2026 Transactions of KNS Autumn Meeting

Hire Your Own Researcher on a Single GPU:

Building a Local LLM Agent for Reactor Physics

디지털기획부 이대영

2026. 10. 28



FNC 주|미래와도전
FNC Technology Co., Ltd.

FNC Technology Co., Ltd.

CONTENTS

1 Introduction

2 Large Language Models

3 Local LLM Deployment

4 Building AI Agents

5 Future Work

FNC TECHNOLOGY CO., LTD.

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

1 Introduction

개발이 쉬워지고 있다.

프로그래밍 언어

- Fortran, C, C++, JAVA, JavaScript, Python
- HTML, XML, 마크다운은 문서의 구조와 의미를 나타내는 마크업 언어이고, JSON, YAML은 데이터를 기술하는 데이터 언어이며, SQL은 데이터베이스와 대화하기 위한 질의 언어임.

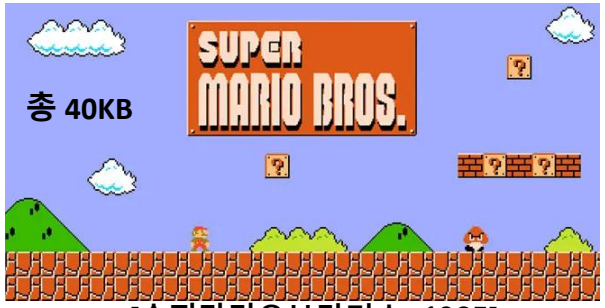
하드웨어, 통신기술의 발전

- 초기: PC 성능이 낮고 비용이 비쌀 때, 인터프리터 언어 (Python, JS 등) 는 "속도가 느려서 실용 불가" C, C++ 같은 컴파일 언어로만 최적화 개발
- CPU 성능, 메모리, 스토리지가 대폭 개선되면서, Python, JavaScript, Ruby 등 인터프리터 언어로도 웹 개발, 데이터 분석, AI/ML, 자동화 스크립트 등을 실제 업무에서 직접 자주 사용

개발을 쉽게하기 위한 노력

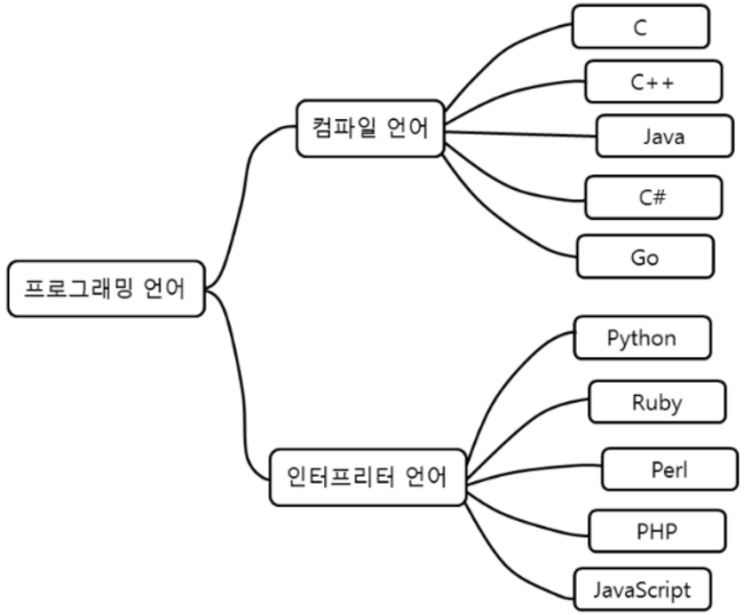
- IDE, 프레임워크/라이브러리(Django, Spring, React) nexacro PLATFORM
- 그리고 구글이 있었다.

이젠 바이브코딩의 시대



[수퍼마리오브라더스, 1985]



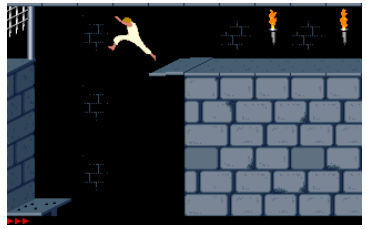


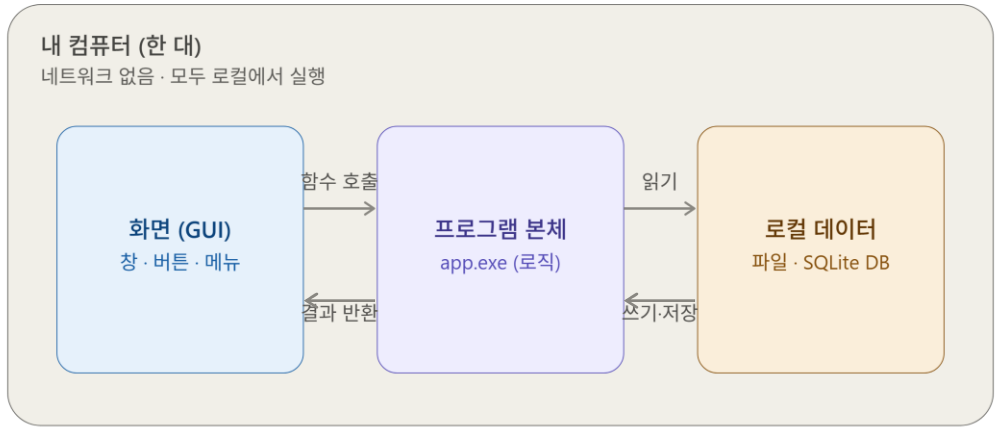
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

1 Introduction

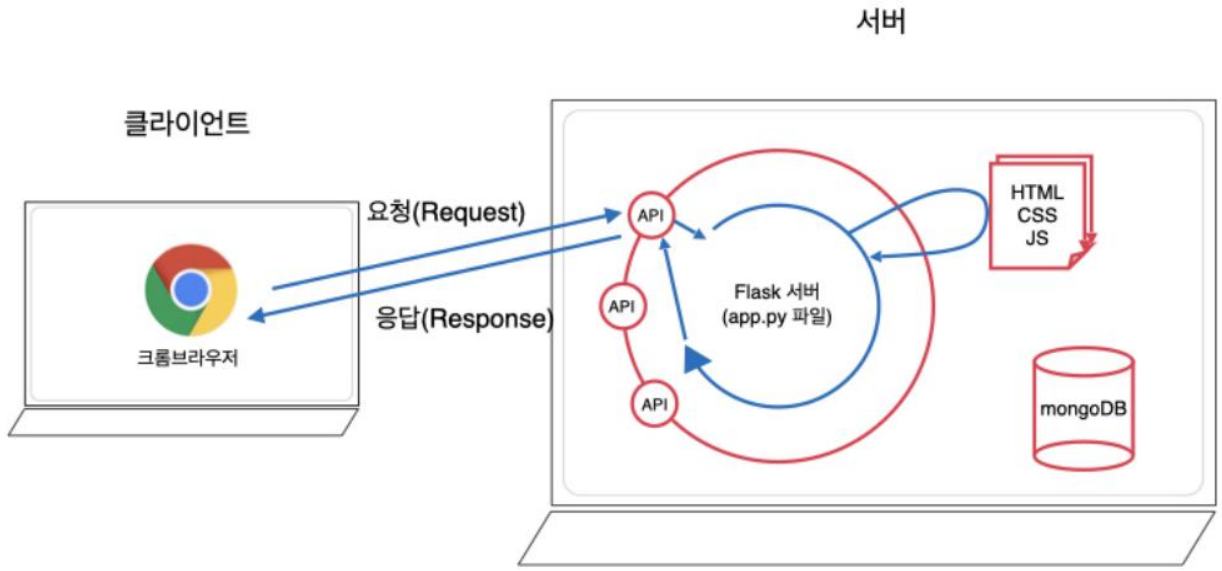
SW 형태의 변화

데스크탑 응용SW

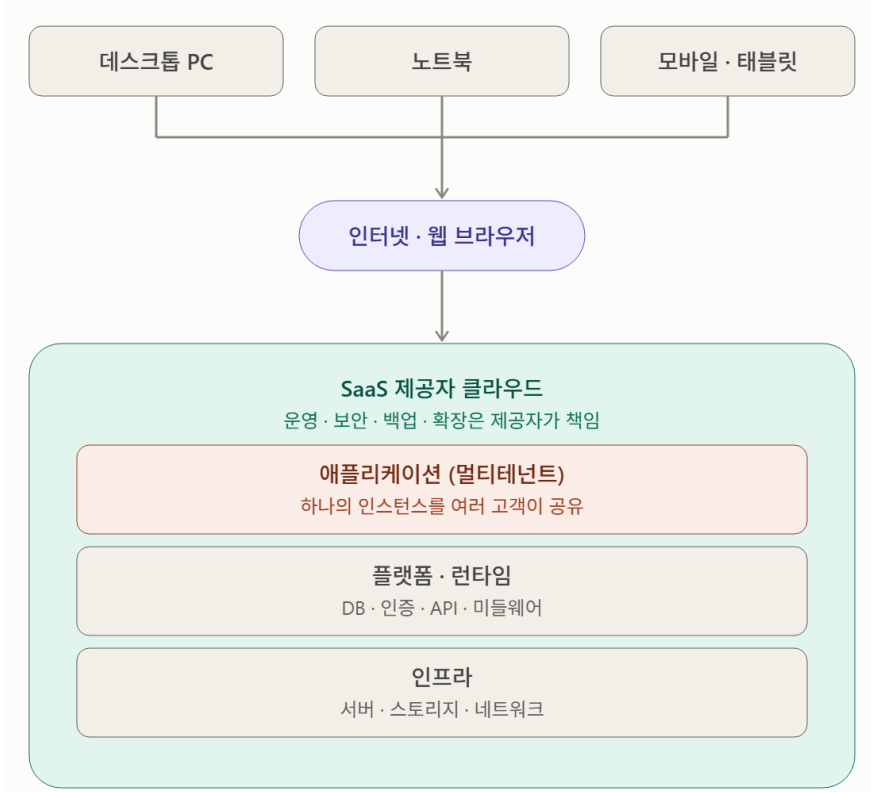




웹 응용SW



SaaS(Software as a Service)



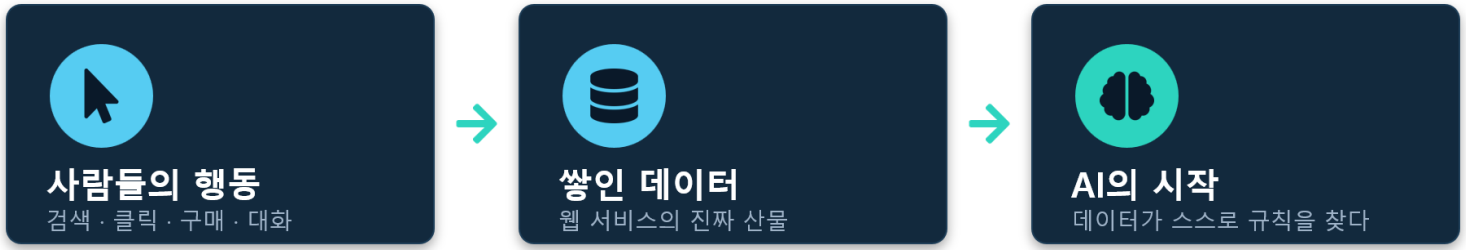
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

1 Introduction

데이터의 시대

웹이 데이터를 모으다.

- 인터넷과 스마트폰이 퍼지자 검색·클릭·구매·대화가 모두 서버에 쌓였고, 웹 서비스의 진짜 산물은 '데이터'가 되었다.





데이터를 이용한 기존의 연구분야

- 통계학: 회귀분석, 분류, 가설검정, 시계열 분석
- Data Science/Mining: 패턴 발견, 예측, 군집화, 추천
- 기계학습(ML)/인공지능(AI): 예측 모델, 분류 모델, 딥러닝

- 분석

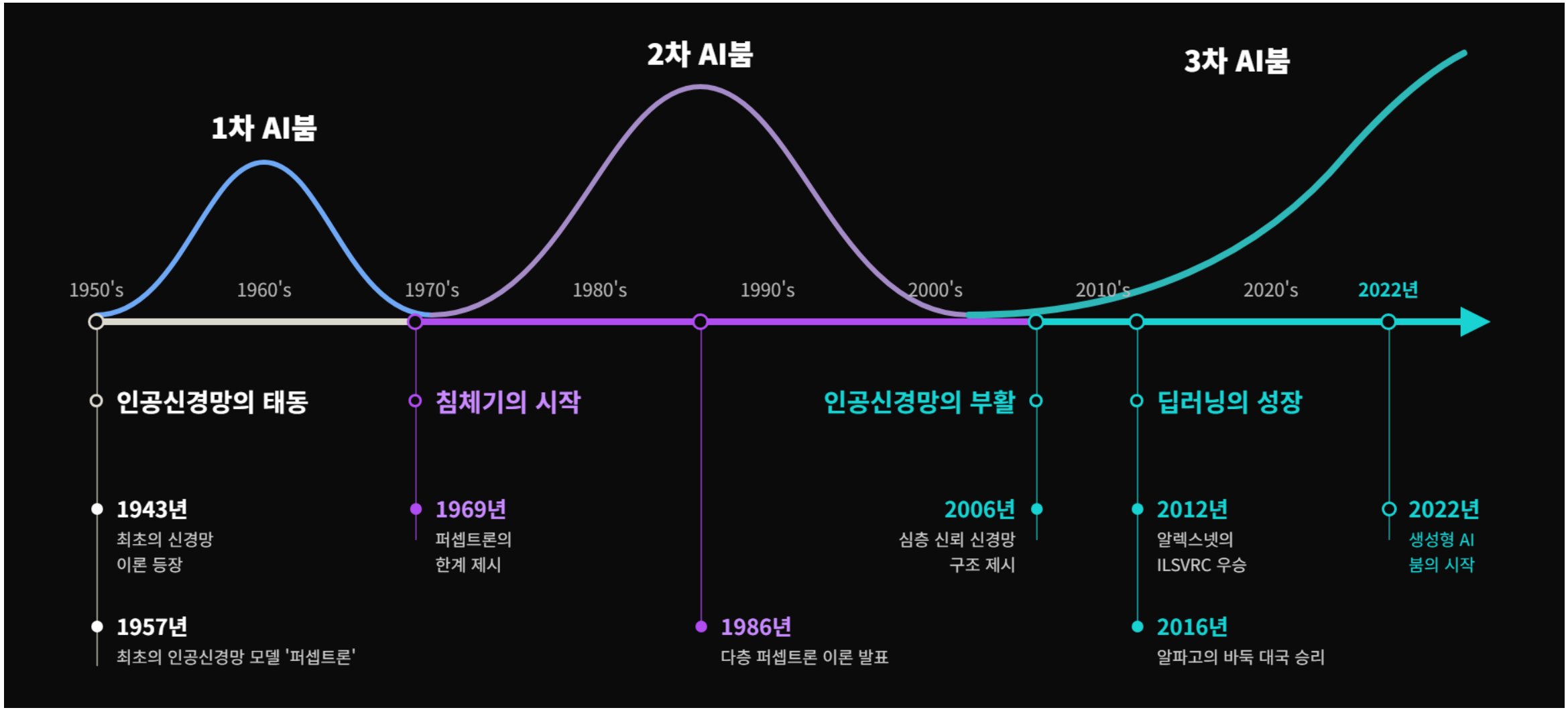
- 데이터베이스/분산컴퓨팅: 대용량 저장, 처리, 질의 최적화 - 관리
- 정보시스템/비즈니스 분석: 의사결정 지원, 프로세스 개선, 데이터 거버넌스 - 활용

규칙을 사람이 다 짜기엔 데이터가 너무 많고 복잡해졌다. 그래서 발상을 뒤집는다.
- 규칙을 짜는 대신, 데이터로부터 규칙을 스스로 찾게 하자. **이것이 AI다.**

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

1 Introduction

AI: 두번의 겨울과 딥러닝



Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

어쩌다 LLM의 시대가 되었나?

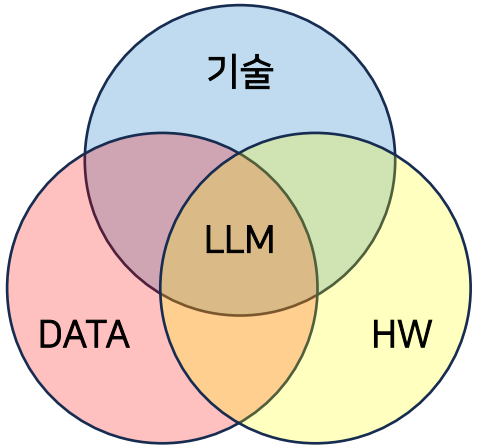
- **"너 인공지능 하니?"**
  - 3년 전, "저 그런거 할 줄 모르는데요."

CNN RNN
LSTM

현재, "ChatGPT 사용하고 있습니다."

- **기술, HW, 데이터 덕분**
  - 인터넷·통신 인프라의 확산으로 대규모 텍스트 데이터가 지속적으로 축적되었고, 이것이 LLM 학습의 핵심 자원이 됨.
  - 대규모 데이터를 수집·저장·분산처리 하는 기술이 발전되면서, 글로벌 규모의 학습 파이프라인이 현실화됨.
  - GPU·NPU 와 같은 AI 가속기의 비약적 발전으로 대규모 신경망 학습 시간이 크게 단축되었고, 수십억~수천억 파라미터 모델 학습이 가능해짐.
  - 메모리 용량, 대역폭, 분산학습 기술의 향상은 초대형 모델을 여러 장비에 나누어 학습·추론할 수 있게 하여 LLM 상용화의 기반이 됨.
  - 결과적으로 "데이터를 활용하는 기술"과 "그 데이터를 학습시킨 하드웨어"가 동시에 발달하면서, 오늘날의 LLM 시대가 본격화됨.







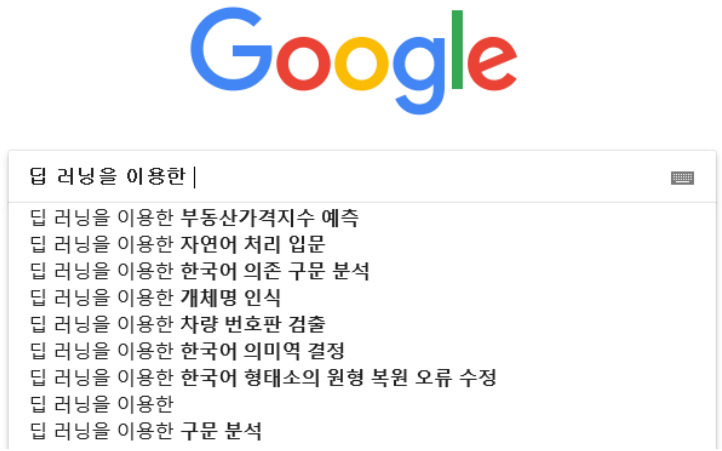
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

언어모델(Language Model, LM)

언어 모델(LM)은 가장 자연스러운 단어 시퀀스를 찾아내는 모델 : 이전 단어들로 다음 단어 예측

- 통계적 모델(SLM): 단어 출현 빈도를 계산하여 확률적으로 다음 단어를 예측 (N-gram 기반, 희소 문제 한계, 1980년대 ~ 2000년대 초)
- 신경망 모델(NLM): 단어를 벡터로 표현하고 RNN/LSTM으로 문맥을 기억하기 시작 (Word2Vec, 장기 기억 한계, 2000년대 초 ~ 2010년대 중반)
- 사전학습 모델(PLM): Attention 메커니즘과 Transformer 구조를 통해 병렬 학습과 깊은 문맥 이해 실현 (BERT, GPT, 2017년 ~ 2019년)
- 거대 모델(LLM): 모델 크기가 임계점을 넘으며 창발적 지능이 발현되고 스스로 사고하는 에이전트 단계로 진화 (GPT-3, DeepSeek, 2020년 ~ 현재).



[Attention is All You Need, 2017] 에서 Transformer 모델 소개

- Transformer의 활용 구조에 따라 3가지 개발 방향이 발생함.

구조	특징	대표 모델	용도
Encoder-only	문맥을 양방향으로 파악(앞뒤 단어 모두 참조)	BERT, RoBERTa	텍스트 분류, 감성 분석, 개체명 인식
Encoder-Decoder	입력을 이해하는 부분과 출력을 생성하는 부분이 나뉨	T5, BART	기계 번역, 요약, 질문 답변(QA)
Decoder-only	이전 단어들을 보고 다음 단어를 생성(한 방향)	GPT, Llama, Claude	범용 생성형 AI, 챗봇(현재 주류)

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

GPT(Generative Pre-trained Transformer)

Attention을 통해

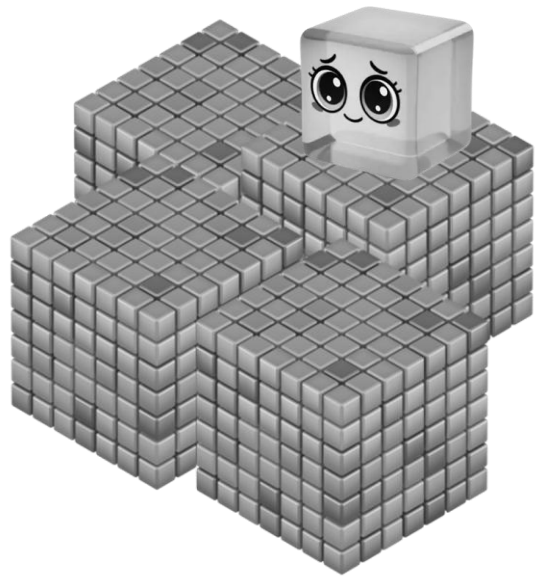
- 이제 말귀를 알아듣네.
- 이제 맥락(Context)을 이해하는구나.

→ 결국, 눈치 빠른 Matrix

```
Parameter containing:
tensor([[  0.4223,   1.0829,  -0.7585,  ...,   1.3011,  -0.2246,  -0.6516],
        [  3.3340,  -0.5697,  -1.1163,  ..., -13.4995,  -3.8654,  -1.6603],
        [  0.1018,   0.1551,  -3.5475,  ...,   5.5360,  -0.5984,   3.6319],
        ...,
        [  0.0933,   0.9214,  -0.1964,  ...,   4.7361,  -1.0311,   0.7439],
        [  1.5408,   0.3737,  -1.5125,  ...,   4.6930,  -0.6784,   1.1959],
        [  0.6573,   0.2671,  -2.0278,  ...,   1.8184,   2.6442,  -0.4709]],
       device='cuda:0', requires_grad=True)
Parameter containing:
tensor([[ 0.0003, -0.0174,  0.0091,  ..., -0.0149, -0.0151,  0.0066],
        [ 0.0276,  0.0565, -0.0226,  ...,  0.0458,  0.0060,  0.0606],
        [-0.0084,  0.0156,  0.0493,  ...,  0.0095,  0.0267,  0.0108],
        ...,
        [ 0.0064, -0.0641, -0.0178,  ...,  0.0659,  0.0482, -0.0747],
        [-0.0164, -0.0333,  0.0448,  ...,  0.0113,  0.0405,  0.0243],
        [-0.0396,  0.0035, -0.0345,  ..., -0.0434,  0.0033, -0.0482]],
       device='cuda:0', requires_grad=True)
```

GPT(Generative Pre-trained Transformer)의 등장

GPT-1 (2018.06)	Transformer 디코더를 대규모 비라벨 텍스트로 사전학습한 뒤 소량 라벨 데이터로 미세조정하는 반지도 전이학습 방식을 처음 체계화한 1.17억 파라미터 모델
GPT-2 (2019.02)	모델을 1.5B까지 키워 별도 파인튜닝 없이 사전학습만으로 여러 NLP 작업을 zero-shot으로 수행할 수 있음을 보인 모델
GPT-3 (2020.05)	175B 규모로 확장해 프롬프트에 예시 몇 개만 넣는 few-shot(in-context learning)만으로 다양한 작업이 가능함을 입증하며 프롬프트 엔지니어링 패러다임을 연 모델
InstructGPT (2022.01)	GPT-3를 SFT → 보상 모델 학습 → PPO 강화학습의 3단계 인간 피드백(RLHF)으로 미세조정해 지시 따르기에 최적화한 모델
ChatGPT (2022.11.30)	GPT-3.5를 RLHF와 대화 데이터로 다시 튜닝해, 전문가용 API였던 LLM을 누구나 대화형으로 쓸 수 있는 대중 서비스로 바꾼 전환점



전후사정(Context)을 잘 알려줘야 이해한다.

- 프롬프트에 전후사정을 잘 담아라.
- 한정된 공간에 전후사정을 어떻게 담을까?
- 모델이 틀려도 시스템이 잡아내도록 환경을 짠다.

→ Prompt Engineering
→ Context Engineering
→ Harness Engineering

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

파운데이션 모델과 서비스

- ChatGPT, OpenAI
- Gemini, Google
- Claude, Anthropic
- Muse Spark, Meta
- Grok, xAI
- Gemma, Google
- DeepSeek, DeepSeek
- Qwen, Alibaba
- Mistral, Mistral AI
- Kimi, Moonshot AI
- GLM, Z.ai
- Command, Cohere
- ERNIE, Baidu
- Hunyuan, Tencent

Muse Spark by Meta

Tencent Hunyuan

Mistral AI

NVIDIA NEMOTRON Revoluoizing / AI Agents for Enterprises

ERNIE



Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

누가누가 잘하나?

26 Oct '26, 현재

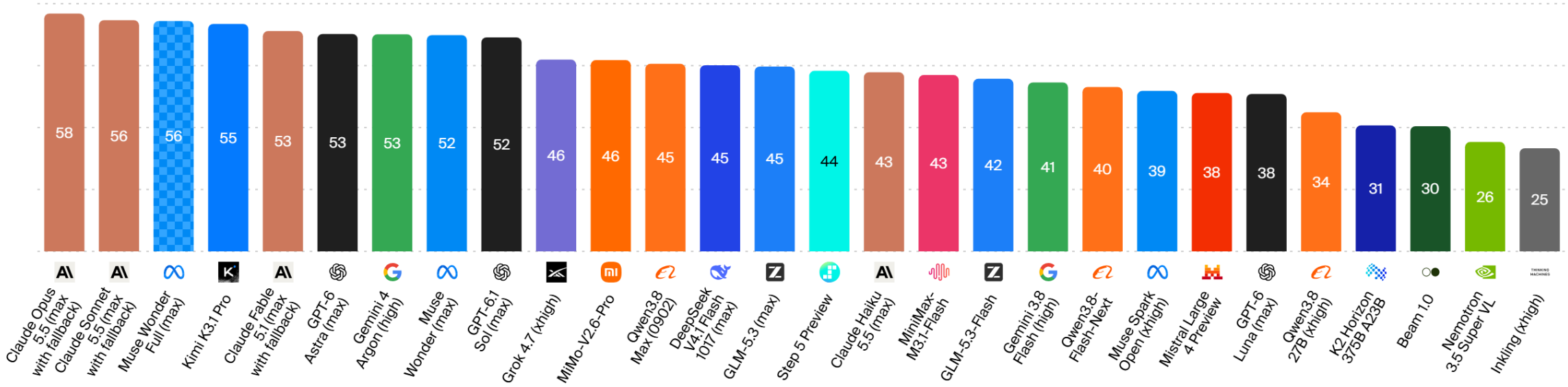
Artificial Analysis Intelligence Index

28 of 708 models

Artificial Analysis Intelligence Index v4.3.2 incorporates 10 evaluations: AA-Briefcase v1.1, GDPval-AA v2.1, AutomationBench-AA, Terminal-Bench 4.0, SciCode, Humanity's Last Exam, GDP.pdf, CritPt, AA-Omniscience, AA-LCR v1.1

No publicly available API

Artificial Analysis



Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

우리에겐 지켜야 하는 것들이 있다. <보안>



[단독] 우려가 현실로…삼성전자, 챗GPT 빗장 풀자마자 '오남용' 속출

정두용 기자
입력 2023.03.30 18:20 수정 2023.03.30 20:35

가 글자크기 공유 인쇄 URL복사 댓글

반도체 사업장 챗GPT 허용 20일, 정보 유출 사고 3건 발생
설비 계측·수율 데이터, 미국 기업에 고스란히 전송…회수 불가
필요시 해당 임직원 징계…사내 전용 AI 서비스 구축 검토



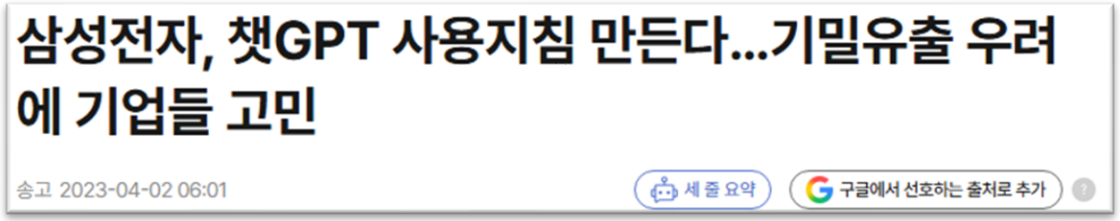
[단독] 삼성전자, 챗GPT·제미나이 사내 사용 전면 허용 추진

배지현 기자 · 2026. 5. 25. 19:01



삼성전자, 외부 생성형 AI 도입으로 AX 본격화

2026/06/11



삼성전자, 챗GPT 사용지침 만든다...기밀유출 우려에 기업들 고민

송고 2023-04-02 06:01 세 줄 요약 구글에서 선호하는 출처로 추가



삼성, 챗GPT 데이터 유출 후 임직원 AI 사용 금지

임대준 기자 입력 2023.05.03 12:46 | 댓글 0 구글 검색 선호 출처로 추가 가

다만, 삼성전자 경영진은 외부 데이터센터에 연결된 클라우드 기반의 인공지능 모델은 정보 유출 우려가 있는 만큼, 자체적인 안전장치를 마련하기로 했다. 사내 전산망에서만 인공지능 모델을 구동하거나, 외부 인공지능 모델의 학습에 직원들이 입력한 정보를 가져다 쓰지 못하도록 제한을 두는 방식 등이다.
국가 핵심 기술로 분류되는 정보는 인공지능 사용 범위에서 제외할 것으로 보인다.

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

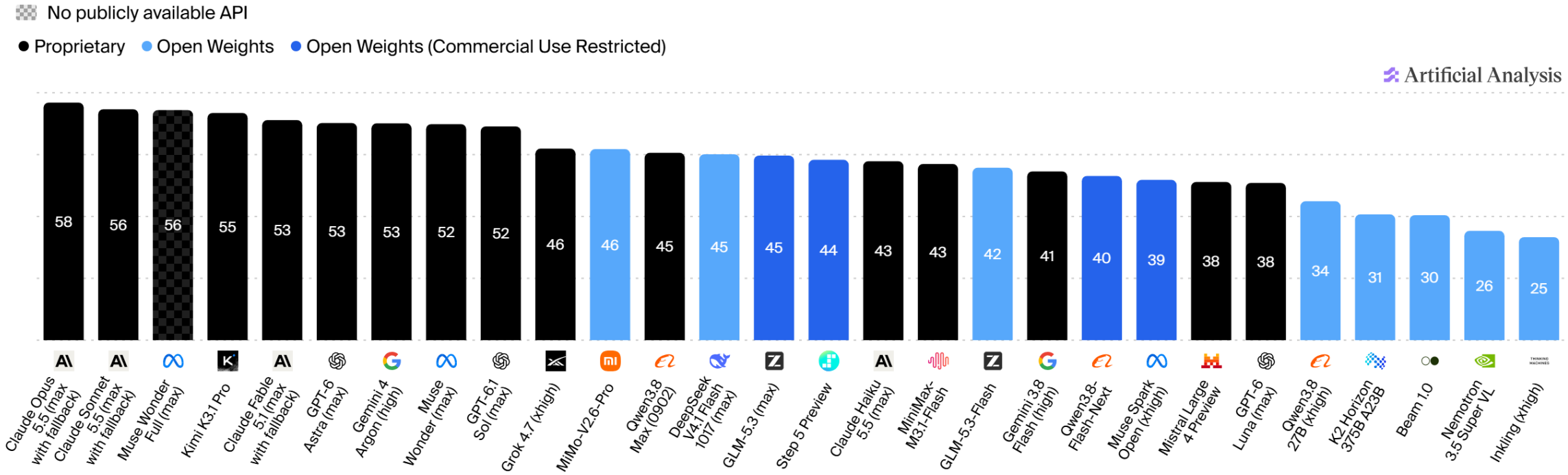
보안이슈로 오픈웨이트 모델 검토

26 Oct '26, 현재

Artificial Analysis Intelligence Index by Open Weights / Proprietary

28 of 708 models

Artificial Analysis Intelligence Index v4.3.2 incorporates 10 evaluations: AA-Briefcase v1.1, GDPval-AA v2.1, AutomationBench-AA, Terminal-Bench 4.0, SciCode, Humanity's Last Exam, GDP.pdf, CritPt, AA-Omniscience, AA-LCR v1.1



Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

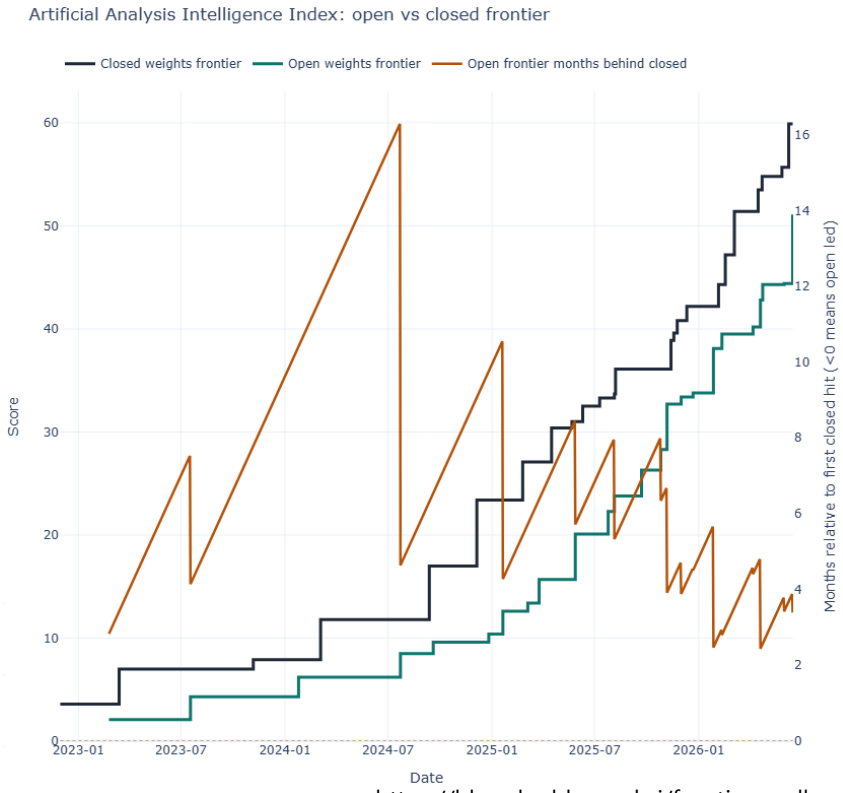
Open Weight Model 선정

Open Weight Model vs Closed Weight Model

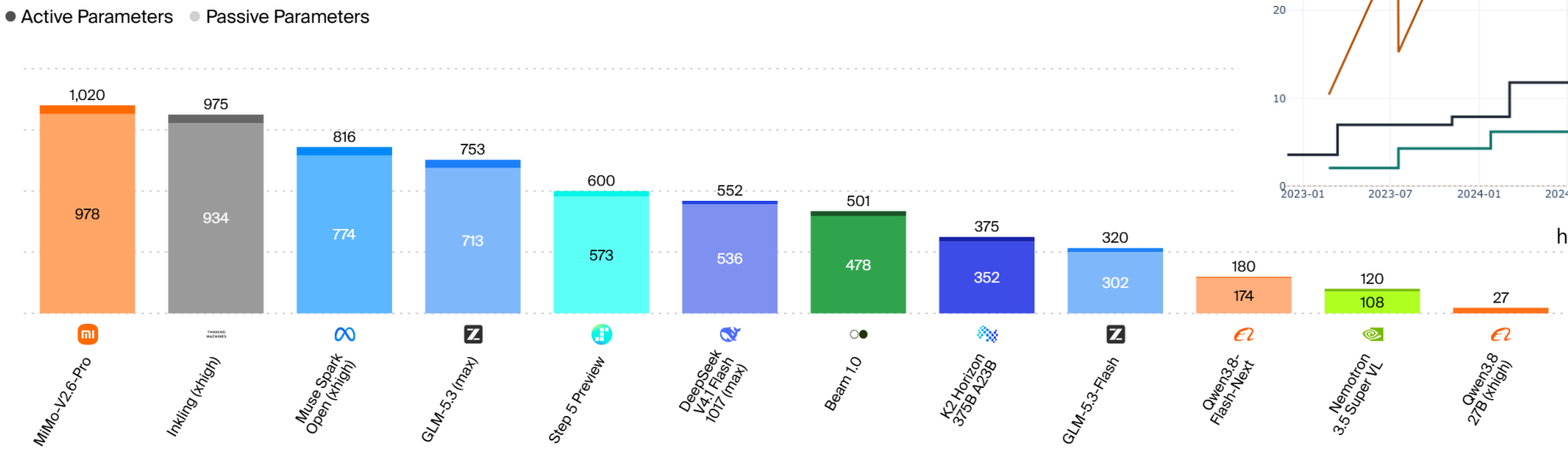
- AAII 지수로 5-6 개월의 성능차이 발생

Open Weight Model 로 Local LLM 구축 시 고려사항

- Hugging Face 등에서 모델 가중치를 내려받아 자체 서버나 개인 PC에서 실행
- 외부 API 의존을 줄이고, 사내 데이터 통제·커스터마이징·온프레미스 운영 가능
- 채팅·요약·번역뿐 아니라, 사내 문서 검색(RAG), 업무 자동화, 도메인 특화 미세조정, 에이전트 시스템 등 필요
- 양자화, 모델크기, Context Window, 사용자 수를 고려한 HW spec 산정



https://blog.doubleword.ai/frontier-os-llm

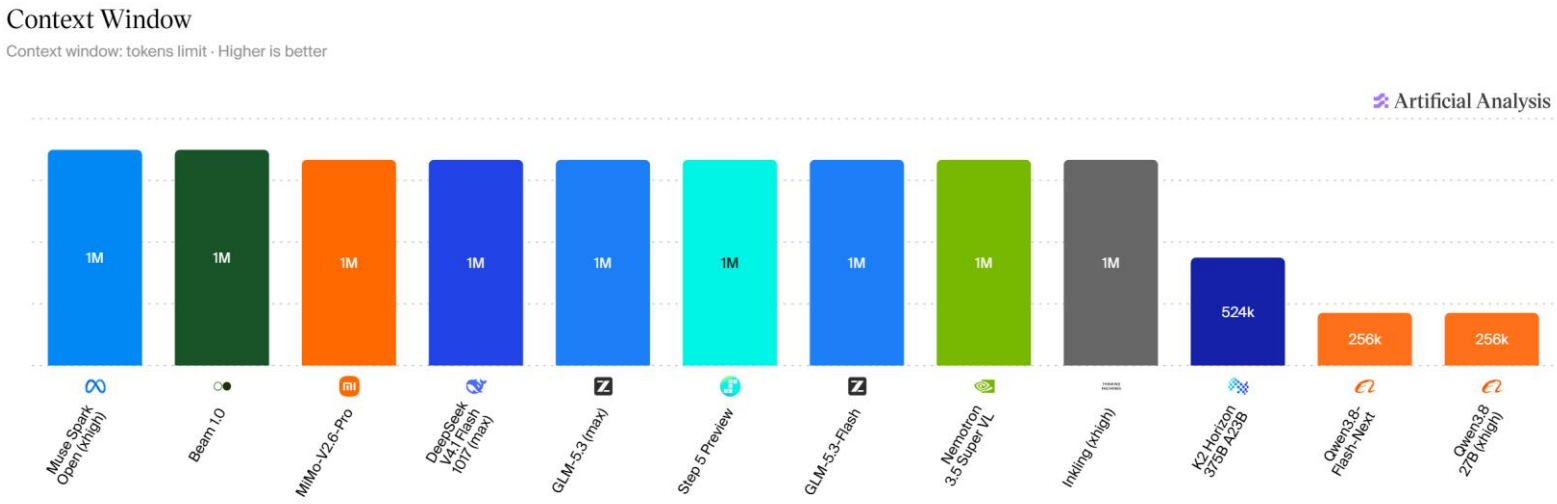


Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

2 Large Language Models

Context Window

- **LLM이 한 번의 요청에서 동시에 처리할 수 있는 토큰의 최대 총합**
  - 시스템 프롬프트, 사용자 입력 (대화 이력, 지시문, 첨부 자료 등), 그리고 모델이 생성하는 출력 토큰을 모두 합한 길이
  - 사용자가 쿼리를 계속 보내면 컨텍스트가 쌓이고 KV cache도 함께 커지며, 최대 길이에 도달하면 보통은 가장 오래된 토큰부터 잘라내거나 요약해서 유지
- **VRAM 사용량 증가의 원인**
  - VRAM 사용량 = 모델 메모리 + Context Length + 임시버퍼
  - 모델 메모리는 고정되더라도 Context Length는 동시 사용자 수만큼 필요
  - LLM 서비스에서 Long Context(1M) 지원을 위해 고용량 HBM/메모리를 더 많이 요구



모델	개발사	Context Window
Grok 4.20	xAI	2,000,000 토큰
GPT-5.6 Sol	OpenAI	1,050,000 토큰
Gemini 3.1 Pro	Google	1,048,576 토큰
Claude Opus 5	Anthropic	1,000,000 토큰
Claude Sonnet 5	Anthropic	1,000,000 토큰
Claude Opus 4.8	Anthropic	1,000,000 토큰
GPT-4.1	OpenAI	1,000,000 토큰
Claude Opus 4	Anthropic	200,000 토큰
GPT-4o	OpenAI	128,000 토큰

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

3 Local LLM Deployment

LLM으로 누가 돈을 벌었나?

인프라 계층

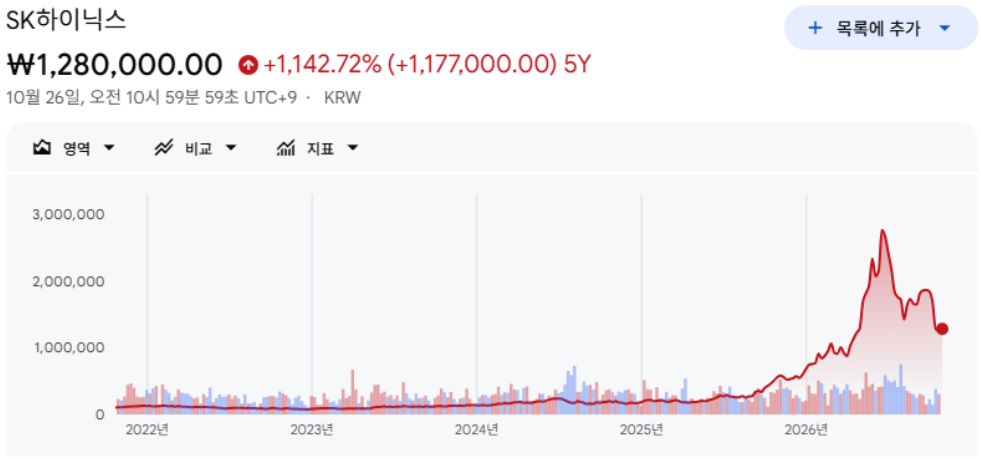
- AI를 물리적으로 떠받치는 기반
- GPU 설계(NVIDIA), 반도체 제조(TSMC), 클라우드(AWS·Azure)
- 현재까지 AI 붐에서 이익을 가장 확실하게 가져간 계층

빅테크 업체의 투자규모

- 2025년 4대 빅테크(MS·알파벳·아마존·메타)의 AI 관련 자본지출(CapEx)은 약 3200억 달러 수준
- 2026년에는 이 규모가 더 커져, 4사 합산 AI 인프라 투자 추정치가 약 6,650억~8,300억 달러



AI데이터센터 : GPU서버 + 저장 + 냉각 + 전력 인프라





Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

3 Local LLM Deployment

NVIDIA GPU

Graphic Processing Unit: AI 가속기

- 화면에 그림, 3D 장면, 영상 같은 것을 빠르게 그려 주는 그래픽 전용 장치
- CPU가 복잡한 작업을 순차적으로 빠르게 처리하는 데 강하다면,
  GPU는 같은 종류의 계산을 대량으로 반복하는 작업을 동시에 처리

구분	용도	대표 제품군
GeForce	개인용, 게임, 일반 그래픽	RTX 5090, RTX 5080, RTX 5070 Ti, RTX 5070, RTX 4090, RTX 4080, RTX 4070 Ti, RTX 4070, RTX 4060
RTX PRO / Quadro 계열	워크스테이션용, CAD·3D·영상 편집	RTX PRO 6000, RTX PRO 5000, RTX PRO 4500, RTX PRO 4000, RTX A6000, RTX A5000, RTX A4000, Quadro RTX 8000, Quadro RTX 6000, Quadro RTX 5000
Data Center / AI 계열	서버, AI 학습, 추론, 데이터센터	A100, H100, H200, B200, A800, L40S, L4, V100, T4









Pascal → Volta → Turing → Ampere → Hopper → Blackwell → Rubin(2026)

- P100 → V100 → T4 → A100 → H100 → B200

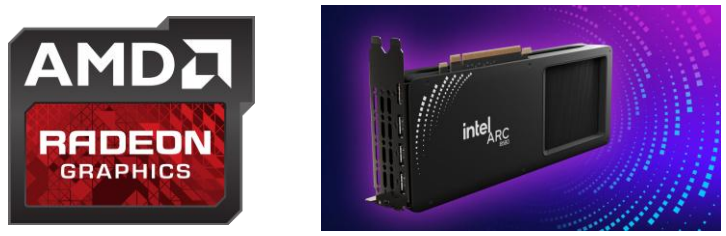
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

3 Local LLM Deployment

NVIDIA GPU

GPU 생산업체

- NVIDIA: GeForce, RTX PRO, Data Center GPU.
- AMD: Radeon, Instinct.
- Intel: Arc, 내장 그래픽도 강함.



왜 NVIDIA 인가?

- CUDA 생태계: GPU를 그래픽 전용 장치가 아닌 범용 병렬 연산 장치로 사용하기 위해 CUDA(2006)를 기반으로 cuDNN(2014), TensorRT(2017년경) 등의 스택을 단계적으로 확장
  - 학계의 GPGPU(범용 GPU 연산) 실험 + 엔비디아의 조기 투자/영입 → 2006년 CUDA 출시.
  - "C/C++ 개발자가 GPU 병렬 연산을 비교적 직관적으로 활용할 수 있게 하자"
  - AI·HPC 개발자들이 이미 NVIDIA 환경에 깊게 묶여 있음(Lock-In)
- 압도적 점유율: GPU 점유율이 80~90%대라, 업계 대부분이 NVIDIA를 기본으로 사용함
- 하드웨어+소프트웨어 패키지: 칩만이 아니라 드라이버, 라이브러리, 레퍼런스 서버, 도구까지 묶어서 제공해 그냥 사서 쓰면 됨

항목	AMD MI300X	NVIDIA H100
HBM	192GB HBM3	80GB HBM3
메모리 대역폭	5.3TB/s	3.35TB/s
하드웨어 가격	$10k~15k	$25k~35k
8-GPU 플랫폼	$160k~200k	$310k~340k
VRAM당 비용	$120/GB	$500/GB

XPU : 추격자들…

- TPU (Tensor Processing Unit) : 구글이 만든 딥러닝 전용 칩
- NPU (Neural Processing Unit) : 스마트폰, PC, 자동차 같은 온디바이스/엣지에서 저전력 신경망 추론
- LPU (Latency Processing Unit) : Groq 같은 회사가 만든, 초저지연 실시간 추론에 특화된 AI 가속기

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

3 Local LLM Deployment

Open Weight Model 구동(Inference)을 위한 GPU

모델 별 GPU 수량 산정 (NVIDIA H100 VRAM 80GB 기준)

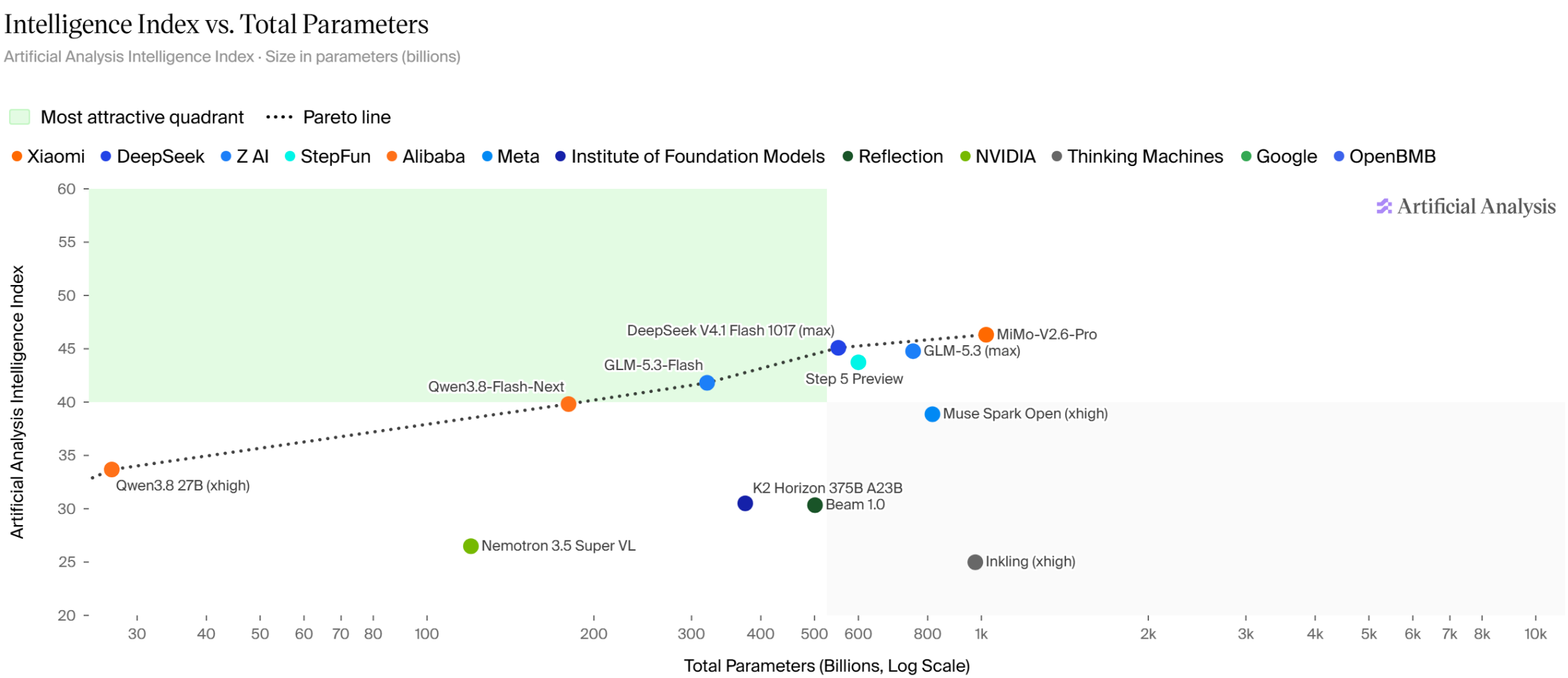
오픈모델	AAII	총 파라미터	활성 파라미터	FP16 가중치 용량	이론상 최소 H100	권장 H100 수량*	INT4 가중치**
Kimi K3	44	2,800B	104B	5,600GB	70장	72~80장	18~20장
MiMo-V2.6-Pro	46	1,020B	42B	2,040GB	26장	28~32장	7~8장
Inkling (xhigh)	25	975B	41B	1,950GB	25장	28장	7장
GLM-5.3 (Max)	45	753B	40B	1,506GB	19장	20~24장	5~6장
Step 5 Preview	44	600B	27B	1,200GB	15장	16장	4장
DeepSeek V4.1 Flash 1017	45	552B	16B	1,104GB	14장	16장	4장
DeepSeek V4.1 Flash	38	552B	16B	1,104GB	14장	16장	4장
Nemotron Ultra	23	550B	55B	1,100GB	14장	16장	4장
MiniMax-M3	29	428B	23B	856GB	11장	12장	3장
K2 Horizon 375B A23B	31	375B	23B	750GB	10장	12장	3장
GLM-5.3 Flash	42	320B	18B	640GB	8장	8~10장	2~3장
Qwen3.8 Flash-Next	40	180B	6B	360GB	5장	6장	2장
Nemotron 3.5 Super VL	26	120B	12B	240GB	3장	4장	1장
Qwen3.8 27B	34	27B	27B	54GB	1장	1장	1장

* "권장 H100 수량"은 KV cache와 실행 여유 공간을 고려한 실무적 추정치, **INT4로 양자화한 용량을 고려한 권장 H100 수량

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

3 Local LLM Deployment

Model Size vs AAII



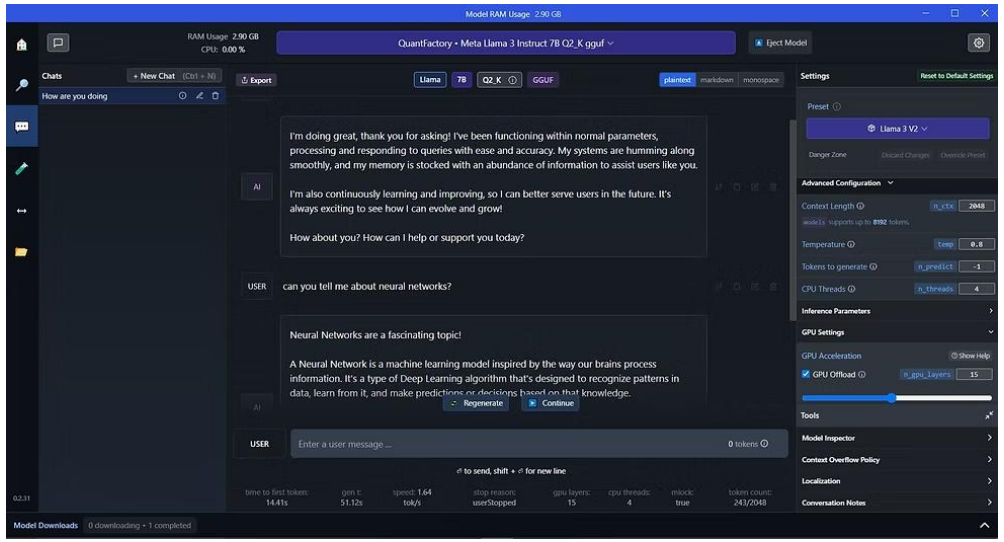
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

3 Local LLM Deployment

쉽게 쉽게 해보자.(개인)

LM Studio

- 모델 다운로드부터 채팅, API 서빙, 플러그인 확장까지 GUI 하나에서 해결하는 로컬 LLM 데스크톱 워크벤치

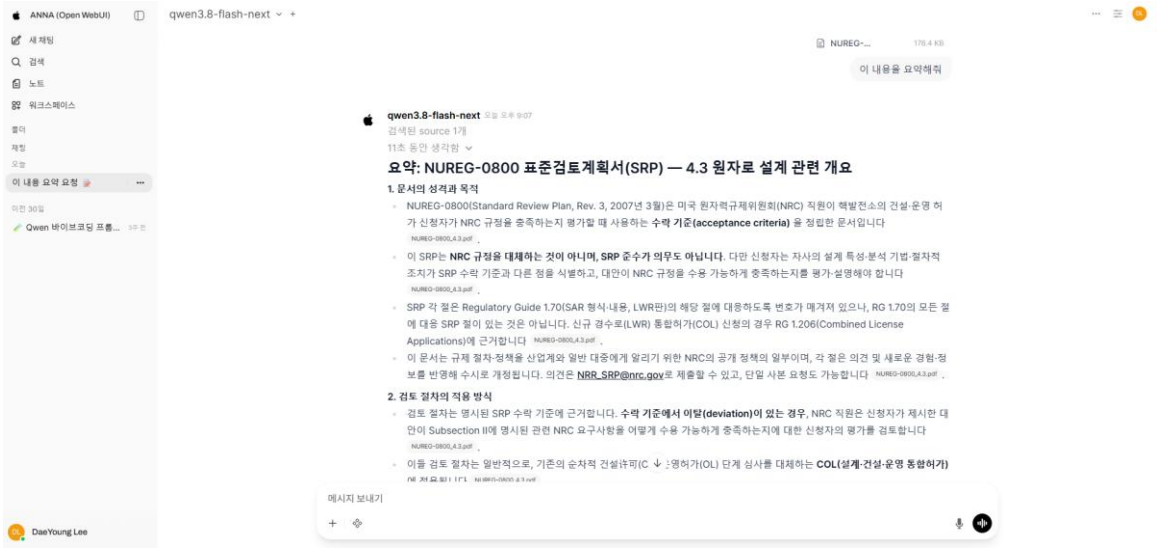


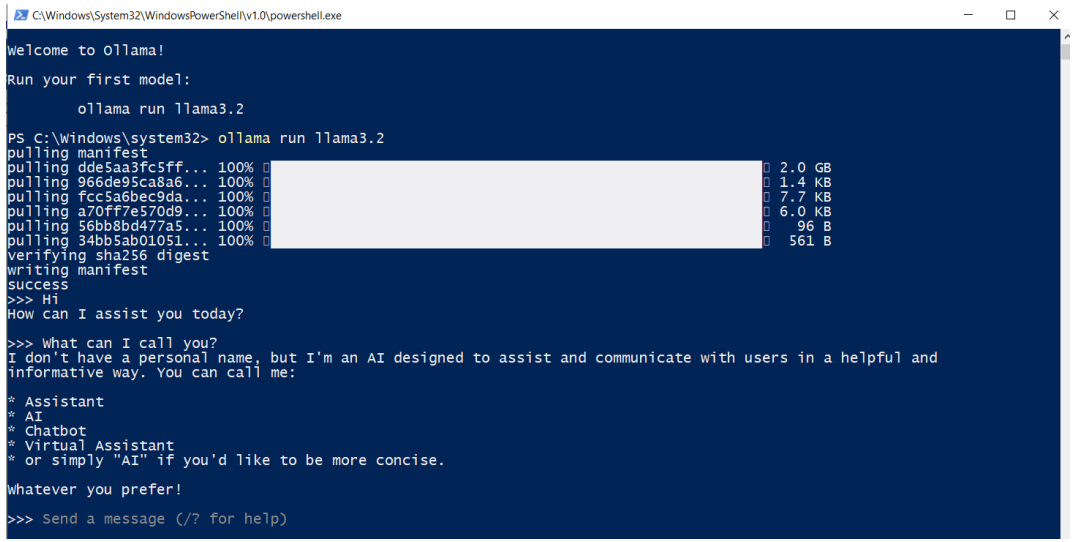
Ollama

- 명령 한 줄로 모델을 받아 백그라운드 서버로 띄우는 오픈소스 런타임으로, 다른 도구들이 붙는 표준 백엔드 역할

Open WebUI

- 모델을 직접 돌리지 않고 Ollama·vLLM 등 서빙 엔진 앞에 붙여 ChatGPT 같은 웹 채팅 화면과 사용자·문서 관리를 제공하는 오픈소스 프론트엔드





Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

3 Local LLM Deployment

쉽게 쉽게 해보자.(팀단위)

Hugging Face → vLLM → Open WebUI

- Hugging Face : AI 모델·데이터셋·데모 앱을 모아둔 오픈소스 허브 + 라이브러리, AI계의 GitHub
- vLLM : 대규모 언어 모델(LLM)의 추론(Inference) 성능을 극대화하기 위해 설계된 오픈소스 엔진

Open WebUI 기본 설정 점검

- 채팅 템플릿: 채팅 템플릿이 그 모델이 학습된 형식과 일치하지 않으면 답변 끊김·반복이 발생
- 생성 파라미터: temperature, top_p, repeat_penalty(다음 토큰을 고를 때의 확률 분포를 조정하는 샘플링 파라미터)를 모델 카드의 generation_config.json 권장값에 맞춰 Open WebUI 모델 설정
- 컨텍스트 길이: vLLM --max-model-len과 Open WebUI num_ctx를 충분히 크고, 서로 같게 조정
- 추론 모델이면 --reasoning-parser를 켜서 thinking 내용이 답변에 유입되는 것 방지

시스템 프롬프트 설계

- Open WebUI는 기본 시스템 프롬프트가 비어 있음.
- 역할과 페르소나
- 응답 언어(한국어 고정)와 어투
- 마크다운·목록 사용 규칙, 답변 길이 기준
- 모르는 내용일 때의 행동(추측 금지, 확인 요청 등)
- 금지 사항과 거절 방식

더 똑똑해지고 있는데, 막상 사용해보면…





Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

3 Local LLM Deployment

완성된 LLM에 새로운 지식을 주입하는 방법

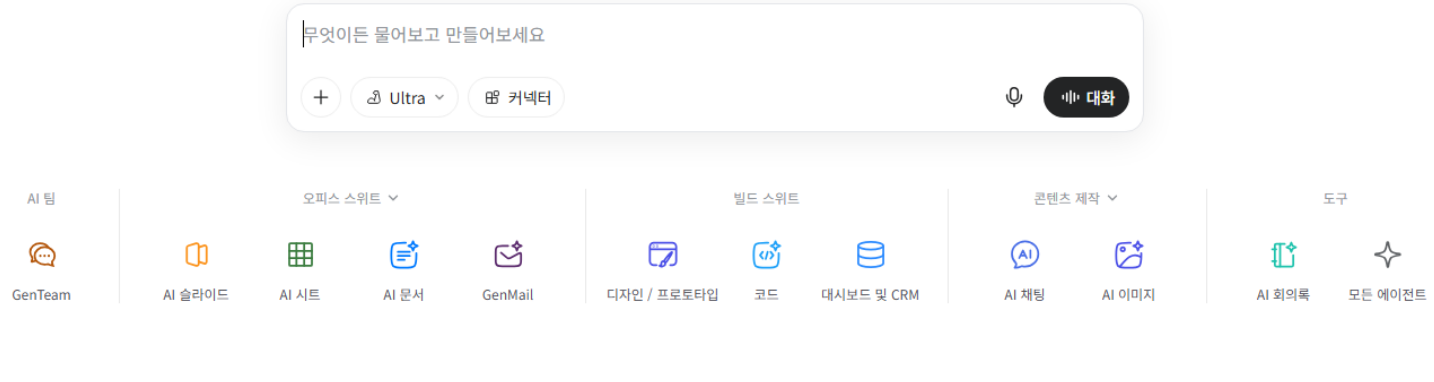
구분	세부 기법	작동 원리	주요 장점	주요 한계점
외부주입 방식 (가중치 고정)	① 인컨텍스트 러닝 (In-Context Learning / Long-Context) 	프롬프트 입력창(Context Window)에 참고할 **문서나 지식을 통째로 넣어**주고 답변을 생성하게 한다.	• 별도의 인프라 구축 불필요 • 즉각적인 테스트 및 적용 가능	• 질의마다 막대한 토큰 비용 발생 • 입력 길이가 길어질수록 추론 속도 저하 및 **중간 내용 망각**(Lost in the middle) 발생
	② 검색 증강 생성 (RAG) 	외부 벡터 DB 및 검색 엔진에 지식을 색인해 두고, 질문과 관련된 내용만 **실시간 검색하여 프롬프트에 결합**한다.	• 최신 정보의 실시간 업데이트 용이 • 출처(근거 문서) 제시를 통한 환각 억제 • **가장 높은 비용 효율성**	• **검색 품질(Retrieval)**이 낮으면 최종 답변도 실패함 • 도메인 고유의 어투나 복잡한 추론 논리 자체를 습득시키기는 어려움
모수적 방식 (가중치 추가)	**③ 지도 미세조정 (SFT / PEFT·LoRA)**	**'질문-답변'** 형태의 정제된 데이터셋을 사용하여 모델의 **가중치 일부(LoRA 등) 또는 전체(Full Fine-Tunning)를 업데이트**한다.	• 도메인 특화 용어, 출력 포맷, 답변 스타일 교정에 탁월함 • 추론 시 긴 프롬프트가 필요 없어 응답 속도가 빠름	• **새로운 지식(Fact)을 주입시키는 데는 취약함** (오히려 기존 지식과 충돌해 환각이 증가할 수 있음) • 지식 변경 시 재학습 필요
모수적 방식 (가중치 수정)	④ 연속 사전학습 (CPT, Continued Pre-Training)	레이블이 없는 대규모 도메인 원문 텍스트(수억~수십억 토큰)를 모델에 그대로 추가 학습시켜 **기반 지식 자체를 확장**한다.	• 원자력, 법률, 의료, 특수 프로그래밍 언어 등 범용 모델이 전혀 모르는 **새로운 도메인 지식을 뇌(가중치)에 각인**시킬 수 있음	• 막대한 GPU 연산 비용(CAPEX) 소요 • 기존에 잘하던 범용 능력을 잃어버리는 **파국적 망각(Catastrophic Forgetting) 위험** 존재
	⑤ 직접 모델 편집 (Model Editing: ROME, MEMIT 등)	특정 사실 관계가 저장된 트랜스포머 내부의 특정 뉴런(FFN 레이어) 위치를 찾아내어 해당 가중치 값만 외과수술처럼 수정한다.	• 전체 재학습 없이 단 몇 초 만에 특정 사실(예: "현재 대통령은 OOO이다")만 정밀하게 수정 가능	• 수만 건 이상의 방대한 도메인 문서를 통째로 주입하는 용도로는 부적합하며, 다중 편집 시 모델 안정성이 저하됨

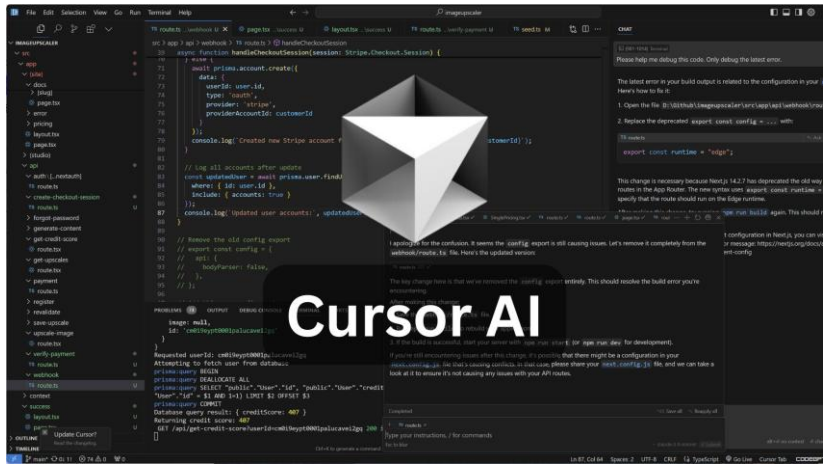
LLM 개발과정: PT → SFT → 강화학습

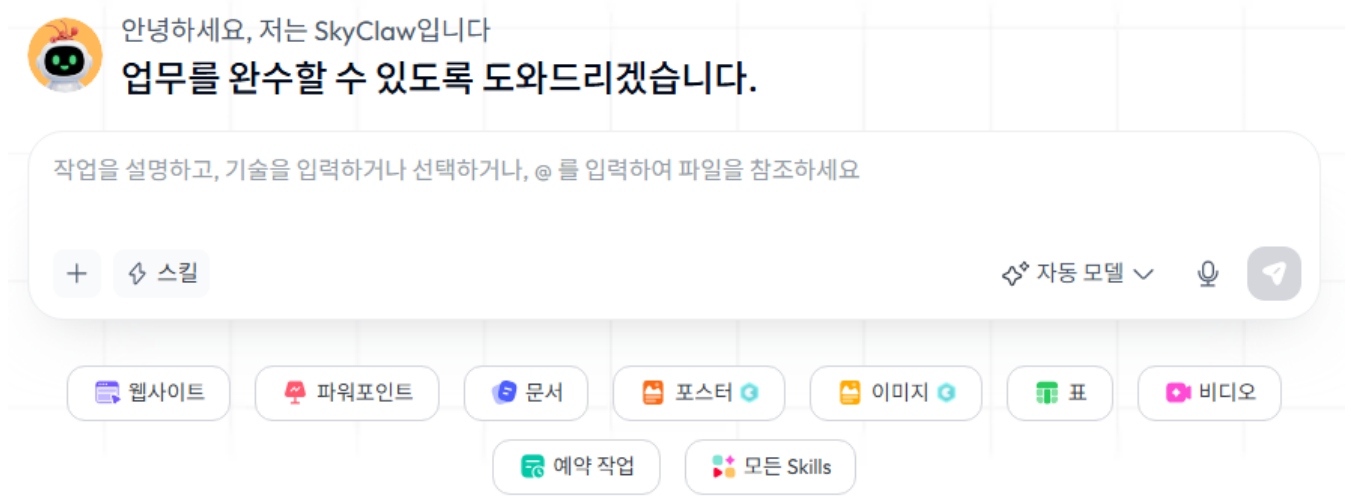
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

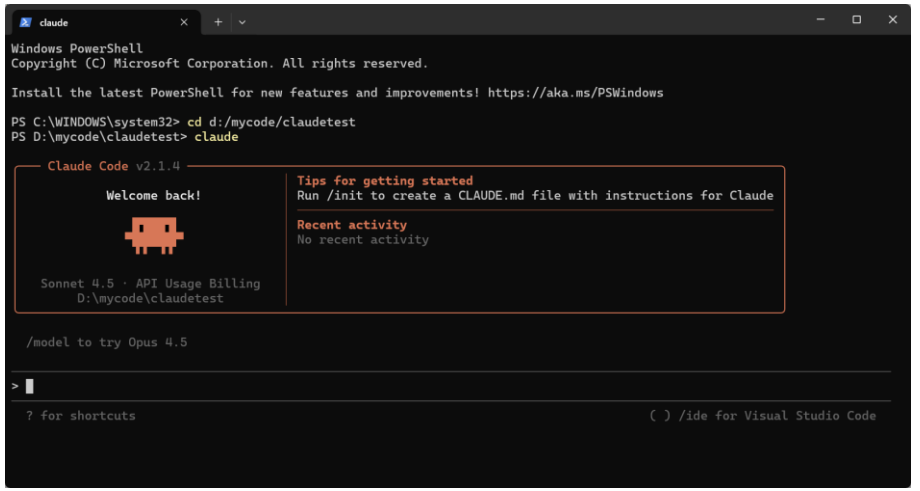
AI Agent 서비스

Genspark AI 워크스페이스 6.0









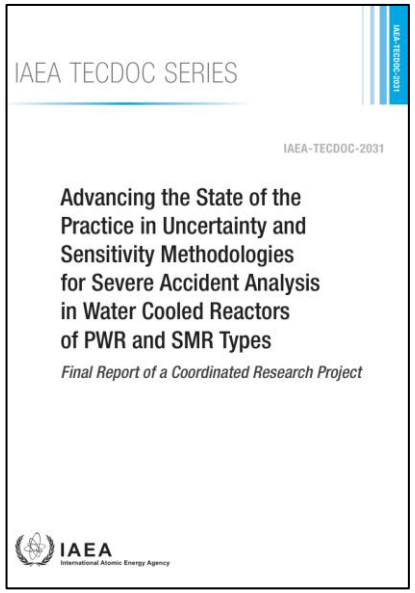
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

RAG(Retrieval-Augmented Generation)

TABLE 5. INPUT PARAMETERS AND RELEVANT PDFs

Institution	Source: code manual, literature survey, expert Judgment, and parametric analysis	
DNPER/PAEC (Pakistan)	MELCOR1.8.6 (**26 parameters**) (PDF type: triangular, log-triangular, normal, lognormal, uniform, log-uniform, beta, discrete): 1) Core melt progression (in-vessel) (17), 2) H_2 combustion in containment (1), 3) Aerosol/fission products release and transport (in-/ex-vessel) (7), 4) Heat transfer to concrete walls (containment) (1)	
ENRRA (Egypt)	ATHLET + SCALE6.3 (**one parameter**: coolant void reactivity): 1) Reactivities at different coolant densities (in-vessel) (14 multivariate Gaussian PDFs)	
KAERI+HYU (Republic of Korea)	MELCOR2.2 (**26 parameters**): 1) Core melt progression (in-vessel) (20), 2) Molten core-concrete interactions (MCCI) in reactor cavity (ix-vessel) (3), 3) H_2 combustion (containment) (1), 4) Fission products release and transport (in-/ex-vessel) (2), 5) Heat transfer to concrete walls (1)	MAAP5.05 (**29 parameters**): 1) Core melt progression (in-vessel) (18), 2) MCCI in reactor cavity (ex-vessel) (3), 3) H_2 combustion (containment) (2), 4) Fission products release and transport (in-/ex-vessel) (5), 5) Heat transfer to passive heat sinks (1)
	PDF type: Triangular, log-triangular, normal, lognormal, uniform, log-uniform, beta, discrete	
KINS+SNU (Republic of Korea)	MELCOR2.2/COOLAP2 (**13 parameters**) (PDF type: triangular, normal, lognormal, uniform): 1) MELCOR2.2 : Core melt progression (in-vessel) (5 parameters); 2) COOLAP2: MCCI in reactor cavity (ex-vessel) (8 parameters)	
SJTU (China)	MELCOR1.8.5 (**18 parameters**) (PDF type: normal, uniform, triangular): 1) Thermal power (1); Core melt progression (in-vessel) (17)	
University of Sharjah (UAE)	3KEYMASTER simulator (**44 parameters**) (PDF type: normal): 1) 44 groups SCALE covariance library for the cross-sections affecting on the fuel early phase fuel temperature	
CNEA (Argentina)	MELCOR1.8.6 (**11 parameters**) (PDF type: uniform, normal, lognormal): 1) Physical models, integrity criteria, and core relocation models (In-vessel) (10), 2) Accident management (1)	
ENSO (Spain)	RELAP/SCDAPSIM/MOD3.5 (**20 parameters**) (PDF type: uniform, normal, lognormal): 1) Physical models, boundary and initial conditions, material properties, and code correlations (in-vessel) (17), 2) Safety systems (3)	



IAEA_TECDOC-2031
PDF

한양대와 원자력연구원이 불확실성 분석을 위해 고려한 인풋변수들에 대해 설명해줘

분석

IAEA TECDOC-2031 보고서에 따르면, 한양대학교(KAERI 협력)와 한국원자력연구원(KAERI)이 중대사고 불확실성 분석을 위해 고려한 주요 인풋 변수들과 그 특성은 다음과 같다.

1. 한양대 및 원자력연구원(KAERI+HYU) 고려 변수

한양대학교와 KAERI는 OPR1000 원전을 대상으로 MELCOR 2.2와 MAAP5.05 코드를 사용하여 불확실성 분석을 수행하였으며, 각각 다음과 같은 인풋 변수들을 선정하였다. +2

- **MELCOR 2.2 관련 변수 (26개):** 노심 용융 진행(In-vessel core melt progression)과 관련된 20개 변수를 포함하여 총 26개의 매개변수를 고려하였다. +1
  - **노심 내 현상:** 입자형 파편물 기공률(PORDP), 파편물 등가 직경(DHYPD), 낙하 파편물 속도(VFALL), 용융물 냉각 열전달 계수(COR CHT), 연료 용해 비율(FUOZR), 복사 열전달 계수(FCELR & FCELA) 등이 포함된다. +2
  - **노외 및 기타:** 원자로 공동 내 용융물-콘크리트 상호작용(MCCI, 3개), 수소 연소(1개), 핵분열 생성물 방출 및 이동(2개) 관련 변수들이 포함되었다. +1

Gemini 3에게 물어보기

+ 도구 빠른 모델

- LLM서비스에 문서를 업로드 하면…
  - 문서 내용 Chunking → 벡터화
  - 사용자 Query 벡터화
  - 두 벡터의 유사도를 통해 관련내용 탐색
- 보통 문서 10개 이상 업로드 불가
- 수많은 문서를 DB화해서 빠르게 찾자.

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

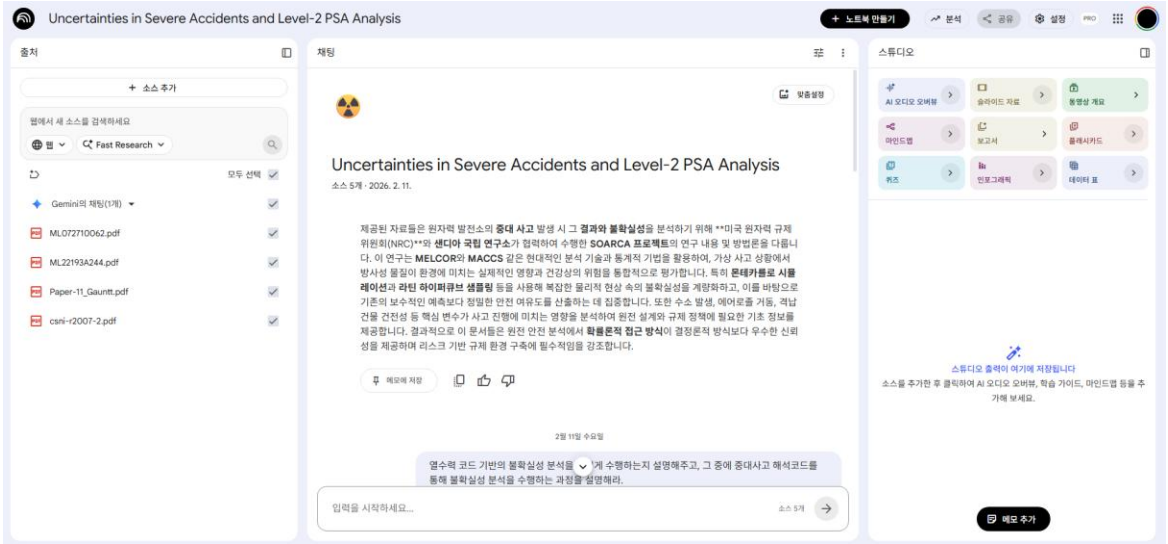
4 Building AI Agents

RAG(Retrieval-Augmented Generation)

NotebookLM, Google, 2024. 06.

- **사용자가 업로드한 자료만을 대상으로 하는 폐쇄형 RAG**
  - Google이 Gemini 모델 위에 만든 소스 기반(source-grounded) RAG 애플리케이션
  - 소스 유형: PDF, Google Docs, 웹 페이지, YouTube 자막, 음성 파일 등
  - 모든 질문·답변은 원칙적으로 등록된 소스 안에서만 수행
- **무료: 최대 50개 소스, 소스당 200MB/50만 단어.**
- **유료: 최대 300개 소스, 소스당 제한은 동일**
- **내부 동작 (RAG 파이프라인)**
  - 소스 업로드 → 문서를 작은 조각(chunk)으로 분할
  - 각 조각을 임베딩(수치 벡터)으로 변환해 색인
  - 질문도 벡터로 변환 → 의미가 가장 가까운 조각 검색
  - 검색된 조각을 프롬프트에 붙여 "이 자료에 근거해서만 답하라"고 지시
  - 답변 문장마다 인용 번호 부착, 클릭 시 원문 위치로 이동
  - "소스에 없으면 모른다고 답하라"는 강한 지시 → 환각 대폭 억제
- **생성 기능**
  - 학습 가이드, 브리핑 문서, 타임라인, FAQ, 마인드맵 자동 생성
  - 오디오 오버뷰: 소스 내용을 두 진행자의 팟캐스트 대화로 변환
  - 구조는 동일: 고정 프롬프트 템플릿 실행 + (오디오는) 음성 합성 모델 통과
- **정해진 자료 더미에서 정확하게 꺼내 쓰는 작업**





Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

RAG(Retrieval-Augmented Generation)

perplexity : 2022.08 설립, 기업가치 $30B, 직원수 300여명

자체 거대 모델 개발이 아닌, 검색을 중심에 둔 "답변 엔진(answer engine)"

- 아키텍처: 검색 증강 생성(RAG)을 도구가 아닌 파이프라인 전체 구조로 채택한 개방형 RAG

동작 흐름 (질문 1건 처리 과정)

- 1단계 - 질의 이해·재구성: 사용자 질문을 분석해 검색에 적합한 질의로 변환
  (필요시 여러 개의 하위 질의로 분해, 대화 맥락을 반영해 대명사 등을 구체화)
- 2단계 - 실시간 웹 검색: 자체 크롤링 색인과 검색 엔진을 통해 관련 문서 수집
- 3단계 - 순위화·선별: 임베딩 기반 유사도와 신뢰도 기준으로 문서 순위를 매겨 상위 문서만 선별
- 4단계 - 프롬프트 조립: 선별된 문서 조각들을 컨텍스트에 넣고 "이 자료에 근거해 답하라"고 지시
- 5단계 - LLM 합성: 모델이 문서들을 종합해 답변 생성
- 6단계 - 인용 부착: 문장마다 각주 형태로 출처 표시, 클릭 시 원문으로 이동

다양한 LLM 사용

- LLM을 부품처럼 교체 가능한 멀티 모델 구조
- 타사 최전선 모델(GPT, Claude, Gemini 계열)을 사용자가 선택 가능
- 자체 미세조정 모델(Sonar 계열)도 병행 운영 - 검색 합성 작업에 특화, 비용·속도 최적화

특징과 한계

강점

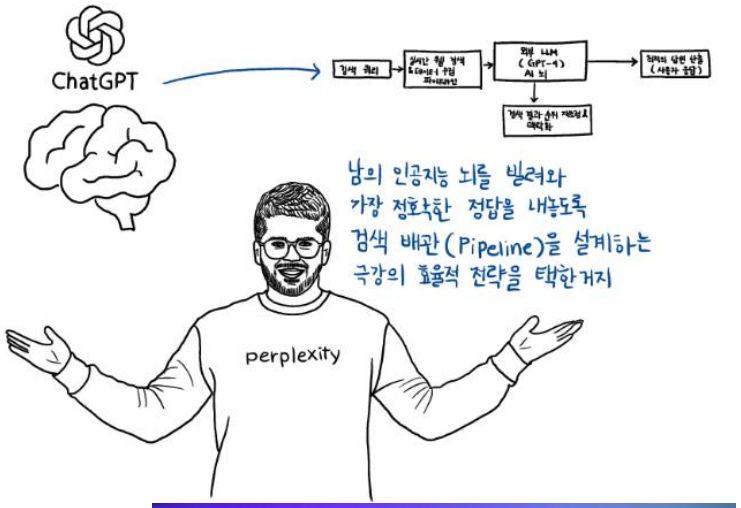
- 최신 정보에 강함 (매 질문마다 실시간 검색)
- 출처 확인 가능 → 검증 용이
- 학습 데이터에만 의존한 환각 상대적으로 억제

한계

검색 결과 품질에 답변 품질이 종속됨

창작, 복잡한 코딩 등 검색이 불필요한 작업에 약함

인용이 있어도 원문을 잘못 요약하는 오류는 여전히 가능





[아라빈드 스리니바스, 94년생]

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

Deep Research(2025.02)

"검색 + 추론 + 반복 탐색 + 종합 보고서 생성" 합쳐진 형태

- 초기: LLM + 웹검색/RAG 형태가 먼저 나왔고, "검색해서 답하기"가 출발점
- 한 번 검색하는 게 아니라, 질문을 쪼개고 여러 번 검색하고 비교·검증하는 에이전트형 리서치로 발전
- 질의 분해 → 반복 검색 → 근거 검증 → 최종 리포트를 자동화
  (인간의 연구과정과 동일하다고 하여 붙여진 이름)
- 구글 Gemini가 먼저 비슷한 기능을 공개했고, 이어서 OpenAI Deep Research가 2025년 2월 대중화



주요 Agent

- Router Agent: 요청 유형 분류
- Clarification Agent: 질문이 모호하면 추가 질문
- Planner/Instruction Agent: 연구 계획과 하위 질문 생성
- Research Agent: 실제 검색·페이지 열기·자료 수집 실시
- Writer/Report Agent: 모은 근거를 정리해서 최종 리포트 작성
- Supervisor Agent: 여러 서브 에이전트를 조정하고 언제 끝낼지 결정

핵심 의미

- RAG는 지식 주입.
- Agent는 도구 사용과 단계적 실행.
- Deep Research는 그 둘을 묶어 리서치 업무 자체를 자동화한 것.

LLM이 "답만 하는 모델" 에서 "조사하고 검증해서 보고서까지 만드는 연구 에이전트"로 진화한 결과

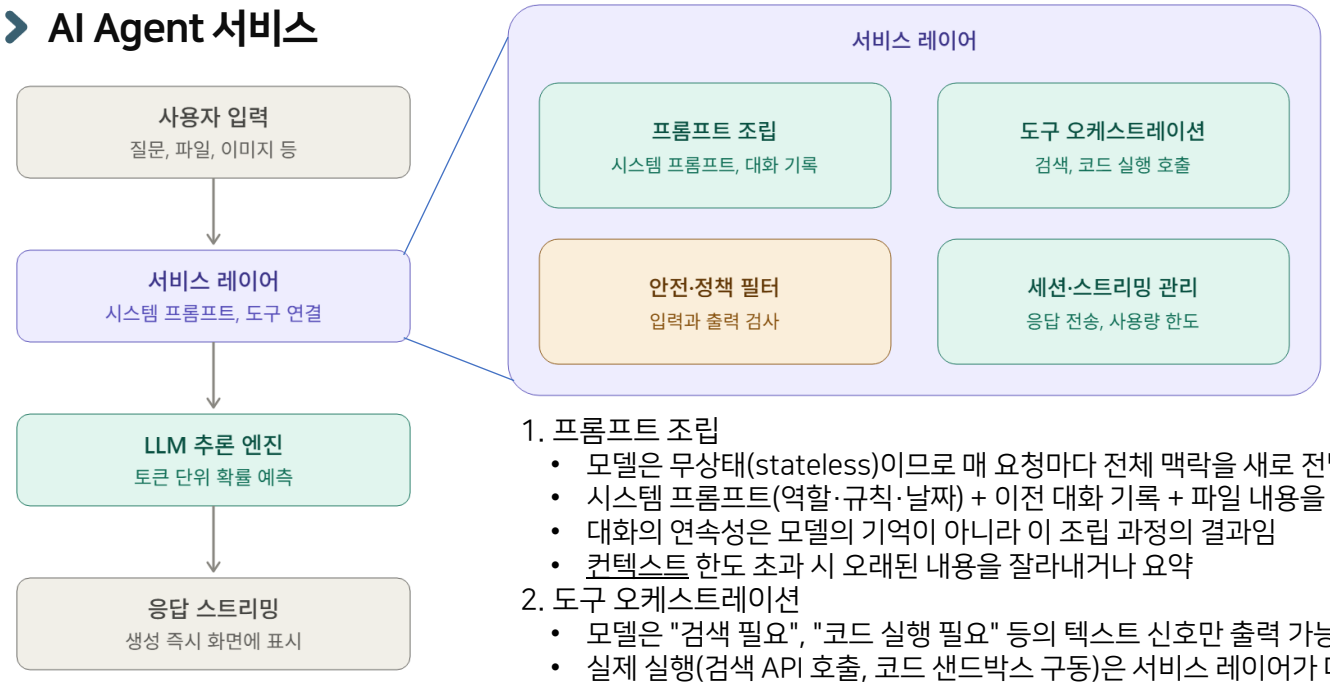
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

AI Agent 서비스의 작동원리

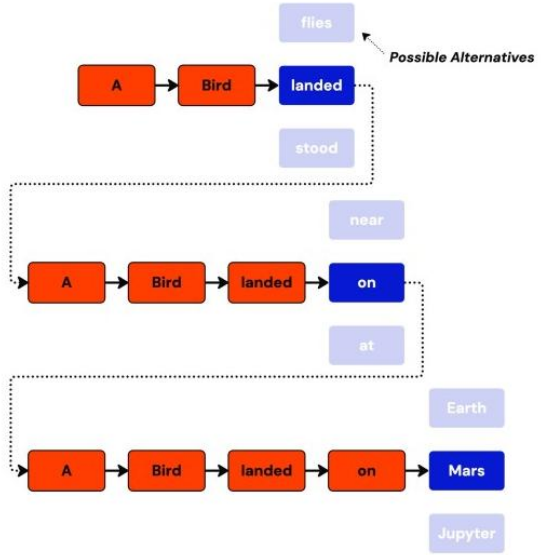
LLM 엔진 : 문장입력 → 토큰화 → 임베딩 → Transformer → 다음 토큰 예측 → 자기회귀 반복 생성

AI Agent 서비스



1. 프롬프트 조립
   - 모델은 무상태(stateless)이므로 매 요청마다 전체 맥락을 새로 전달해야 함
   - 시스템 프롬프트(역할·규칙·날짜) + 이전 대화 기록 + 파일 내용을 하나의 입력으로 결합
   - 대화의 연속성은 모델의 기억이 아니라 이 조립 과정의 결과임
   - 컨텍스트 한도 초과 시 오래된 내용을 잘라내거나 요약
2. 도구 오케스트레이션
   - 모델은 "검색 필요", "코드 실행 필요" 등의 텍스트 신호만 출력 가능
   - 실제 실행(검색 API 호출, 코드 샌드박스 구동)은 서비스 레이어가 대행
   - 실행 결과를 다시 모델 입력에 넣어 답변 생성을 이어감
3. 안전·정책 필터
   - 모델의 자체 판단과 별개로 작동하는 이중 방어선
   - 입력과 출력을 분류기(classifier)로 검사
   - 금지 콘텐츠 차단, 민감 주제 감지 시 추가 지침 주입
4. 세션·스트리밍 관리
   - 생성된 토큰의 실시간 화면 전송(스트리밍)
   - 대화 저장·불러오기, 요금제별 사용량 한도 적용
   - 서버 부하 분산(로드 밸런싱), 장애 시 재시도 등 운영 업무 담당

Autoregressive Model (Left-to-Right Prediction)



Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

Harness Engineering이 필요해

› LLM은 엔진이고, 하네스는 그 엔진을 실제 작업에 연결하는 운전 시스템이다.

1단계 · 완료
LLM 서빙

→

2단계 · 개발이 필요해
하네스 (Harness) - 모델을 둘러싼 제어·연결·감시 계층

→

결과
신뢰할 수 있는 작업 수행

① 컨텍스트 관리

무엇을 알려줄 것인가

시스템 프롬프트, 대화 메모리, 문서 검색(RAG)으로 모델에 필요한 정보를 선별해 넣는다.

② 도구 연결

무엇을 할 수 있는가

함수 호출, MCP, 코드 실행 환경으로 모델이 외부 시스템을 읽고 조작하게 한다.

③ 오케스트레이션

어떤 순서로 일하는가

계획, 반복 루프, 재시도, 멀티 에이전트 분업으로 긴 작업의 흐름을 제어한다.

④ 가드레일

무엇을 막을 것인가

권한 범위, 입출력 검증, 안전 필터, 사람의 승인으로 위험한 행동을 차단한다.

⑤ 평가·관측

잘 되고 있는가

로그, 트레이싱, 평가 데이터셋, 사용자 피드백으로 품질을 측정하고 개선한다.

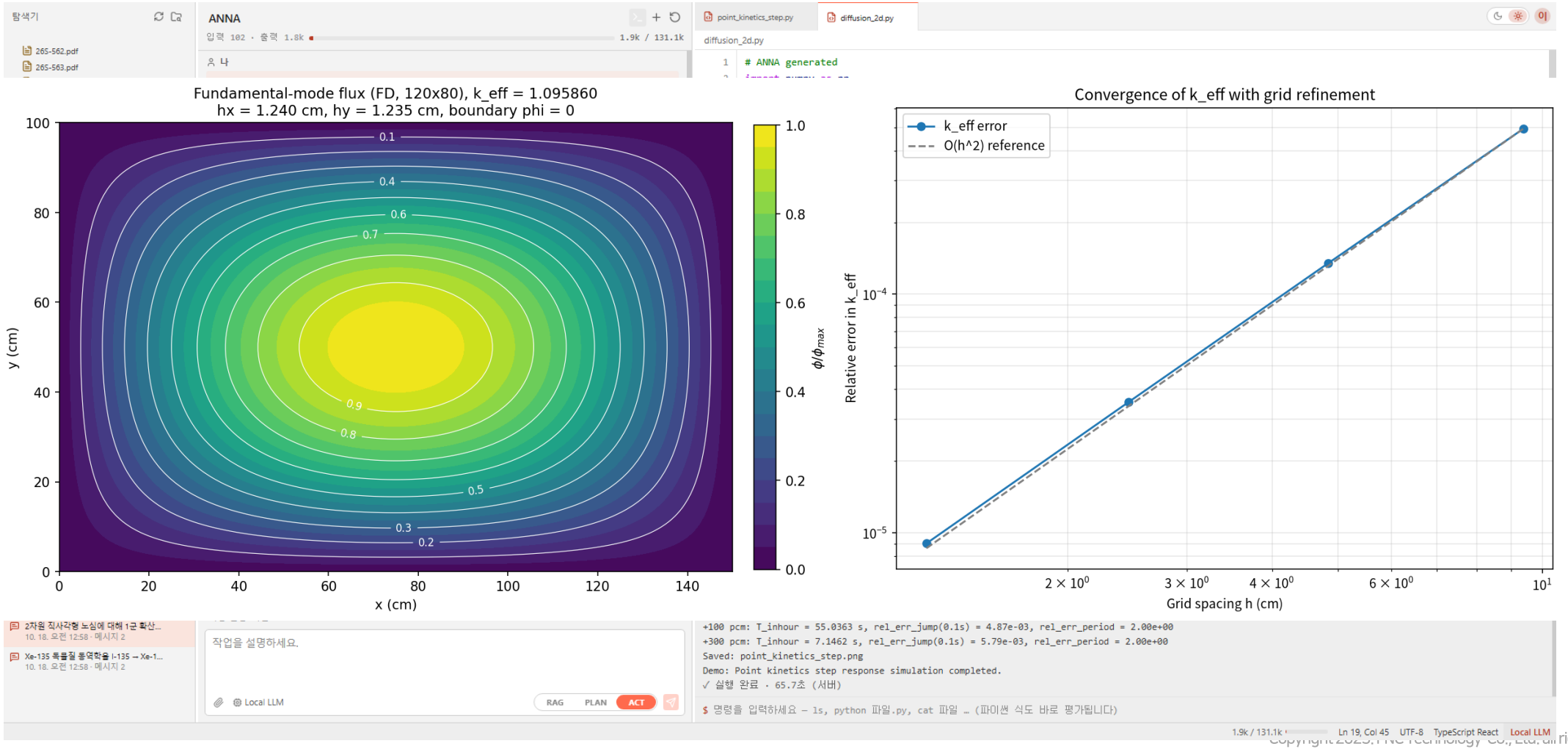
핵심 같은 모델이라도 하네스의 완성도가 결과물의 품질과 신뢰성을 좌우한다.

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

ANNA Code

- 2차원 직사각형 노심에 대해 1군 확산방정식을 유한차분법으로 풀고, 거듭제곱법으로 k_eff와 기본모드 플럭스를 구해라. 계산된 k_eff를 해석적 임계 버클링으로 얻은 값과 비교해 오차를 출력하고, 격자 세분화에 따른 수렴 그래프를 그려라.

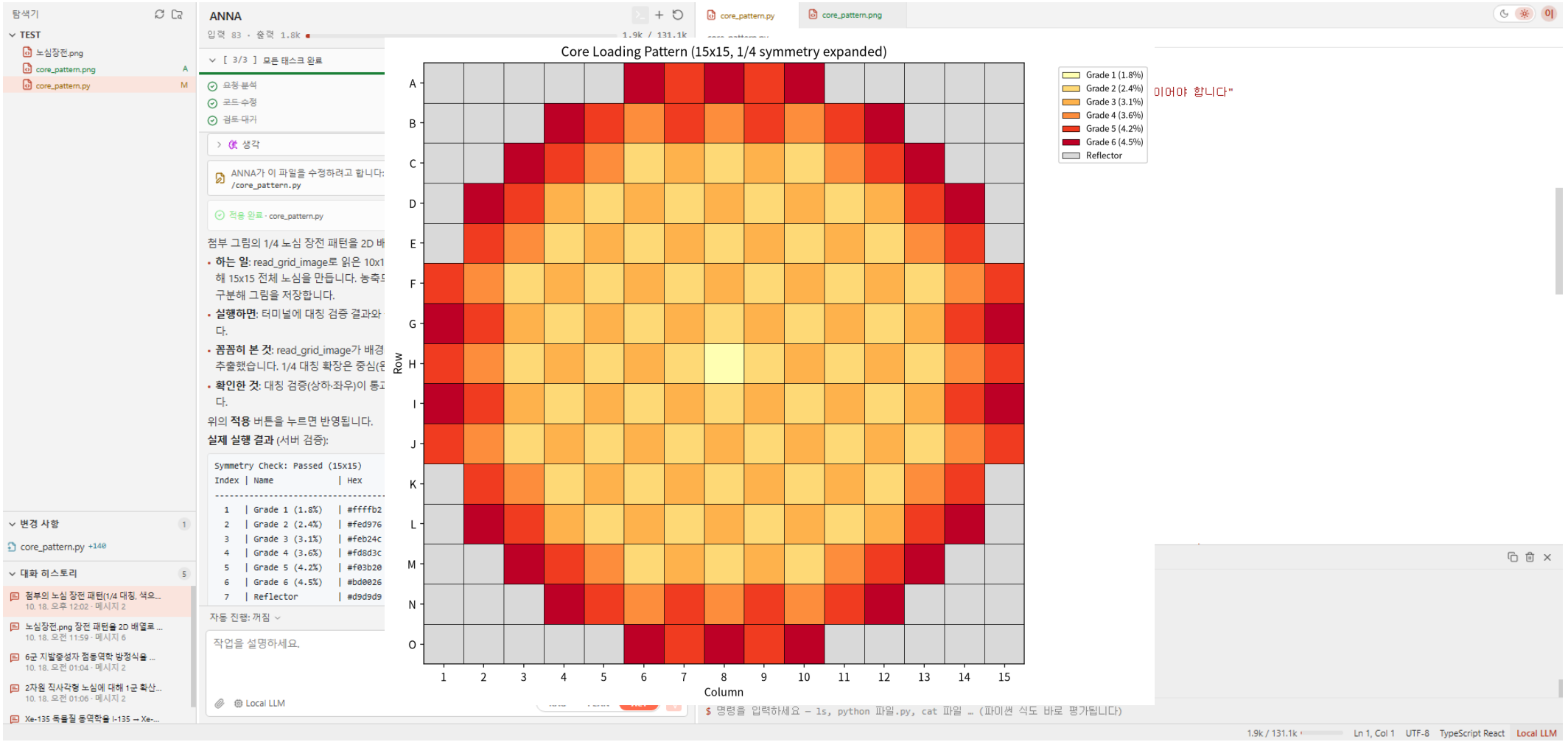


Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

ANNA Code

› 이 장전 패턴을 2D 배열로 읽어내고, 농축도별 분류가 맞는지 표로 출력한 뒤, 전체 노심으로 대칭 확장해 matplotlib으로 다시 그려라.

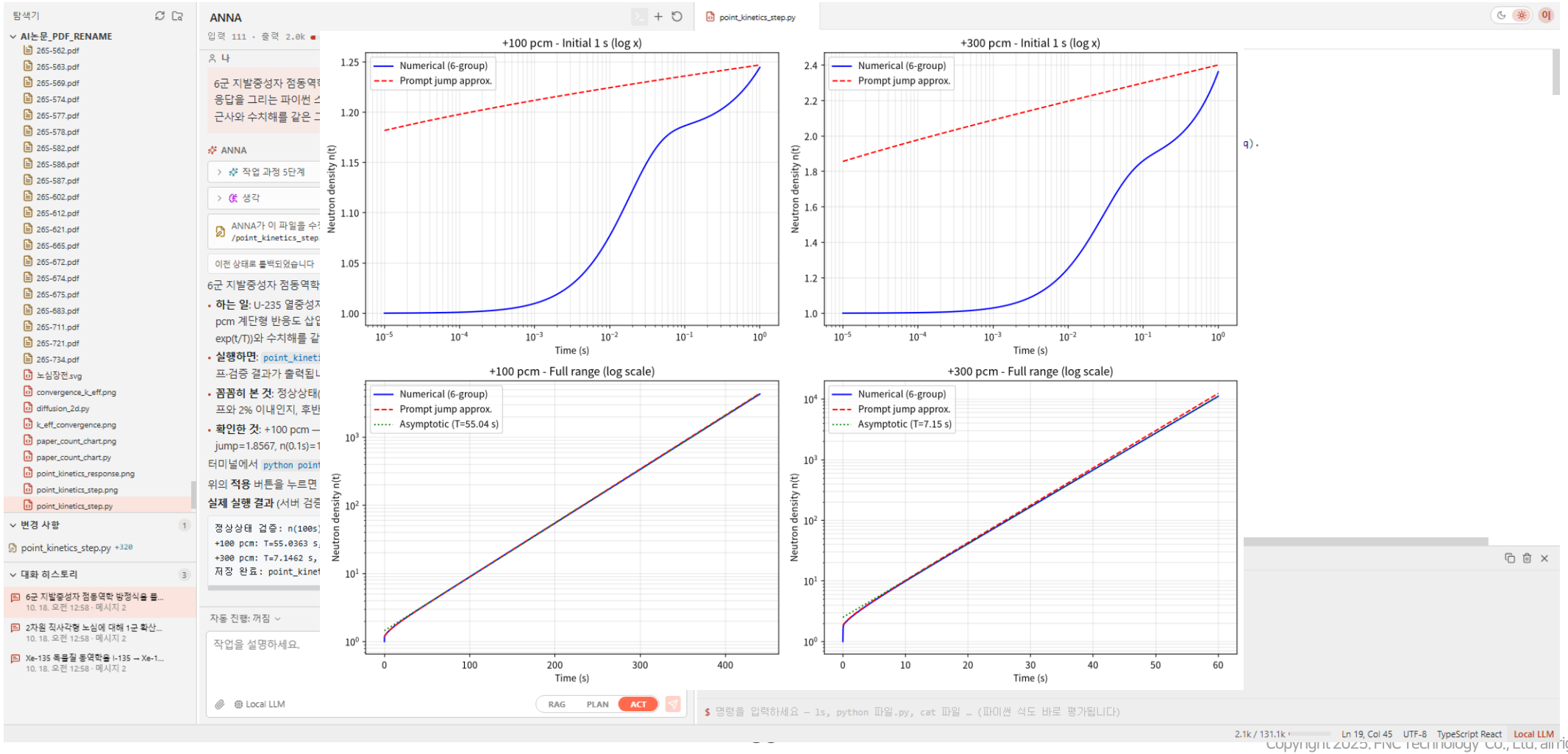


Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

ANNA Code

› **6군 지발중성자 점동역학 방정식을 풀어 계단형 반응도 삽입(+100 pcm, +300 pcm)에 대한 중성자 밀도 응답을 그리는 파이썬 스크립트를 작성해라. 지발중성자 상수는 U-235 열중성자 값을 쓰고, 즉발점프 근사와 수치해를 같은 그래프에 비교해라. 단일 파일, scipy만 사용.**

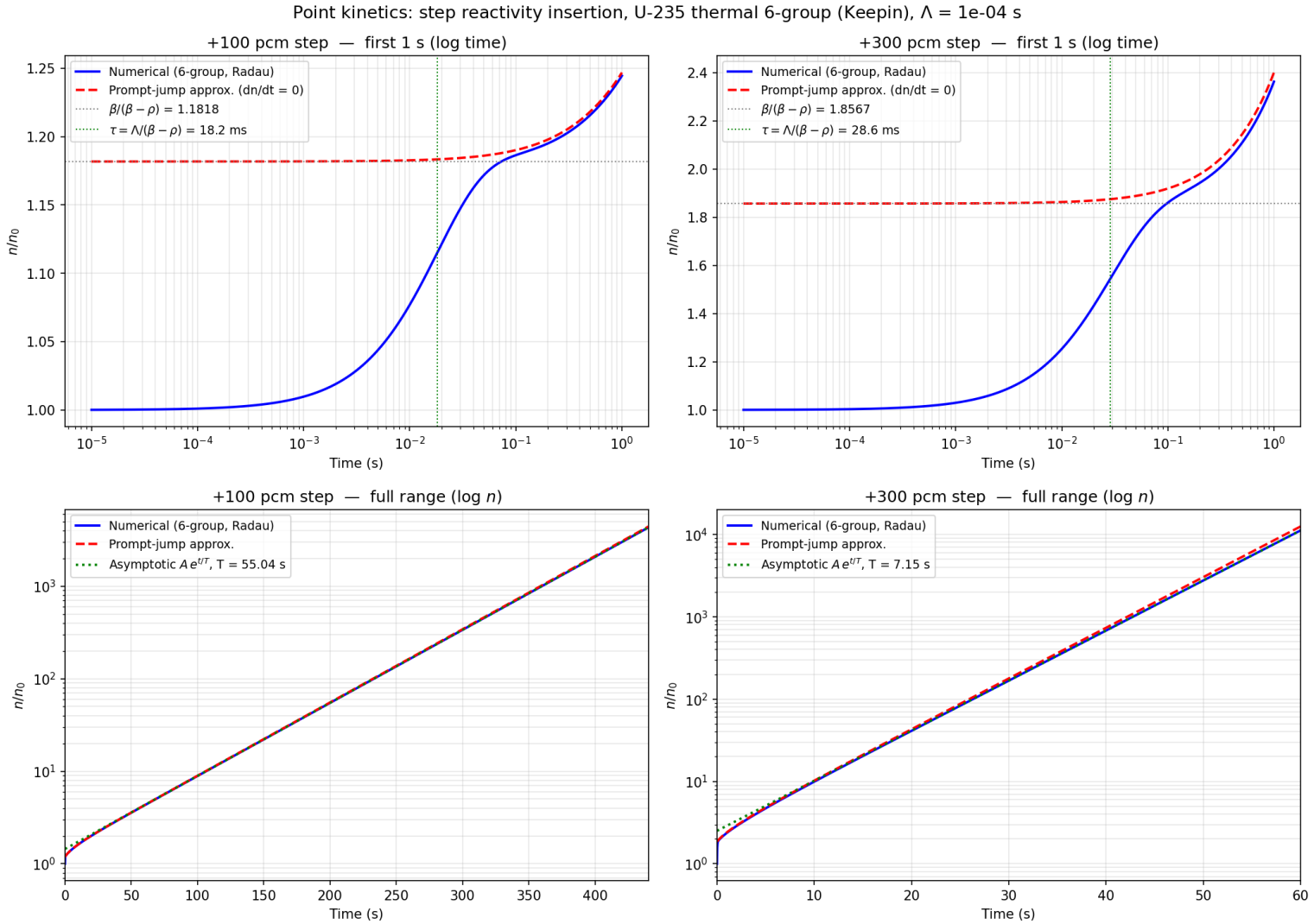


Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

ANNA Code

- 6군 지발중성자 점동
  작성해라. 지발중성자
- Claude 의견
  - 핵심 물리와 수치해
    이어붙인 형태여서 "
  - 물리 모델과 수치 결
    이중 스크립트 구조
    그대로 t_eval로 쓰



= 파이썬 스크립트를
, scipy만 사용.

서로 다른 두 스크립트를

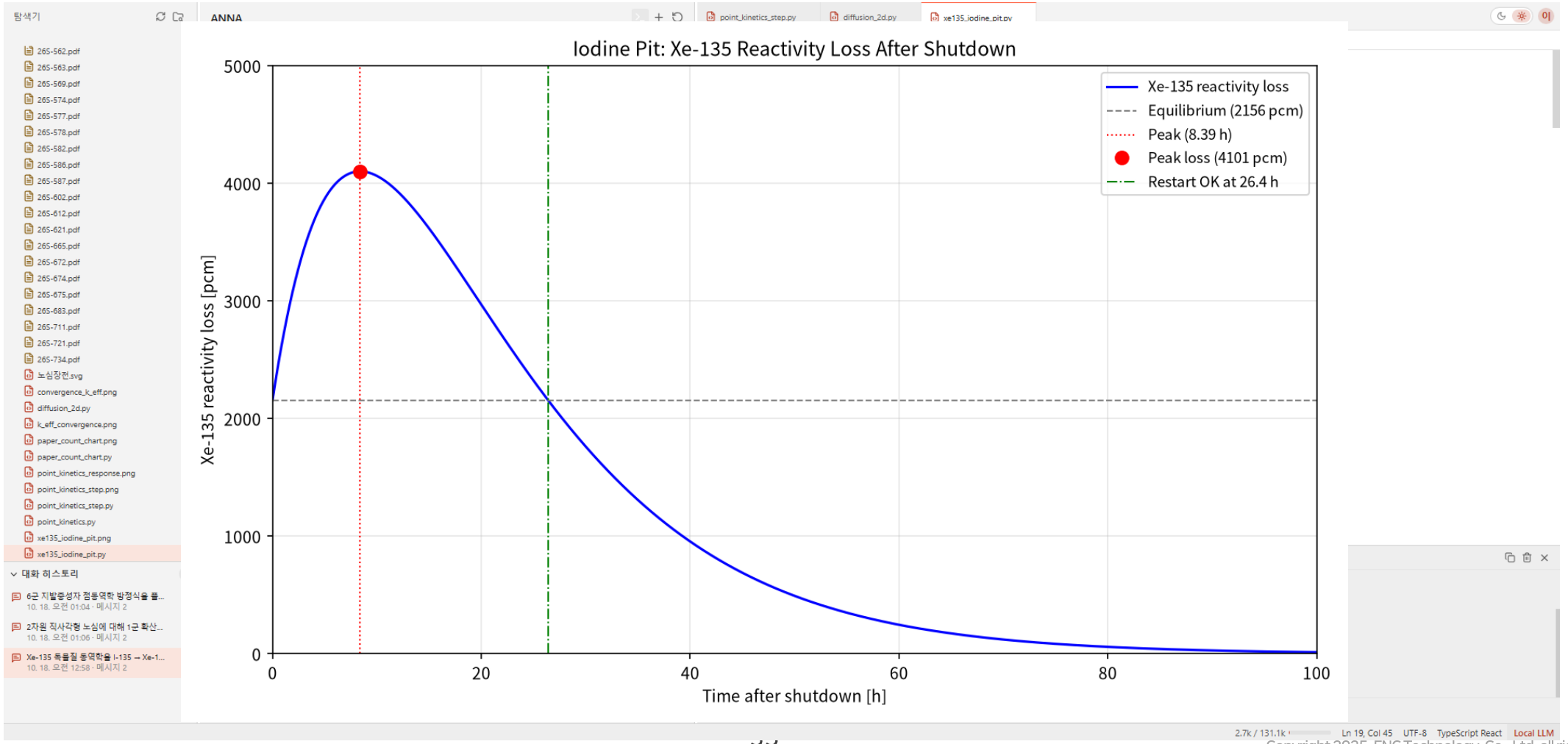
력의 부호 오류, 그리고
on이 입력 t 배열을

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

4 Building AI Agents

ANNA Code

- Xe-135 독물질 동역학을 I-135 → Xe-135 사슬로 모델링해, 정격 출력 운전 후 정지 시 요오드 피트가 발생하는 과정과 반응도 손실(pcm)을 시간에 따라 그려라. 출력 감소 후 재기동 가능 시점을 자동 판정하는 함수도 포함해라.



Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

ANNA Code

Xe-135 독물질 동역학을 I-135 → Xe-135 사슬로 모델링해, 정격 출력 운전 후 정지 시 요오드 피트가 발생하는 과정과 반응도 손실(pcm)을 시간에 따라 그려라. 출력 감소 후 재기동 가능 시점을 자동 판정하는 함수도 포함해라.

Claude 의견

물리 모델, 상수, 해석해, 수치해 모두 올바르며 그래프의 수치도 독립적으로 재계산한 값과 일치한다. 요구사항 중 "요오드 피트 과정과 반응도 손실 그리기"는 충실히 수행했고, "재기동 가능 시점 자동 판정 함수"는 동작은 하지만 일반성이 부족하다. 전체적으로 적절한 결과물이나, 재기동 판정 함수의 설계와 그래프의 정보량에서 보완 여지가 있다.

수식·상수·수치가 모두 검증되는 올바른 구현이며, 그래프가 보여주는 평형 2156 pcm, 피크 4101 pcm(8.39 h), 재교차 26.4 h는 독립 계산과 일치한다. 다만 재기동 판정 함수는 잉여반응도와 정지 후 플럭스를 인자로 받는 일반형이 아니라 특정 기준(L_eq, φ = 0)에 고정된 축소판이고, I-135·Xe-135 농도의 시간 변화를 보여주지 않아 "과정"의 묘사가 부족하다. 프롬프트의 요구를 대체로 충족하는 적절한 결과물이되, 앞선 확산방정식 결과물보다는 요구사항 해석의 폭이 좁다.



Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

5 Future Work

AI의 빈익빈 부익부

AI는 이미 자본·데이터·역량을 갖춘 쪽에 더 큰 이익을 안겨주고, 그렇지 못한 쪽과의 격차를 더욱 심화시키는 구조로 발전 중

영역	부익부	빈익빈
자본·컴퓨팅	최신 모델 학습에 필요한 자본(CAPEX)과 GPU를 감당할 수 있는 소수 기업(미국, 중국)이 시장을 선점한다.	후발 주자와 스타트업은 진입 비용 자체를 감당하지 못해 종속적 위치에 머문다.
데이터·사용자	사용자가 많을수록 데이터가 쌓이고, 모델이 좋아져 다시 사용자가 몰린다.	데이터가 적은 서비스는 품질 격차가 벌어져 사용자를 더 잃는다.
개인·노동	AI 활용 역량이 뛰어나고 고성능 모델·무제한 토큰 구매력을 갖춘 개인은 다중 에이전트를 구동하며 생산성과 소득을 극대화한다.	교육 기회가 부족하거나 비용 문제로 저가·무료 모델의 호출 제한에 묶인 개인은 산출물 품질에서 밀려나며 일자리와 경쟁력을 잃는다.
국가	컴퓨팅파워·인재·데이터를 보유한 국가가 기술 표준과 규칙을 정한다.	그 밖의 국가는 해외 플랫폼에 의존하며 기술 종속이 깊어진다.

자본 → 더 나은 모델 → 더 많은 사용자 → 더 많은 데이터 → 더 큰 자본

Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

5 Future Work

AI 발전은 현재 진행형(ING)

AI의 대중화를 위해 다양한 HW 개발 중

- 서버 중심의 GPU 시장에서 DGX Spark, RTX Spark같은 개인용 AI 워크스테이션 보급
  - 20코어 Arm CPU, Blackwell 아키텍처 6,144개 CUDA 코어, 128GB LPDDR5x(273 GB/s)
- GPU 가격은 언제까지나 고가로 유지되기는 어렵다.
  - 26년 6말~7초 미국 캘리포니아 연방법원에 소비자들과 소규모 PC 관련 사업자들을 중심으로 집단소송(Class-Action Lawsuit)이 제기, 메모리 3사가 인공지능(AI)용 고대역폭 메모리(HBM) 전환을 핑계 삼아 일반 소비용 D램(DDR3, DDR4, DDR5 등)의 생산량을 의도적으로 제한하고 공급 부족을 유발하여 가격을 비정상적으로 폭등시켰다고 주장



Frontier 모델에 필적하는 오픈웨이트 LLM 출시

- Qwen 3.8 Flash-Next 모델의 경우 180B 임에도 고도의 성능을 보여줌
- 오픈웨이트 시장에서는 작고, 똑똑한 모델을 만드는 것이 대세 : 다운로드 증가

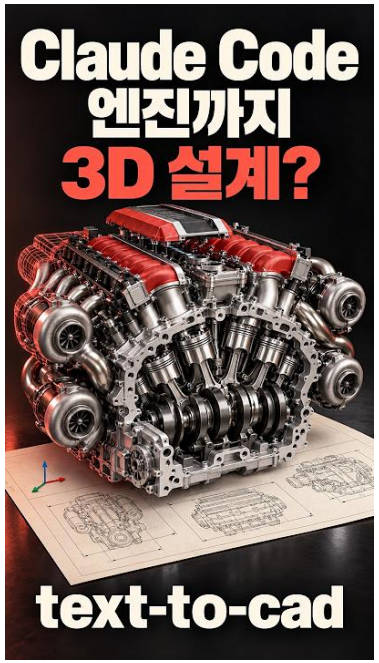
SW 개발분야 에서도 LLM을 최대한 활용하여 그 동안 불가능했던 기능 개발 중

- Text to CAD
- Diagram Design
- Reverse Engineering Anything









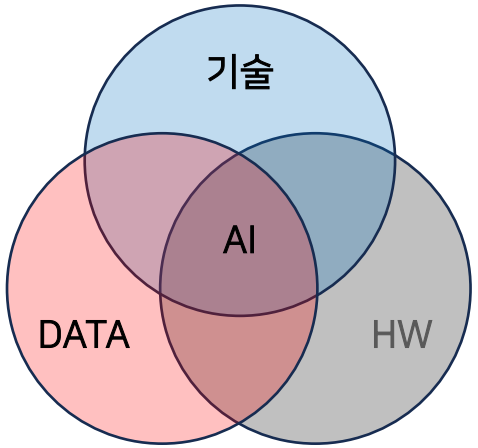
Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

5 Future Work

Our Mission

우리가 준비할 수 있는 건: 데이터와 기술 영역

- HW: 우리의 영역이 아니다. GPU 한두 장 도입이 최선이다.
- 모델: GPU 한두 장으로 돌아가는 똑똑한 범용 LLM은 반드시 나온다. 작고 똑똑한 오픈웨이트 모델이 이미 대세이니, 기다려보자.(관찰)
- 데이터·기술: 범용 모델이 아무리 좋아져도 암묵지의 전문영역까지 커버할 순 없다.
  이것은 우리만 할 수 있고, 우리가 해야 한다.

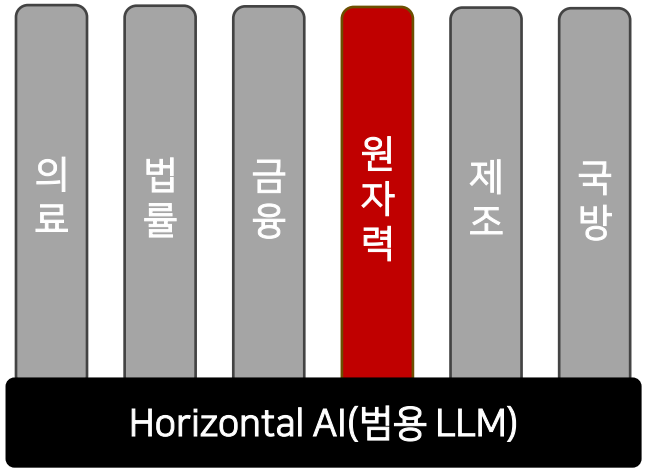


무엇을 준비할 것인가?

- 데이터: 데이터 플랫폼/온톨로지
  - 핵설계보고서(NDR)·노심 운영 보고서·장전패턴 이력의 DB화, 격자·노심 코드(예: CASMO/SIMULATE, DeCART/MASTER 등)의 입출력 데이터셋, 노내 계측·노심 추적(출력분포, 붕소농도, 임계 데이터)과 계산값의 비교 데이터셋, 노심설계 Q&A·평가(벤치마크) 데이터셋
- 기술: 하네스 엔지니어링/미세조정
  - 모델을 노심설계 업무에 연결하는 하네스 — 설계보고서·절차서 기반 RAG·에이전트(ANNA), 노심 해석 코드를 도구로 연결하는 MCP(입력 생성→실행→결과 추출 자동화), 격리 실행·검증·자가수정으로 설계 계산의 신뢰성을 확보하는 Coding Agent, 노심 설계 용어·입력 포맷에 맞춘 도메인 미세조정(SFT)

Horizontal AI(범용 LLM) 위 제조·국방·의료·법률·금융·원자력 Vertical AI

- 각 산업은 범용 LLM 위에 자기 도메인의 Vertical AI를 쌓고 있다.
- 원자력은 전문 지식의 밀도가 높고 폐쇄망 환경이라 로컬 Vertical AI의 필요성이 가장 큰 분야이다.
- Local LLM 기반 AI Agent를 원자력 설계·해석·규제·운영을 지원하는 원자력 전문 AI 연구자로 키우자.



Copyright 2025. FNC Technology Co., Ltd. all rights reserved.

ABOUT

(주)미래와도전
FNC Technology Co., Ltd.

본사

16954 경기도 용인시 기흥구 흥덕1로 13, 32층 (영덕동, 흥덕아이티밸리 타워동)
+82-31-8065-5114

대전지사

대전광역시 유성구 대덕대로 593, 10층 1004-1호 (도룡동, 대덕테크비즈센터)
+82-42-867-5114

미래에너지기술연구소

경기도 용인시 기흥구 탑실로 46, 미래에너지기술연구소
+82-31-8005-5618

설비개발연구센터

경기도 용인시 기흥구 탑실로 46, 미래에너지기술연구소
+82-31-8065-5114

UAE 아부다비지사

Office 904, Capital Tower, Al Rawhah, PO Box 95133, Abu Dhabi, UAE
+971-2-643-5144

미국 메릴랜드지사

11140 Rockville Pike, Suite 380A, Rockville, MD 20852 USA
+1-202-577-5537



THANK YOU

FNC TECHNOLOGY CO., LTD.

