요한 곳에 적절히 사용만 하면 되는 것입니다.

9) DeepSeek는 AI LLM시장의 생태계의 리더가 된 것을 의미하는 것입니다. 소위말하는 정치적, 도덕적 우위에 선 것이라 평가하게 되는 것입니다. OpenAI도 처음에는 이 LLM모델(초-거대 언어모델, 뇌, Foundation)을 공개하기로 했다가 갑자기 유료로 돌아섰다. 이유는 충분히 이해가 됩니다. 너무 많은 자본과 GPU리소스 사용비용을



감당하기 어려웠을 것이다. 하나 만들어서 테슬라 자유주행을 위한 기반, 마이크로소프트 검색 및 오피스의 기반 모델로 사용하려고 했는데 유료로 전환됨으로써 OpenAI 샘 알프만과 테슬라 CEO 일론 머스크와의 불화로 이어졌고, 급기야 테슬라는 자유주행을 위한 Foundation 모델을 재-구축하기 위해 다시 중복해서 엄청난 투자를 해야하는 부담을 안게된 것이었다 말씀드렸습니다.

10) 기존 오프소스 Meta(facebook)이 LIama와는 어떻게 다른 것인가? 답은, LIama가 오픈소스이지만 개발을 위해 엄청나 GPU를 필요로 하는 것은 ChatGPT와 마찬가지이고 심지어 성능이 ChatGPT에 비해 뛰어나지 않았다는 것입니다. 이것은 다음 장에서 보다 세부적으로 설명할 것이다.

4. DeepSeek 모델공개에 DeepSick한 4곳

사전지식으로 LLM(초거대언어모델)은 AI의 기초|기반 모델로 사람으로 비유하면 "뇌(Brain)"에 해당합니다. 그래서, 사람들이 이게 너무너무 중요하다라고 했고, 이것을 만들지 못한 것에 대한 한탄들을 하고 있었다고 말씀드렸습니다.

1. 미국: AI 종주국이라고 떠들면서, 돈을 엄청 투자해 AI를 통해 패권국 유지하겠다 합니다. 자본주의면서 관세장벽 및 중국이 AI를 하지 못하도록 오만가지 장난질| 갑질을 했대습니다. 이런 미국의 갑질에 질려있던 사람들이 약 80억원의 가치를 가진 DeepSeek를 Open Source형태로 갖게되어 미국의 ChatGPT를 7개월 정도 뒤졌었더라고 이제 ChatGPT를 딸라갈 수 있다는 희망을 다른 모든 나라와 AI기업들에게 부여한 것이다. 미국은 AI분야에서 기술패권을 누리기 어렵게 된 것입니다.

소위 AI기술에 대한 도덕적인 해게모니 싸움에서 미국을 중국이 넘어



서게 된 것입니다. AI의 전체 서비스 생태계 구성에서, 연합학습을 통해 지적인 고도화를 통해 DeekSeek가 ChatGPT를 앞설 수 있다는 가능성이 높아진 것입니다. 심지어 무료입니다. 각 나라별, 산업체별 DeepSeek를 이용해 이미 LLM서비스를 시작하거나 DeepSeek기반의 독자적이고 특화된 LLM을 만들 것입니다. 이것을 기반으로 LLM모델개발 필요없이 서비스체계로 완전히 넘어가 DeepSeek기반 수 많은 AI-Agent들이 출시될 것입니다.

DeepSeek는 AI 서비스 생태계의 중요한 도덕적 헤게모니를 확보한 것입니다.

DeepSeek의 대표적인 업적인 하드웨어 기반 최적화 부분에 대해서 알아보겠습니다.

중국의 AI 기업인 DeepSeek는 미국의 수출 제한으로 인해 NVIDIA의 최상위 GPU인 H100을 직접 사용할 수 없었습니다. 대신, DeepSeek는 NVIDIA의 H100에 비해 선능이 대폭 낮추어진 H800 GPU를 활용하여 AI 모델을 개발하고 최적화 했습니다. H800 GPU는 H100의 하위 버전으로, 미국의 대외 수출 규제를 통과하기 위해 성능이 일부 제한된 모델입니다. H100은 4세대 Tensor FP8 정밀도의 트랜스포머 엔진을 탑재하여 대규모 언어 모델의 훈련 속도를 이전 세대보다 A100보다 최대 9배 향상시키는 등 최상위 성능을 제공합니다. 반면, H800은 이러한 최상위 성능을 일부 제한하여 수출 규제를 준수하도록 설계되었습니다. 이것만 중국이 공식적으로 수입할 수 있었던 것입니다.

DeepSeek는 이러한 제한된 하드웨어 환경에서도 소프트웨어 최적화와 프로토콜 개선을 통해 효율적인 모델 훈련을 달성했습니다.

DeepSeek는 2,048개의 H800 GPU를 사용하여 6,710억 개의 매개



변수를 가진 DeepSeek-V3 모델을 훈련하는 데 280만 GPU 시간을 소요했습니다.

반면,

Meta는 16,384개의 H100 GPU를 사용하여 4,050억 개의 매개변수를 가진 Llama 3 모델을 훈련하는 데 3,080만 GPU 시간을 소모했다.

이것때문에 Meta(Facebook)의 LIama모델의 공개해도 사람들은 그리 열광하지는 않았습니다. ChatGPT에 비해 하드웨어가 그리 적게 들지는 않았기 때문입니다.

이러한 비교를 통해 H800과 H100 간의 성능 차이를 간접적으로 확인할 수 있습니다. H800은 H100에 비해 성능이 제한적이지만, DeepSeek는 소프트웨어 최적화 등을 통해 이러한 하드웨어 한계를 극복하고 효율적인 최고의 LLM AI 모델 훈련을 달성했습니다.

중국에 대해서 좋지 않은 감정을 가진 사람들이 DeepSeek에 대한 보안우려 등 바보같은 비판들을 합니다. 정치를 혐호하는 사람들 대표적인 말이 "정치가 밥먹여 줍니까?" 하는 말일 것입니다. DeepSeek가 80억짜리 밥을 전-세계인들에게 준 것입니다. 80억짜리 밥을 먹고 우리는 더 다양한 일을 할 수 있습니다. 밥먹여주는 것이 정치입니다. DeepSeek에 대한 여러 정치적인 판단에 휘둘리지 마시고, 이것 이용해 당장 돈 벌 생각을 해야 합니다. 돈벌어 주는 것이 정치입니다.

2. NVIDIA: 우린 너무 잘나가..V100->H100->H200->GH200, Blackwell들을 만들어내면서 기세 등등했습니다. 예를들면, 한국기업들이 일부러 NVIDIA를 찾아가 차기버젼 Blackwell을 예약으로 사고싶다고 이야기하면 이미 미국기업들이 모든 선주문-예약이 끝났고, 너희들에게는 찌끄러기 몇개 줄께!!!. 심지어 선-입금을 해야하지 않을까? 하면서...



ChatGPT가 학습에 1억달라 들었다면, DeekSeek는 600백만달러 들었답니다. 구축 단가가 즉 1/10도 않되는 가격이다. 결국, 그만큼 저렴한 비용으로 LLM모델을 만드는것이 가능해지면서, NVIDIA의 거품에 대해서 이야기하게 되었고, 이것으로 주가가 하루 아침에 폭락한 이유입니다.

또한 밑에 AI Engineer와 관련되어 있기도 하지만 혹시 GPU를 많이 팔기 위해 NVIDIA는 혹시 병렬처리 라이브러리인 CUDA를 최적화하지 않은 것 아닌가? 라는 강한 의심을 받기에도 충분합니다.

3. OpenAI-ChatGPT: 처음 OpenAI는 Open Source를 타겟으로 모델을 만들고 있었었다. 그런데, 아마도 무슨 이유인지 Close로 변경하여 상업화했고, 사용자들에게 엄청난 비용을 받고 있습니다. 아마도 그것은 수 많은 GPU를 사용해야 하는 비용때문에 어쩔 수 없이 상업화를 한 것으로 추정됩니다. 그런 후 여러자랑질을 하면서 너희들을 이제 우리를 따라올 수 없다. 너희는 그저 API Key를 얻어서 사용료를 지불하면 될 것이다라고 떠든 것이다. 우리는 당신들의 영원한 AI Foundation Model이 될 것이고, 너희는 우리의 영원한 노예가 될 것이다.

LLM은 Foundation Model이기 때문에 테슬라CEO 일론 머스크도 이 프로젝트에 초기 참여했으나 이 LLM모델이 Close로, 상업화용으로 변경되면서 나중에 테슬라 Auto-Poilot에서 API형태로 비용을 지불하고 호출해 사용하게 되었던 것입니다. OpenAI 샘 알트만과 테슬라 일론 머스크의 불화의 숨은 내막이고, 자신의 이름값을 이용해 초기투자를 받도록 해주었던 일론 머스크가 OpenAI 이사회에서 탈퇴하고 나오게 되었고, 나와서 자신의 회사 독자적으로 LLM+Auto-Poilot를 개발하고 있다. 테슬라 입장에서 엄청난 중복투자가 되었던 것입니다.



4. AI Engineer: 그동안 AI을 한다고 거들먹 거리며 새파랗게 젊은 이들에게 억대 연봉을 주어가면 대우했었습다. 그런 대우를 받으면서 이 엔지니어들이 한 것은 Meta(Facebook)에서 공개한 LIama 모델을 Fine Tuning, RAG나 하면서 우리는 GPU가 부족하니 LLM대신 SLM(Small Language Model)를 만드는 것이라 했습니다. 이들이 단지 CUDA에 당연시 만족하고, GPU숫자가 부족하다고 불평불만할 때 DeepSeek 엔지니어들은 CUDA의 하위 수준에서 동작하는 Assembly 수준 저수준 언어체계는 PTX (Parallel Thread Execution)을 통해 최적화했습니다. 완전히 기술을 위해 그들의 열적을 갈아넣은 것입니다.

PTX는 GPU버전별 의존성, 호환이 자라 않되어 특정 GPU에 PTX버젼으로 최적화하면 다음 새로운 Arch.의 GPU가 나왔을 때 다시 최적화해야 하는 문제가 있습니다. 그러나 이런 문제의 여지에도 불구하고 DeepSeek엔지어들은 진정하게 자신들의 기술과 열정을 갈아서 초초초-최적화를 단행해 적은 GPU로 이런 엄청난 DeepSeek를 만들어 낸 것입니다. 솔직히 말하면 우리나라나 다른 나라의 AI엔지니어들을 중국기술자만큼 자신들을 갈아서 기술개발을 하지 않았던 것이 밝혀진 것입니다. 이것이 만방에 들어날까봐 전전긍긍하고 있는 것입니다.

위 제가 한 말들을 완전하게 이해하려면 AI 개발 플렛폼인 Tensorflow, Pytorch등과 GPU만드는 회사들의 라이브러리(ex. CUDA), 물리 GPU간의 수직 연동계통을 이해할 필요가 있습니다.

일반적으로 각 기업들이 지원하는 AI가속기(GPU)들을 사용하기 위해서는 아래와 같은 연결체계가 있습니다.





그리고 NVIDIA 제품에 한정해서 좀더 자세히 정리하면 아래와 같습니다.



여기에서 DeepSeek의 논문에서 보면 여러가지 소프트웨어 구조상의 뛰어난 아이디어들을 구현했다고 설명하고 있지만, 특히 AI Engineer들의 열정을 갈아서 만들었다는 부분은 위의 표의 PTX부분입니다.

PTX(Parallel Thread Execution)은 GPU에서 수행되는 병렬 프로그램을 위한 가상 기계 및 명령어 집합 아키텍처(ISA)입니다. PTX는 고수준의 CUDA(하드웨어가 변경되도 상위 체계가 하위 GPU와 호환성을 많이 유지하는, 다른 말로 최적화다 않된)코드가 GPU에서 실행되기 전에 거치는 중간표현 언어로, 하드웨어 독립적으로 작성된 코드가 하드웨어 의존적인 실제 GPU아키텍처에 맞게 최적화된는 기계어로 변화시키는 일종의 Assembly 수준의 언어입니다.



PTX로 프로그래밍은 하면 GPU하드웨어가 새로운 버젼으로 변경되면 다시 프로그램을 새로운 버젼의 PTX에 맞추어 일일이 다시 변경해야 하는 또다시 열정을 갈아넣어야 하는 작업을 해야함을 의미합니다.

여기서, 우리는 이러한 의심을 해봅니다. NVIDIA는 어짜피 GPU를 많아 팔아야 하는 회사로 바로 CUDA를 최적화하지 않았지 않았을까? 하는 것입니다.

그리고, AI Engineer들에게는 맨날 DataSet 타령, GPU숫자타령만 했지 왜 DeepSeek 엔지니어들 처럼 기술적으로 열정과 노력을 갈아서 CUDA가 아닌 PTX를 프로그래밍하면서 GPU사용량을 1/10로 줄였을 때 너희는 도데체 무엇을 하고 있었는가 묻는 것입니다.

과연 너희들을 그동안 그렇게 많은 년봉을 받으면서 도데체 무었을 하고 있어던 것인가?라는 질문을 반족 하는 것입니다.

너희들은 왜 PTX를 손댈생각을 하지 못했던 것이야...

할 능력도 없었고, 열정도 없었다는 것인가?

돈만 낭비하고 있었던 것 아닌가? 우리가 뭘 모른다고 그동안 뻥치고 있던 것 아닌가? 대답하시오. 라고.

PTX Assembly의 최적화를 DeepSeek엔지니어들이 해낸 것이다. 그리고, 이것을 전세계에 또 무료로 공개한 것이다. 박수...박수...박수...

DeepSeek엔지니어들을 노력과 선한 행동이 아래와 같은 성과로 나타난 것이다.

DeepSeek|2,048개의 H800| 6,710억 개의 매개변수| 모델훈련 280만 GPU 시간|



반면,

Meta LIama 3|16,384개의 H100|4,050억 개의 매개변수| 모델훈련 3,080만 GPU 시간|

한국내에서 중국이 DeepSeek를 만들었기 때문에 무엇인가 거짓으로 말하는 것이 있거나 공개하지 않는 것이 있을 것이다. 보안위협이 있을 것이니 조심해야 한다든지..말들이 많습니다. 이런 사람들의 쓸모없는 말에는 귀를 귀우릴 필요가 전혀 없다.

중요한 것은 80억짜리 제품을 지구상에 인류가 모두 비용없이 사용할 수 있도록 오프소스로 열어 둔 것이다. 쓸데없는 폄홰에 신경쓰지 말고 당장 이것을 가지고 무엇을 할 것인가에 집중해야 한다.

이 분야 최고인 ChatGPT에 비록 지금은 7월정도 기술이 뒤진다 하더라고 인류의 집단지성+연합학습(Federated Learning)으로 AI의 가장 중요한 모델인 "뇌"를 우리나라 뿐만 아니라 다른 나라들도 구축할 수 있는 중요한, 매우 중요한 기회인 것이다.

이러한 기회를 이용해서 그렇다면,

1. 개인은

2. 기업은

3. 국가는

AI 전략을 현 시점에서 어떻게 전략을 수정하고 전개해 나가야 할까요?

우선

(2016년)AlphaGO---(2022년)ChatGPT---(2025년)DeepSeek

의 시대는 AI의 발전에 있어서 중요한 변곡점으로 완전히 전혀 다른



사업적 시각으로 대처해야 합니다.

III. 개인들 활용전략

개인은 모든 일상의 지적, 비-지적 활동을 기록하는 DeepSeek기반 PAA(Personal AI Avator)를 가지는 것입니다. 핸드폰에 DeepSeek를 설치해서 각종 SNS활동, 핸드폰 주소록, 핸드폰 사진들, 각종 수업을 받았을 때 기록한 노트필기 내용을 학습시키고, 내가 만든 모든 Office 파일들을 학습시키고, 나의 모든 전화하기, 핸드폰 SNS/SMS Messange등의 음성활동 학습시키고, 나의 모든 보이스활동을 학습시키고, 나의 먹는 것, 입는 것, 자는 것등, 나의 일상의 모든 상태와 나의 모든 심리적 상태를 알려서 전이학습시키는 것입니다. 이것은 단순한 AI비서를 넘어서, 나 자신의 모든 지적, 감정적 활동을 알고있는 완전한 나 자신의 분신을 만들수 있어야 합니다. 데이타를 보관하면 엄청난 기록장치가 필요하지만 DeekSeek의 모델은 학습만을 통해 새롭게 Update하지 모델 사이즈 자체는 늘어나지 않습니다.

이 AI Avator 분신의 지적능력을 LLM 모델들 평가할 수 있는 시스템을 이용해서 평가하면 단순히 특정 시점에만 시험을 잘쳐서 능력이 절대적으로 평가받지 않고, 초등학교에서 대학교, 그 이상의 학습과정의 역사를 기록하고 나중에 평가받을 것입니다.

이 분신은 나중에 나의 제품이 됩니다.

이 분신에 로봇을 입으면 나의 로봇제품이 됩니다.

나는 죽어도 나의 AI Avator는 영원히 살아남을 것입니다.

내가 죽으면 나는 이것을 다른 사람의 AI Avator등과 연합학습(결혼) 시켜서 나의 자식 AI Avator를 만들 것입니다.

나의 자식은 내 AI Avator를 이어받으며, 나의 모든 지적, 정서적 상태를 이어받겠되는 것입니다.



원래 Meta Facebook은 긍극적으로 개인별 AI Avator를 만들고 싶었던 것으로 추정됩니다.

또 다른 새로운 SNS Platform의 도래가 된 것이다. 카카오, 라인과 같은 기존 SNS가 아닌, AI Avator를 하나의 AI-Agent화하는 새로운 SNS Platform이 나올 것입니다. 아니면 우리가 많들든가요...흠...

학교는, 교육부는 "Digital 교과서" 추진했던 것은 "AI Avator 교과서"로 전환해야 할 것입니다. 세상이 너무 빨리 바뀌네요.

긍극적으로 PAA(Personal AI-Agent)가 나의 생산한, 내가 상속할 수있는 제품입니다.

이 개념을 확장하면, 이제 기업이나 종교, 학습단체들을 자신의 설립자나 위인들을 AI Avator로 만들 수 있습니다.

현대그룹은 DeekSeek기반 정주영회장 Avator를 만들어 살아생전의 정주영회자의 어록을 학습시키고, 정주영 회장님 사후에 현대그룹의 여러 경영환경이 변화된 내용들을 전이학습시키면 완전한 정주영회장이 다시 살아 돌아오시게 할 수 있습니다.

이러한 질문도 가능합니다. 질문할 수 있다.

회장님, "현재 우리그룹과 국내외 경제상황이 이렇게 변화하고 있습니다. 회장님께서 우리가 어떻게 대치해야 하는지를 알려주십시오."

종교단체들도 이제는 종교지도자의 역할이 끝이 날 것입니다. 종교지도자들의 모든 문제의 시작들은 자신이 모시는 종교지도자와 다르게 자신의 욕심이 포함된 말을 추종자들에게 전파하면서 원 종교지도자 또는 하나님의 말씀을 왜곡하는 것이었습니다.



기독교 성경을 완전하게 학습시킨 예수님AIAvator,

유교경전을 완전하게 학습시킨 공자님AIAvator,

불교경전을 완전하게 학습시킨 부처님AIAvator,...등등

이렇게 질문하는 것이 가능할 것입니다.

"예수님 입장에서 현재의 이러한 상황에 대해 어떻게 판단하십니까?"

라고." 이제 거짓말들은 다 들통날 것입니다.

이순신 장군님도, 김구선생님도, 안중근장군도,...

종교지도자들을 포함하여 모든 지식노동자들의 거짓말하는 시대는 끝났습니다.

개인화된 AI Avator든, AIFriends, AICompanion,....이건 이름은 상관없습니다.

IV. 기업들 활용전략

(2016년)AlphaGO---(2022년)ChatGPT---(2025년)DeepSeek

그동안 기업들은

(2016년 AlphaGO)가 나왔을 때 기업들은 AI의 가능성에 대하 논의는 많이 했었습니다. 그런데, 사용된 GPU의 어마어마한 수량에 좌했었습니다. 그래도, 그럼 현 시점에서 우리는 무엇을 해야할까? 해서 나중에 AI-Agent로 사용한 Edge기술들 눈(CNN), 입 Speech-to-Text(RNN-T), 번역(RNN), ...등등을 만들고 있었답니. 동시에 Alpha-GO는 아직 바둑밖에 두지 못하내 하면서 안이한 세월을 보내는 사람들도 있었구요.



(2022년에 ChatGPT)가 나오면서 기업들은 한마디로 AGI(AI General Intelligence)의 후보를 보면서 감탄하면서도 또 동시에 좌절을 더 경험했습니다. ChatGPT학습에 사용된 어마어마한 DataSet과 어마어마한 GPU Farm수량, 기술자 인적구성, 투입된 자본금 등에위해서.

그래도 Naver와 같은 일부 회사들을 한국형LLM을 만들어 보겠다고 시도는 하고 있었습니다. 그러나 그 나머지 작은 회사들은 좌절했습니다. 그래도 이렇게 앉아 잊지만을 않아야 했기에 Meta(Facebook)이 만들어 공개한 LIama모델을 가지고 와서 Transfer Learning의 Fine Tunning에서 특정 기업용 SLM(Small Language Model)을 만들어 팔기도 했습니다. 아니면 아예 ChatGPT에 돈을 지불하고 API 호출을 하고, 현재 최신의 정보를 인터넷검색하여 결과를 증강하는 RAG(Retrival Augumented Generative)기능으로 AI Agent들을 만들고 있었습니다.

(2025년 DeepSeek)가 오픈소스로 공개되었다. 이제 기업들은 무엇을 해야할까요?

이제 한국형LLM을 만든다는 것이 끝난 것이다.

이제 AI관련 기술과 인력이 없고, GPU가 없고, DataSet이 없어 LLM을 못만든다는 말도 할 수 없습니다.

1. 일단 기업내부 활용을 위해 자기 기업내 모든 지식자산들, DB, Office문서, 각종 기술자료들을 DeepSeek에 학습시켜 우리 기업만들 특화LLM을 만들어야 한다. 이제 지식의 유지와 발전에 사람에게 의존하지 말고 DeepSeek에 기록해 노-하우와 Knowledge-Base를 구성해야 합니다.



2. DeepSeek를 이용하여 타 기업용 특화LLM 대리학습시켜 주는 솔루션을 판매할 수 있다. 이 작업은 DeepSeek Full State Model만큼의 GPU 자원을 필요로 한다. 돌려서 특수기업특화LLM을 만들어 주고 좀던 작은 규모의 GPU에서 돌 수 있도록 응축LLM을 만들어 줄 수 있는 비지니스 진행이 가능할 것입니다.

DeepSeek의 Full State Model를 담을 수 있는 GPU Farm을 가지고 있다면 개인들의 모든 문서정보들을 학습(Transfer Learning)시키고, 이것을 다시 응축시켜 작은 모델로 만들어서 제공해 주는 비지니스를 진행할 수 있다는 것입니다.

3. 이제 진정한 AI-Agent 서비스의 시대로 전환하는 것입니다. AI-Agent라 불리건, AI-Operator, AI-Service, AI-Inference Service라 불리건 상관없습니다. 기업 특화LLM에 눈, 귀,...등을 역할을 하는 Edge Device에 응축된 DeepSeek LLM엔진을 탑재해 서비스나 제품을 만들 수 있을 것입니다. ChatGPT에게 비용을 지불하는 API Call이 없기때문에 아주 싸게 만들 수 있습니다.

4. DeepSeek의 Full Model를 수행시킬 수 있는 GPU들을 충분하게 가지고 있는 기업이라는 Apple의 AppStore, Google PlayStore처럼 AI-Agent들을 서비스화해 판매하는 AI-Agnet Store를 만드는 비지니스를 진행할 수 있을 것이다. 단 이때 DeepSeek를 돌리는 리소스를 여러 AI-Agent들이 공유하면서 무료로 제공할 수 있을 것입니다.

5. 비슷한 일을 하는 업체들을 예를들면, 현대자동차, 기아자동차, 썅용자동차들은 DeepSeek기반 자율주행AI를 표준화 후 오픈하고 서로 합병(연합학습)시켜 테슬라의 Auto-Pilot에 도전할 수 있을 것이다.



V. 국가의 전략

(2016년)AlphaGO---(2022년)ChatGPT---(2025년)DeepSeek

우리나라는

AI Hub --- 광주AI데이타센터 --- 국가AI데이타센터를..

(2016년 AlphaGO) 당시 많은 사람들은 AI가능성에 눈을 뜨기 시작했다. 이때는 AlphaGO는 단지 LLM과 같은 AGI가 아닌 바둑만 잘 두는 ASI라는 방심하고 있었다고 했습니다.

이 시점에 AIHub가 만들어졌고, AI를 위한 DataSet도 많이 구성하여 공유했습니다. 사람들이 GPU가 부족하다 아우성을 쳐서 국가가 대신 GPU를 구매한 후 1~2개의 GPU를 1년동안 무상으로 제공하기도 했습니다. 1~2장이지만 많은 국민들과 연구자들에게 GPU를 사용할 수 있도록 한 단순한 숫자가 평가시 국민편익증진에 많이 도움이 되었다고 평가받았습니다. DeepSeek를 통해 보면 진짜 아무일도 아니었습니다.

(2022년 ChatGPT)가 나올 땨ㅐ 쯤에 국가광주AI데이타센터가 만들어 졌습니다. 저는 이곳에 Graphcore IPU를 납품한 경험이 있었고, 센터의 설립이 끝난 후 차세대 방향을 잡는 자문단에 참여를 했기 때문에 이 센타의 현황과 미래전략에 대한 어느정도 깊은 이해를 하고 있었습니다.

차세대 자문단 회의에서 저는 아래와 같은 내용을 강하게 주장했습니다.

ChatGPT가 나온 시점에서는 예전의 소량의 GPU를 국민편익증대라



는 평가목표를 두고 개인연구자나 기업들에게 1년간 무료로 제공하는 방식은 이 시점에서는 잘못된 접근방식이다. 이제는 모든 전체 GPU팜을 오직 "한국형 LLM"을 개발하기 위해 집중해야 한다. 그래서, 하나의 큰 프로젝트를 만들고, 한국형 LLM을 만들고 싶어하는 모든 회사들이나 연구소등이 연구원들을 광주AI데이타센터에 파견해서 공동으로 연구해서 한국형LLM을 개발하고 그 결과를 모두 참여자와 심지어 참여하지 않는 모든 국민과 기업, 연구소들이 공유해야 한다고 줄기차게 주장했지만 받아들이지 않았습니다. 이유는 그들은 LLM이 단지 언어모델이라는 것 외에 이게 모든 AI응용의 Foundation Model이고 인간의 뇌에 해당하다라는 인식이 없었기 때문입니다.

이 시점에 기업들은 나름 한국형 LLM을 만들겠다고 대대적인 투자도 진행했습니다만 괄목할 만한 성과는 없었습니다.

이 기간에 AI관련 모든 세미나에서는 오직 초-거대AI=LLM(초-거대언어모델)이라고 모든 사람들이 이것의 개발에만 집중하고 있었습니다. 이것을 개발하기 위해서는 서로 힘을 합쳐야 한다. 국가는 우리를 위해 보다 많은 GPU Farm를 프리로 ㅜ제공해야 한다고 외치고 있었습니다.

이 시점의 끝에 국가가 나서서 지속적인 투자를 해야 하는 시점에 윤석열정부에서 R&D예산 삭감을 단행하면서 AI관련 연구자나 기업들에게 엄청난 좌절감을 안겨 주었습니다. 아마도 윤석열정부는 국가가 나서서 많은 연구투자를 하지 않는 일본정부의 모델을 상정하고 있었나 봅니다. 국가가 나서서 R&D를 많이 투자하지 않는 일본의 경제는 지금 어떤가요?라는 질문이 필요합니다. 아니면 한국의 미래산업까지 완전히 망가트리려 일본에 바치려는 친일파들의 전략이 아니었을까도 의심합니다.



(2025년 DeekSeek)가 발표된 후 국가는 전혀 다른, 완전히 새로운 AI지원전략을 가져야 합니다. 몇 가지 원칙들과 집중해야 하는 방향 및 방법에 대해서 생각을 풉니다.

1) 더 이상 한국형LLM을 만들 필요가 없습니다. 이제는 완전히 AI-Agent 시대로 전환해야 합니다.

2) 더 이상 기업의 AI Model학습용으로 국가 AI GPU Farm을 제공할 필요가 없습니다. DeepSeek가 1/10이면 훈련이 가능하다라는 것을 배웠기 때문에 이 정도의 학습용 GPU는 개별 기업들이 준비할 수 있도록 해야합니다.

3) 국가는 AI 서비스인 Inference체계, AI-Agent개발체계를 도와주는 방향을 GPU리소스를 사용할 수 있는 체계로 바뀌어야 합니다. 국가가 나서서 AI-Agent Store를 만들어야 합니다.

4) 국가는 AI 연합학습을 독려하고 개별적으로 학습한 모델들을 연합학습으로 통합할 수 있는 체계를 만들어야 합니다. 관련하여 각종 AI모델들의 표준화를 통해 연합학습과 통합이 가능하게 해야 합니다.

5) 로봇을 대표되는 AI Edge Device들도 표준화를 독려해야 합니다. 표준화된 로봇과 표준화된 AI 강화학습 모델을 오픈소스로 공개하면 마찬가지로 연합학습의 효과를 얻을 수 있는 것입니다. 이제 우리는 DeepSeek를 통해서 우리는 로봇과 Edge Device의 뇌를 얻었습니다. 단지, 여기에 Edge Device를 붙여서 AI 상품화를 가속화해야 하는 것입니다.

여러가지 AI Edge Device 에 탑재가 가능할 수 있도록, 국가는 연합학습을 통해 지속적으로 똑똑해질 DeepSeek를 만들도로 독려하고, 학습완료된 모델을 작게 응축해서 다양한 Edge Device에 포팅이 될



수 있는 파생응축버젼들을 국가가 GPU팜 리소스를 통해 만들어 제공해야 합니다.

6) 연합학습에 지속적으로 독력하고, 연합학습이 가능하도록 눈, 코, 입, 귀, 손, 발등에 해당하는 Edge Device들을 위한 AI모델들을 표준화하고, 각 기업들이 자신들의 상황에 맞는 특화된 것을 만들고, 국가는 다시 이것들을 연합학습시켜 더 똑똑한 Edge Device용 AI 모델들을 만들내야 하는 체계로 완전히 전환해야 합니다.이 연합학습에 참여하는 기업들과 연구소들에게만 국가의 GPU Farm을 무료로 사용할 수 있도록 제한해야 합니다.

7) 계속 이야기 되었지만 AI Edge Device로 대표되는 로봇, 전기자동차,...등의 표준화도 매우 중요합니다. 이 장치들이 표준화되면 표준 AI강화학습 모델을 가지고 각자 특화 학습한 후에 연합학습을 통해 더더 똑똑한 AI 모델을 만들어 갈 수 있기 때문입니다.

사람의 머리속 생각들은 엄청나게 다양합니다. 그런데 사람의 몸은 남자와 여자로 크게 둘로 나뉘고 구조도 매우 통일되어 있지 않습니까? 로봇도 전기자동차도 이것이 가능합니다. 표준화된 로봇을 다양한 환경에서 표준화된 강화학습모델로 학습시키 후 다시 연합학습할 수 있는 체계로 독려해야 합니다.

8) 국가의 투자 방법도 획기적인 전환이 필요합니다. 국가가 1조원을 투자해서 GPU를 산다라고 가정하면, 여기서 가정이 필요한 것이 국가가 1조원을 투자할 때 이 모든 금액을 GPU사는데 사용하지 않을 것이기 때문입니다. 데이타센터 땅사고, 건물짓고, 인력구성하고, 적어도 5년간 유지보수 비용, GPU를 운영하기 위한 서버, 네트워크, 보안장비들...등등 하다보면 아마도 1/4정도가 GPU사는 비용일 될 것입니다.



하여튼 가정을 빼고 순수하게 1조원 투자해 NVIDIA H100 GPU사면, H100이 한 5,000만원 정도 하니까, 대략 20,000개 정도 살 수 있겠습니다. DeepSeek가 H800을 2,048개 사용했으니 AI학습이나 응축하는 프로젝트 10개 정도 돌릴 수 있겠습니다. 다시 1/4를 적용하면 2.5개 정도 프로젝트 돌릴 수 있는 금액이네요. 턱없이 부족합니다. 또한 AI-Agent 서비스 즉, Inference용으로는 상용해야 하는 모델이 GPU에 로딩되었을 때 이 GPU는 학습용으로 동시 사용할 수 없습니다.

방법은 아래와 같습니다.

국가는 적어도 10조는 투자해야 합니다. 그리고 이 금액이 순전히 GPU와 부대장비 사는데 집중되어야 합니다. 데이타센터와 전기료, 운영비는 예산을 별도로 잡아야 합니다.

그리고, 민간재원들 들여와야 합니다. 그래도 크다고 하는 Naver, Kakao, KT, SKT 같은 통신사들의 수준으로는 어림도 없습니다. 이 분야는 국가의 인프라다라고 차원을 달리 생각해야 합니다. 그래서 도로, 항망, 공항과 같은 인프라에 대대적으로 투자할 수 있는 메쿼리인프라와 같은 급의 회사들의 투자를 유도해야 합니다.

그리고 대대적인 국민들의 의식도 고양시켜 "국민채권" 또는 SPC를 설립해 "국민주식"형태로 국가AI데이타센터 구축에 국민들도 참여할 수 있는 기회를 열어야 대대적인 투자가 가능합니다.

9) GPU구매도 NVIDIA 위주에서 벗어나야 하고 다른 회사들 예를드면 AMD GPU, Intel GPU, Graphcore IPU등과 경쟁시켜야 합니다. 국가주도 AI데이타센터가 만들어 지면 AI Inference서비스에 집중해야 한다고 말씀드렸습니다. 그래서 GPU리소스도 AI학습용(Training)으로는 NVIDIA GPU를 1/2정도 구매하고, AI서비스(Inference)용으로 기타 3개 회사의 GPU를 1/2정도 구매하면 회사들간의 경쟁을 통해서 가격인하를 유도할 수 있을 것입니다.



국산 NPU들은 Edge Device용으로 최적화해서 용도를 집중하는 것이 좋고 정책도 그렇게 가야 합니다.

정리하면

AI Training용은 NVIDIA GPU,

AI Inference용는 AMD, Intel, Graphcore GPU/IPU등로

Edge Device용으로는 국산 NPU들을

사용하면 되는 것입니다.

읽어 주셔셔 감사합니다.

이상. 끝.