

HyperCLOVA: 초대규모 AI의 현재와 미래

하정우 PhD

네이버 AI Lab 연구소장, AI Tech 위원회 좌장

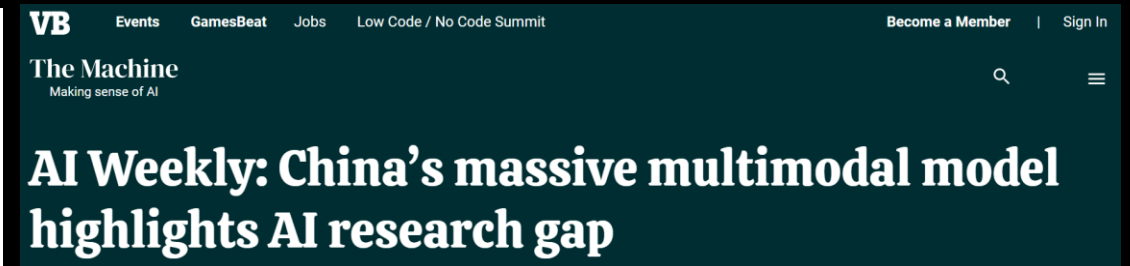
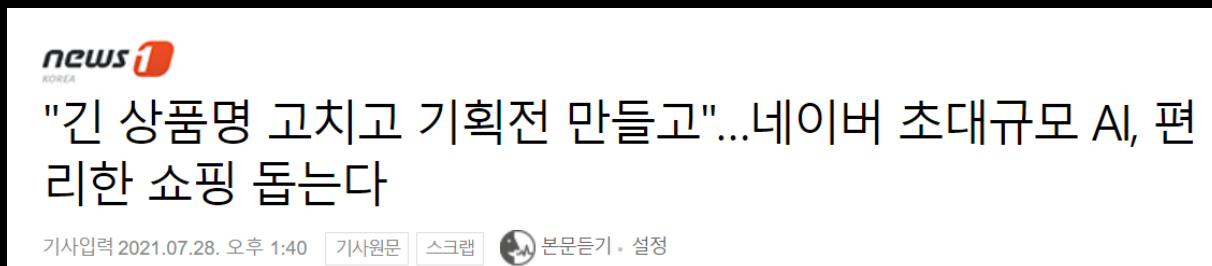
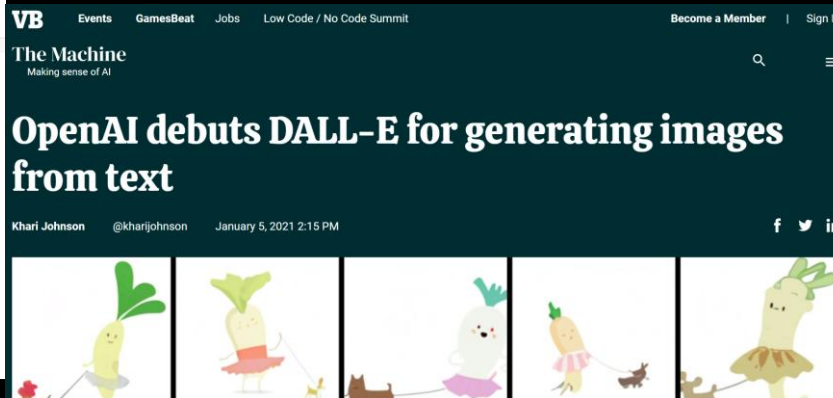
서울대-네이버 초대규모 AI연구센터 공동센터장, KAIST-네이버 초창의적 AI연구센터 공동센터장

AI미래포럼 공동의장

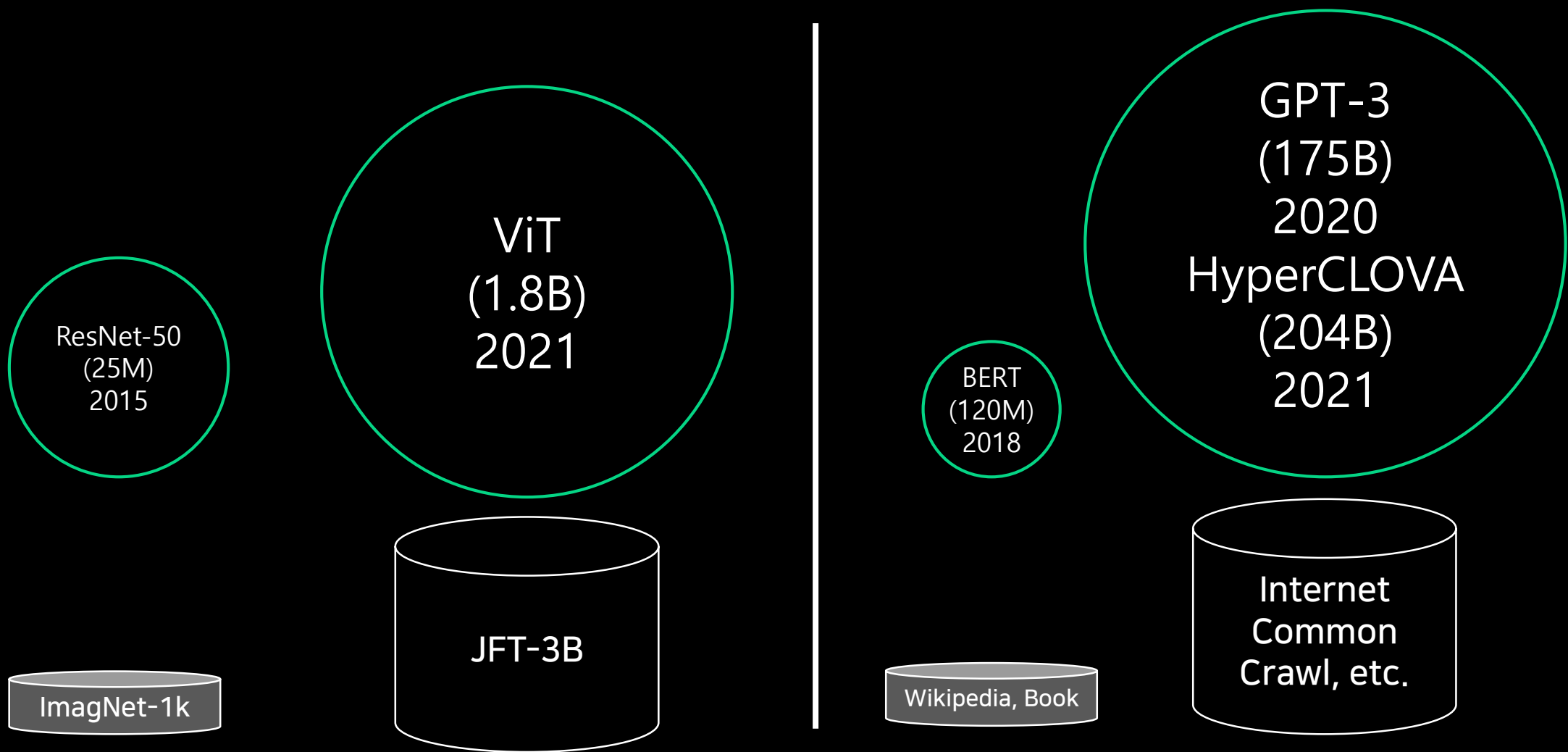


초대규모 AI 시대

초대규모 AI의 시대, 전세계가 초대규모 AI 경쟁 진행중



모델과 데이터의 초대규모화



모델과 데이터의 초대규모화

MS-Nvidia가 함께 개발한 Turing-NLG 530B (GPT3의 3배)

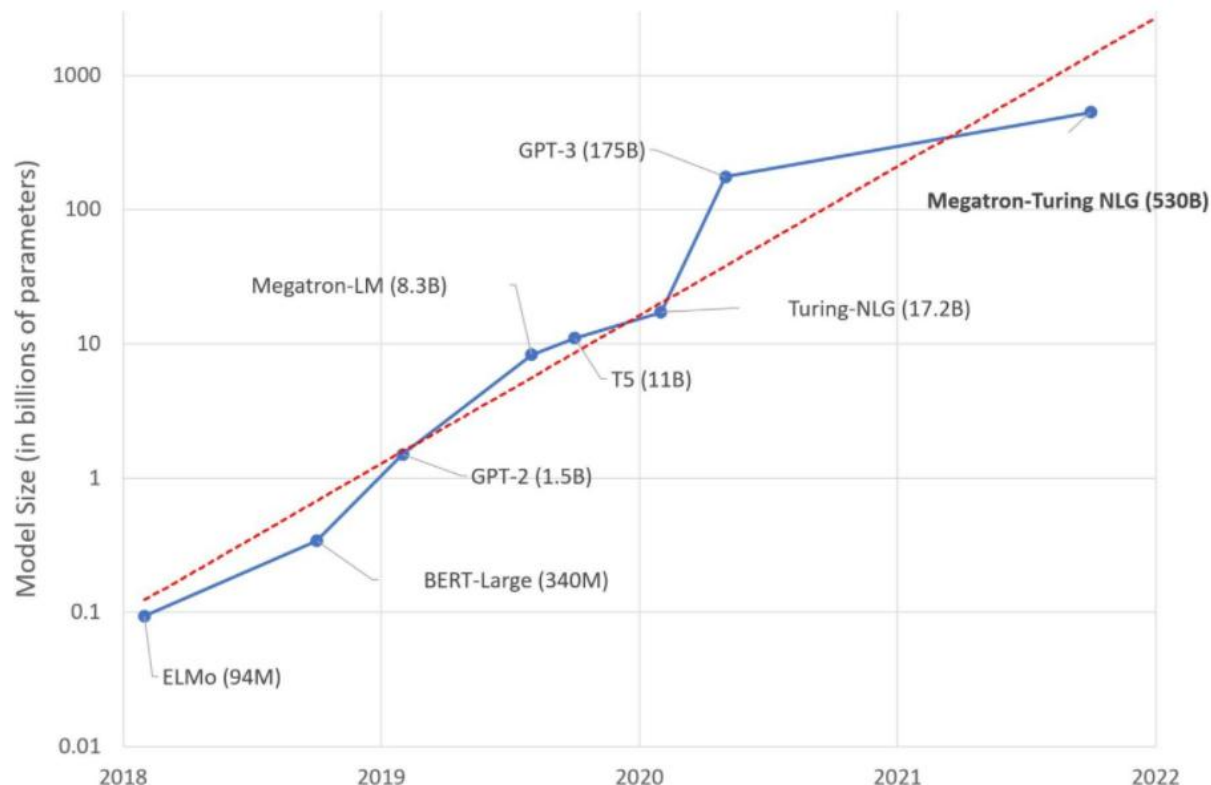
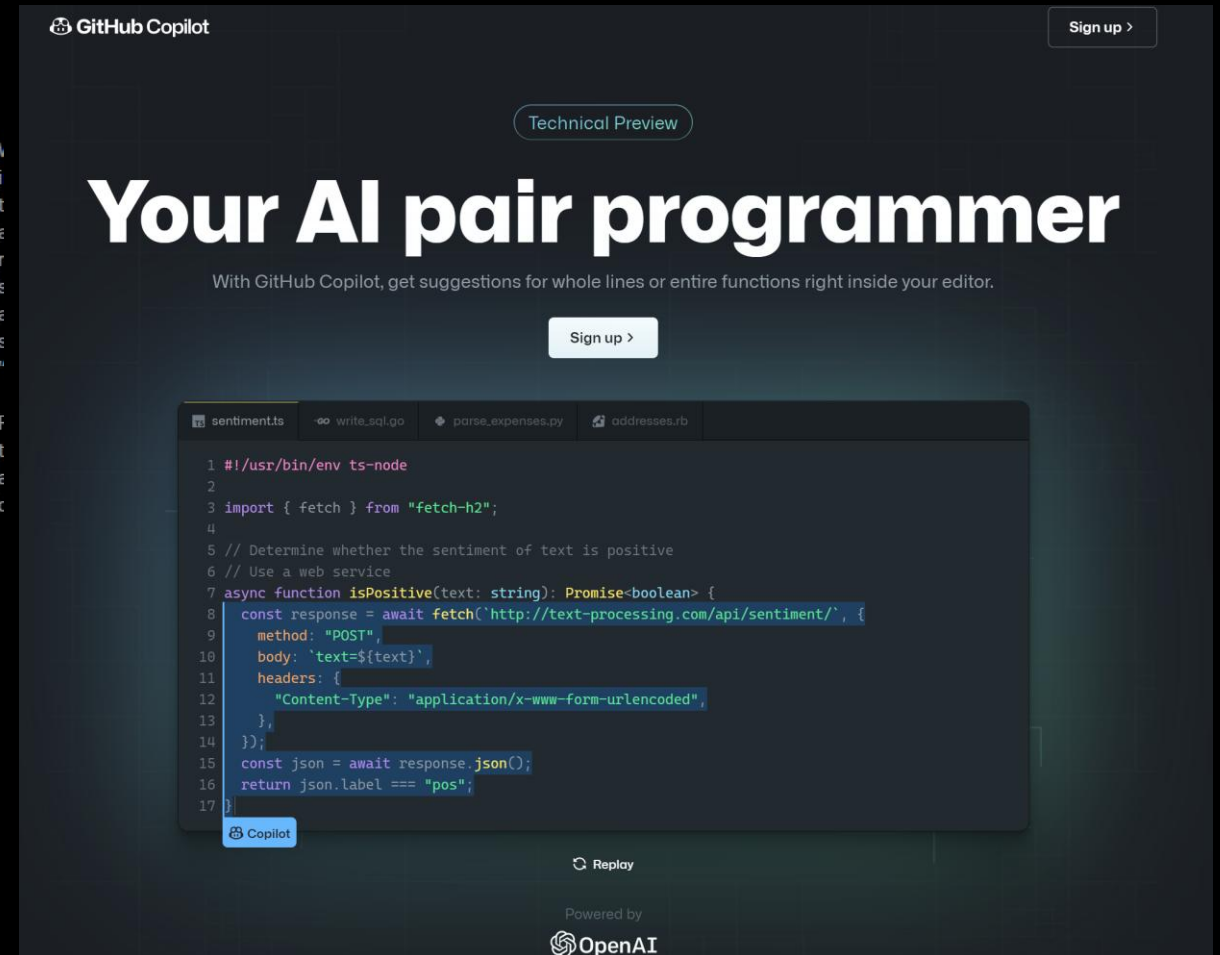
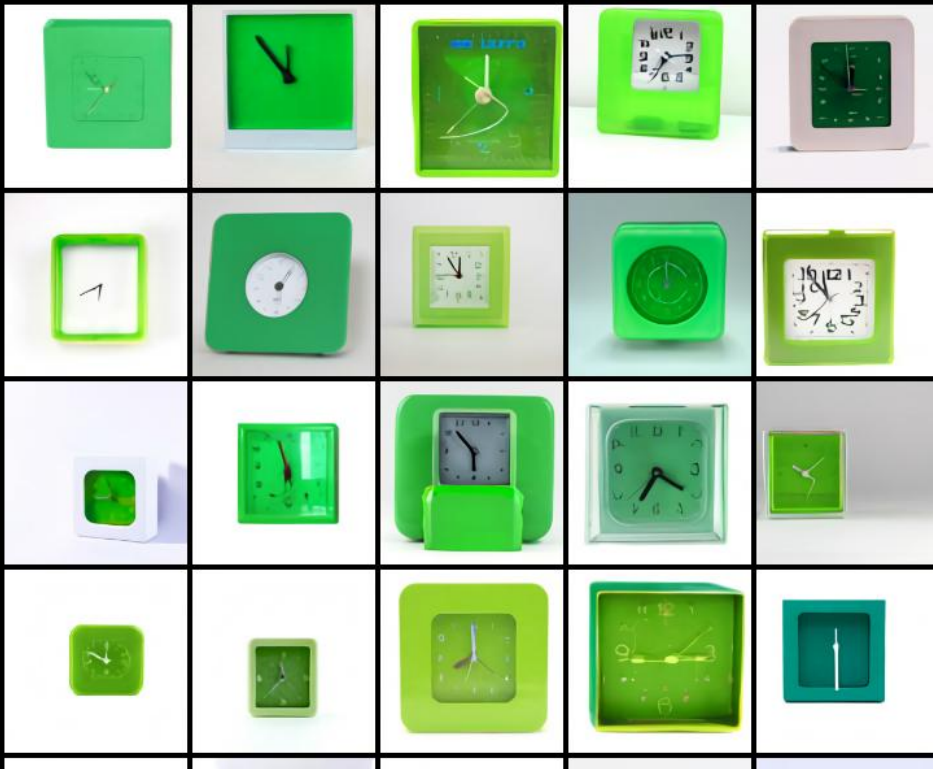


Figure 1. Trend of sizes of state-of-the-art NLP models over time

DALL-E & CODEX (Copilot)

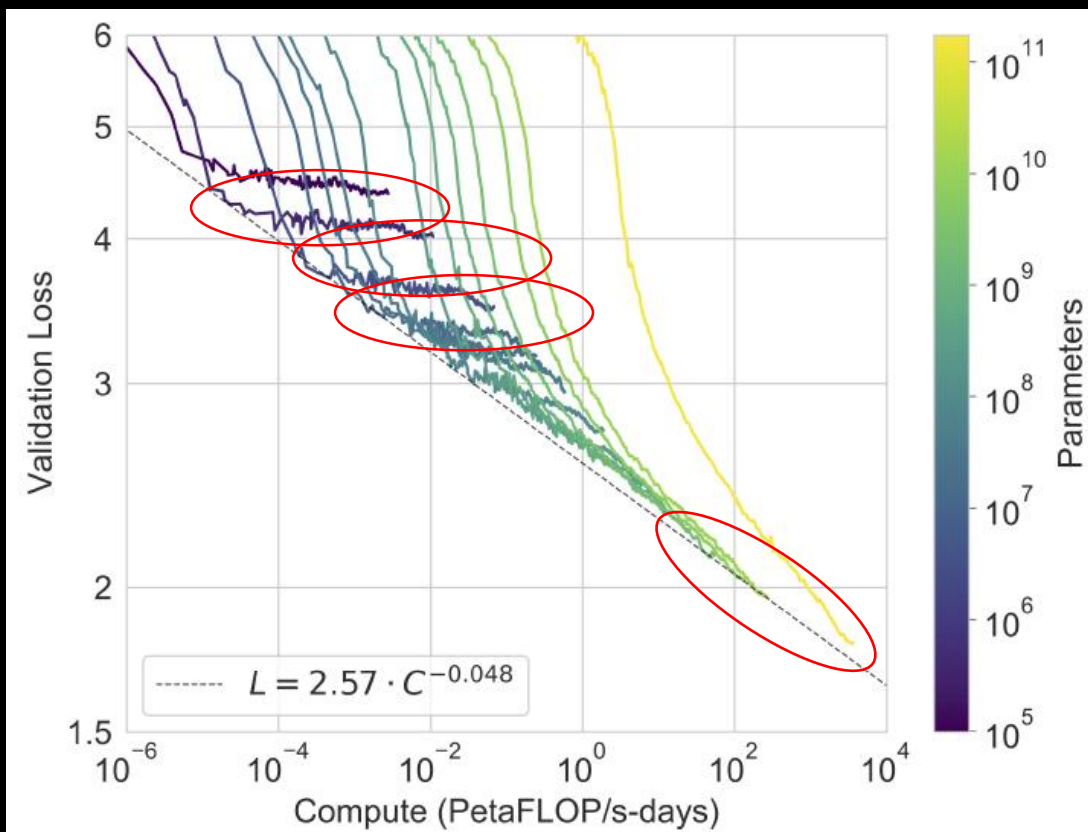
TEXT PROMPT a square green clock. a green clock in the shape of a square.

AI-GENERATED
IMAGES

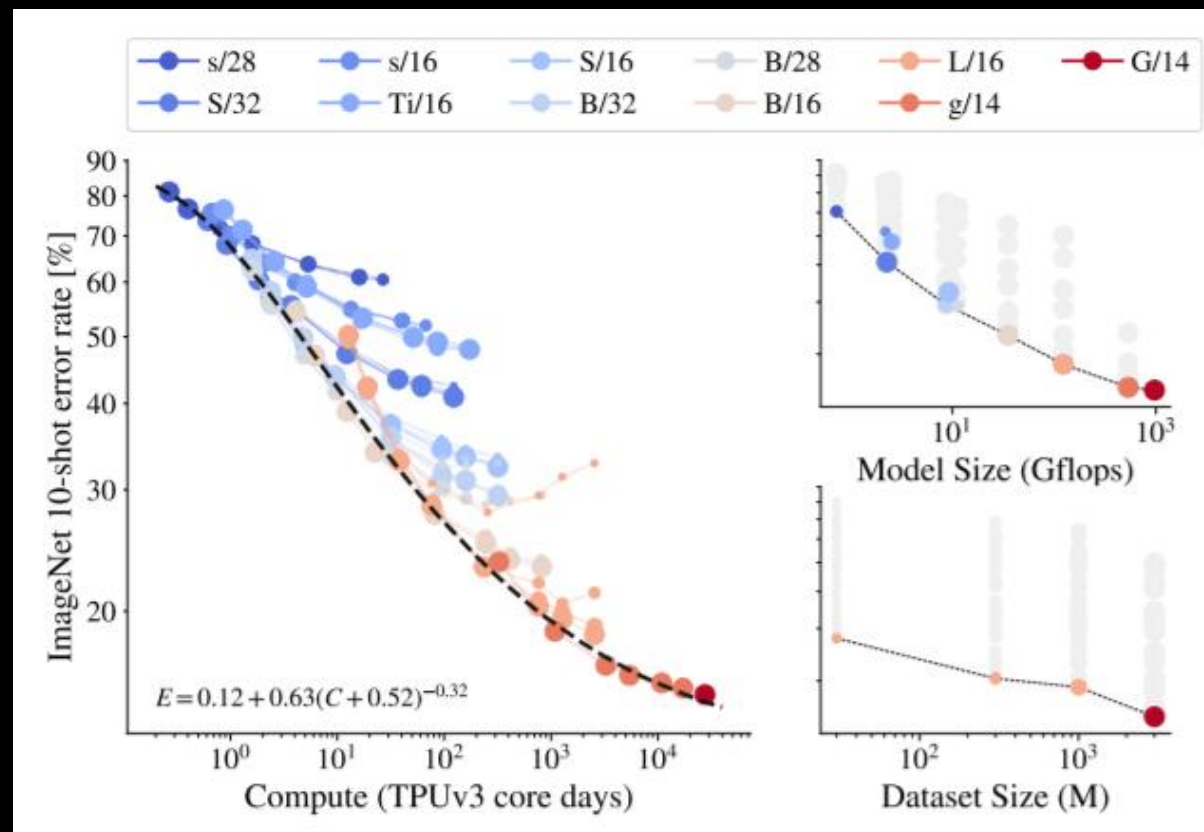


Hyperscale의 중요성 (Scaling Law)

모델과 데이터 크기 증가 → 불가능하던 것이 된다!



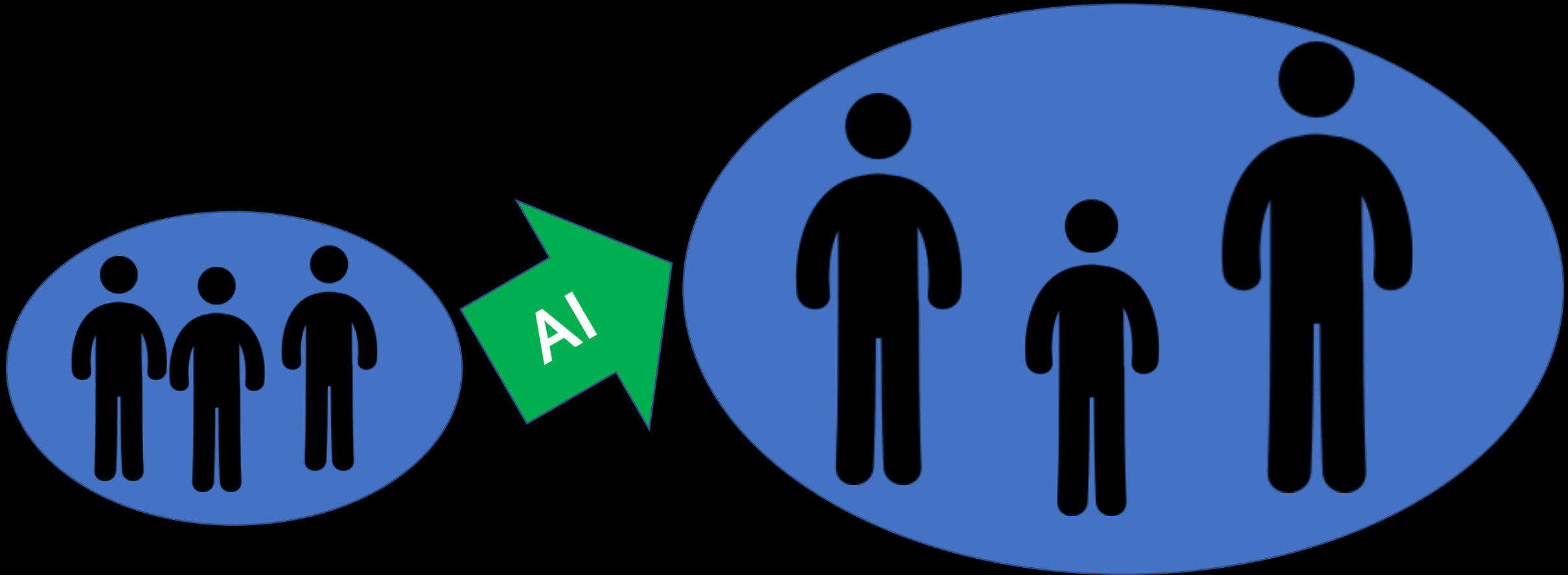
[Brwon et al. NeurIPS 2020]



[Zhai et al. arXiv 2021]

CLOUD + AI: 플랫폼 성장의 핵심 기술

사용자, SME(중소상공인), 크리에이터등의 성장 지원 → 지속가능한 동반성장



NAVER CLOVA

CLOVA EVERYTHING, MEETS AI

네이버와 라인의 AI 기술을 연구하고 서비스하는 조직이며
한국과 일본에서 가장 큰 규모의 AI 조직으로
글로벌 시장을 선도하기 위한 다양한 도전을 계속하고 있습니다.



CLOVA Dubbing



CLOVA Note



CLOVA Chatbot

CLOVA
Contact Center

CLOVA
OCR



CLOVA DEVICES

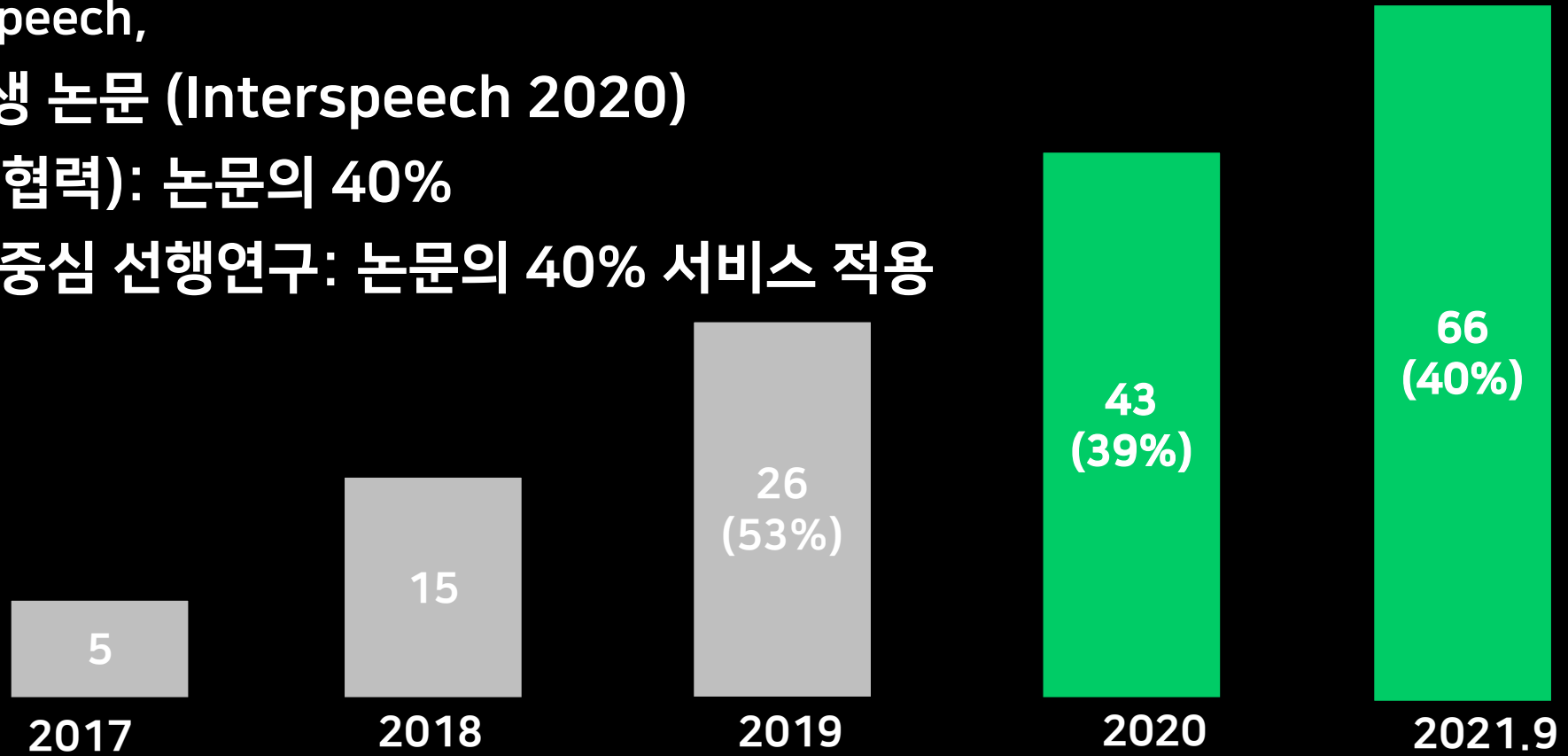
CLOVA 케어콜: COVID-19를 위한 AiCall

- 국내 최초 COVID-19 능동감시자를 위한 전화기반 건강상태 확인 AI



학술적 AI연구성과

- 세계최고 AI 학회 논문 발표
 - NeurIPS, ICML, ICLR, CVPR, ICCV, ECCV, ACL, EMNLP, AAAI, ICASSP, Interspeech,
- 최우수학생 논문 (Interspeech 2020)
- 인턴(산학협력): 논문의 40%
- 실제문제 중심 선행연구: 논문의 40% 서비스 적용



AI 서비스 개발 패러다임의 변화

모든 유형의 데이터

HyperScale AI

(Foundation model [Bommasani et al. arXiv 2021])

↓
디지털파일

↓
Q&A

↓
쇼핑추천

↓
고객응대

↓
수요예측

↓
회의록

초대규모 AI의 필수요소

**Researchers
Engineers**

Data

**Infrastructure
(GPUs, Networks,
Storage, Space, ...)**

HyperCLOVA

- (아마도) 최초로 Top AI학회에 발표되는 수십B급 초대규모 비영어 AI
- 한국어 82B GPT3: 13B와 175B 사이 모델의 성능 분석 (39B, 82B)
- 한국어 특성을 고려한 모델 입력 및 전처리 기법 (Tokenization) 설계
- 최신 Prompt 기법들의 GPT3에서 효과
- HyperCLOVA Studio: No Code AI의 실현가능성

[Kim et al. EMNLP 2021]

What Changes Can Large-scale Language Models Bring?
Intensive Study on HyperCLOVA: Billions-scale Korean Generative
Pretrained Transformers

Boseop Kim^{*,1} HyungSeok Kim^{*,1} Sang-Woo Lee^{*,1,2} Gichang Lee¹ Donghyun Kwak¹
Dong Hyeon Jeon³ Sunghyun Park⁴ Sungju Kim^{1,3} Seonhoon Kim³ Dongpil Seo¹
Heungsub Lee¹ Minyoung Jeong¹ Sungjae Lee¹ Minsub Kim¹ Suk Hyun Ko¹
Seokhun Kim¹ Taeyong Park¹ Jinuk Kim¹ Soyoung Kang¹ Na-Hyeon Ryu¹
Kang Min Yoo^{1,2} Minsuk Chang² Soobin Suh^{1,3} Sookyo In^{1,3} Jinseong Park^{1,3}
Kyungduk Kim^{1,3} Hiun Kim¹ Jisu Jeong^{1,2} Yong Goo Yeo¹ Donghoon Ham¹
Dongju Park¹ Min Young Lee¹ Jaewook Kang¹ Inho Kang^{1,3} Jung-Woo Ha^{1,2}
Woomyoung Park¹ Nako Sung¹

NAVER CLOVA¹ NAVER AI Lab² NAVER Search³ Search Solutions, Inc.⁴

AI 학습을 위한 초대규모 인프라스트럭처

외부 인프라 활용이 아닌 직접 초대규모 GPU 인프라 구축 설치

NAVER

ANNOUNCING
NAVER ADOPTS DGX
SUPERPOD TO CREATE
LANGUAGE UNDERSTANDING
AI SERVICES

#1 Internet Tech Platform in Korea

Construct AI R&D Belt Across Asia and Europe

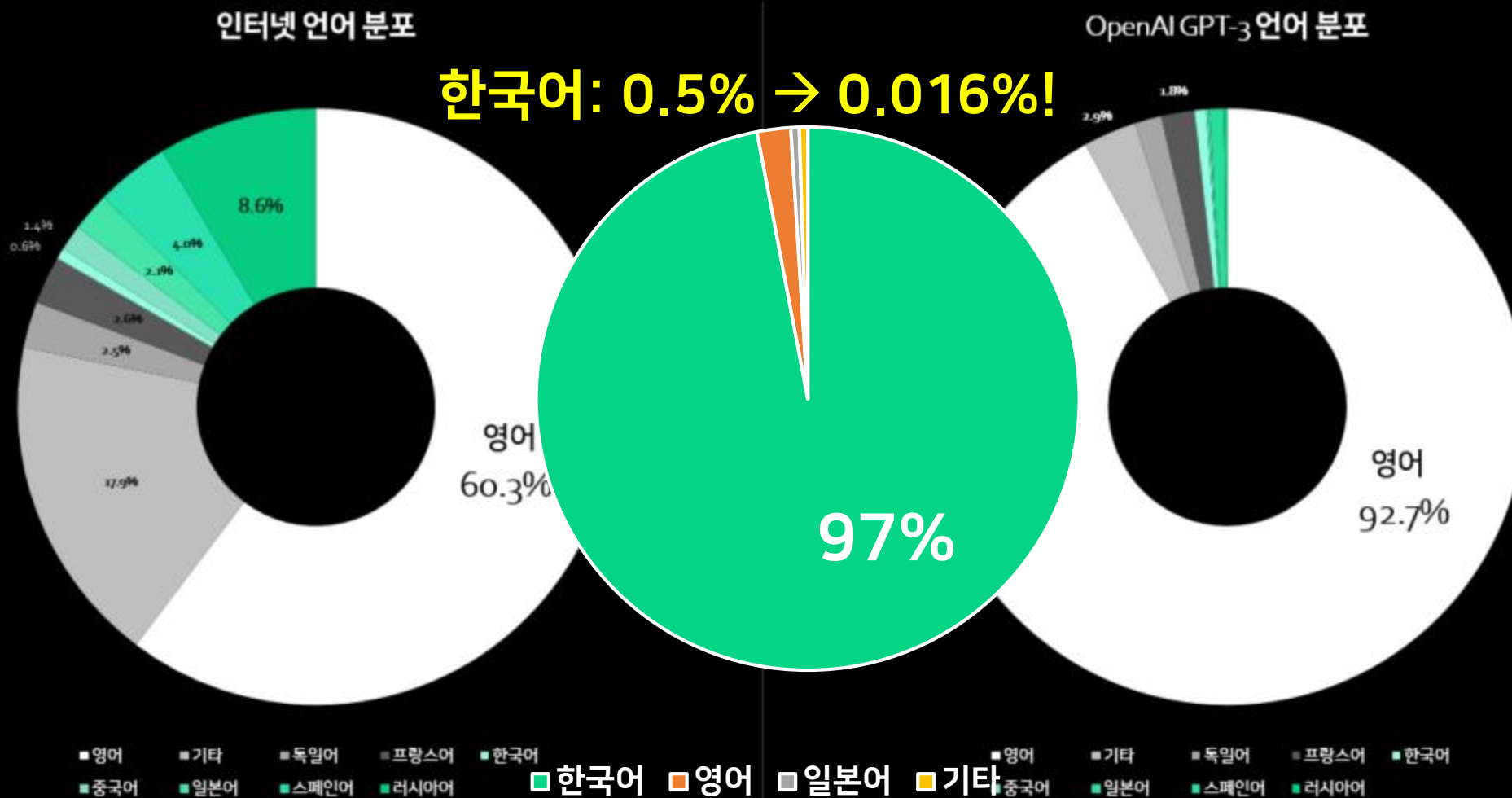
Develop Giant AI Language Models

DGX SuperPOD to Build Global-Scale
AI Supercomputing Infrastructure



- J. Huang's Keynote @ NVidia GTC 2021

AI주권: 한국어를 위한 No. 1 AI



https://en.wikipedia.org/wiki/Languages_used_on_the_Internet

https://github.com/openai/gpt-3/blob/master/dataset_statistics/languages_by_word_count.csv

NAVER Data

한국에서 가장 많은 사용자가 사용하는 플랫폼 네이버는 고품질의 풍부한 데이터 매지니먼트 역량을 보유 하고 있습니다.

[Kim et al. EMNLP 2021]

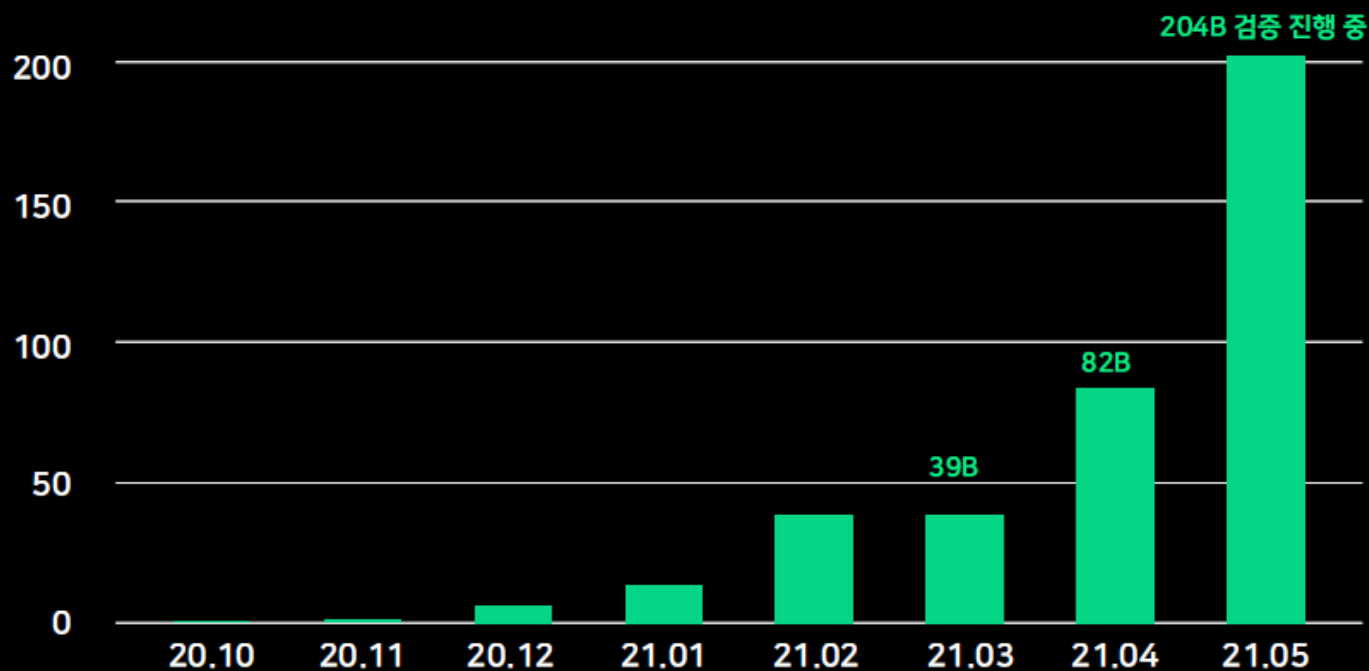


HyperCLOVA의 한국어 모델

모델 학습 현황

760M 모델을 사용하여 학습을 위한 실험을 진행하였으며, 단계적으로 모델의 크기를 키워가면서 82B 모델까지 검증을 완료 하였고, 이 경험을 바탕으로 204B 모델 검증 진행 중입니다.

[Kim et al. EMNLP 2021]



언어 모델 생산 현황 (크기)

HyperCLOVA Studio

몇 가지 예제만으로 필요한 AI 서비스를 만들 수 있습니다.

[Kim et al. EMNLP 2021]

Input

감정 분석기

#준여든다

이것은 문장 감정 분석기 입니다.

문장: "기분 진짜 좋다"
감정: 긍정

문장: "아오 진짜 짜증나게 하네"
감정: 부정

문장: "이걸로 보내드릴게요"
감정: 중립

문장: "새로 나온 뮤비 봤어? 장난아님"
감정: 긍정

문장: "택배가 도착했어요"
감정: 설레임
###

→ 지시문 및 설명

→ 예제

Submit



문장: "어? 이쁘다^^"

감정: 기쁨

###

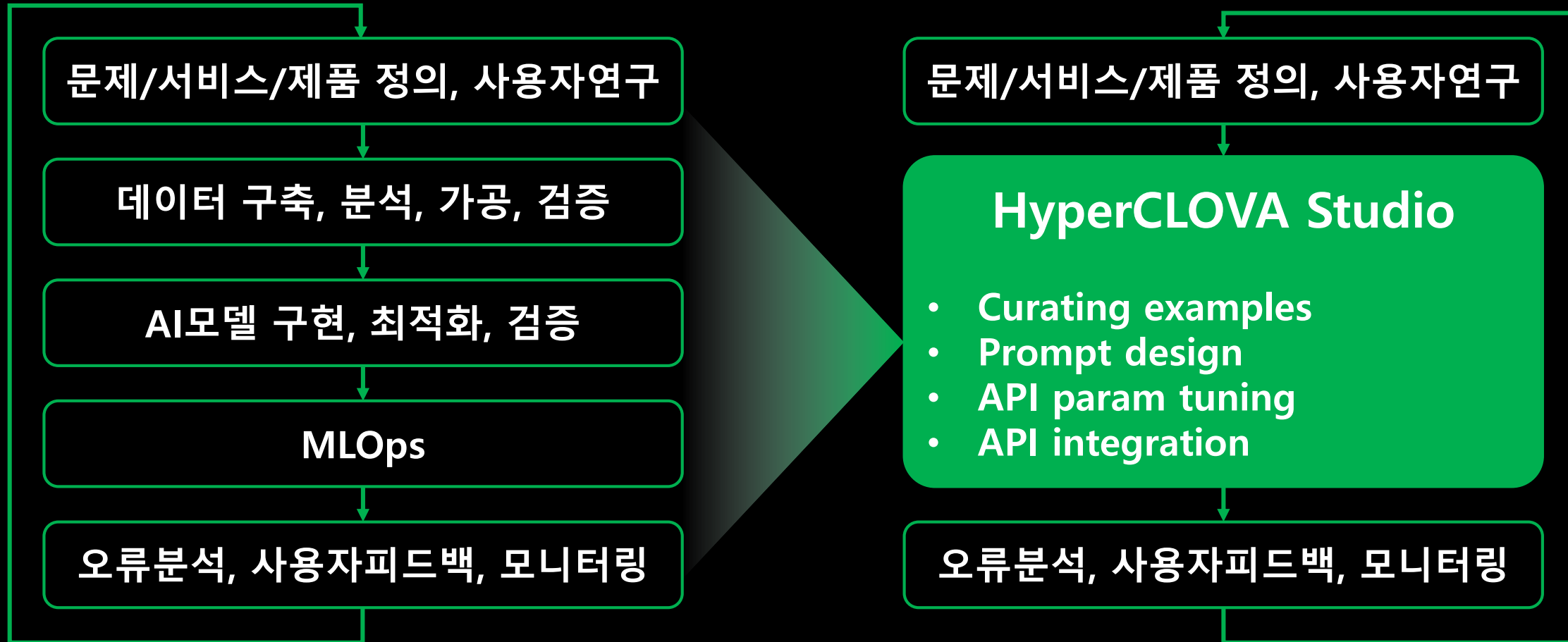
문장: "오늘 날씨 정말 좋다"

감정: 긍정



HyperCLOVA Studio

HyperCLOVA가 가져온 새로운 AI서비스 개발 패러다임 변화



HyperCLOVA Studio

손쉽게 만들어진 여러 AI 들은 별도의 Production Pipeline 없이 네이버, 라인, 생태계 안에서 완벽히 지원되며, 여러 환경의 디바이스 등으로 바로 배포하거나 사용할 수 있습니다.

<http://playground.nap.svc.ad1.io.navercorp.com/>



HyperCLOVA 활용 SME를 도와주는 글쓰기

창작을 보조해 주는 글을 작성합니다.

[Kim et al. EMNLP 2021]

CLOVA AiMD

쇼핑 기획전 제목 생성

#여성간절기자켓

#간절기아우터

#간절기점퍼

#강아지간식

#고양이간식

“
아직 바람이 차다.
완내스 봄자켓
”

“
댕냥이와 함께하는
즐거운 간식시간!
”



노출 적합도 채택률

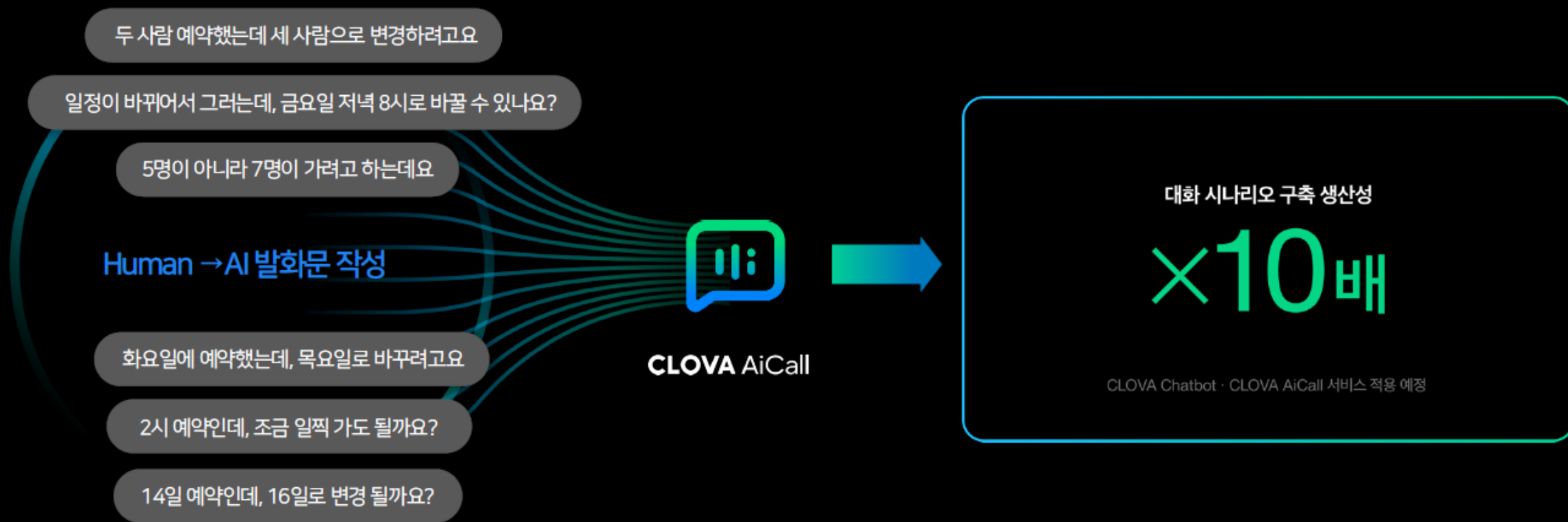
99%

스마트스토어 적용 예정

HyperCLOVA 활용 학습을 위한 데이터셋 자동생성

데이터 생성에 드는 비용과 시간을 줄여주어 AI 개발 Process를 가속화 할 수 있습니다.

[Kim et al. EMNLP 2021]



KLUE: 한국어 AI 성능 평가를 위한 데이터

#측정할수 없다면, 개선할 수 없다

[Park et al. NeurIPS 2021]

#누구나 무료로 자유롭게 쓸수있는 라이선스!

#Leaderboard: <https://klue-benchmark.com/>

#Data: <https://github.com/KLUE-benchmark/KLUE>

#Huggingface: <https://huggingface.co/klue>

Topic classification
(TC)

Semantic Textual
Similarity (STS)

Natural Language
Inference (NLI)

Named Entity
Recognition (NER)

Relation Extraction
(RE)

Dependency Parsing
(DP)

Machine Reading
Comprehension (MRC)

Dialog State Tracking
(DST)

현재 초대규모 AI의 한계

- Short memory
- Halluciation
- Vulnerability
- Biased / Toxic outputs
- Unstable training
- Multimodality

네이버 AI윤리 준칙: 사람을 위한 AI

네이버가 서울대학교 AI정책 이니셔티브 (SAPI)와 함께 만든 AI윤리준칙

<https://www.navercorp.com/value/aiCodeEthics>

사람을 위한 AI 개발

다양성의 존중

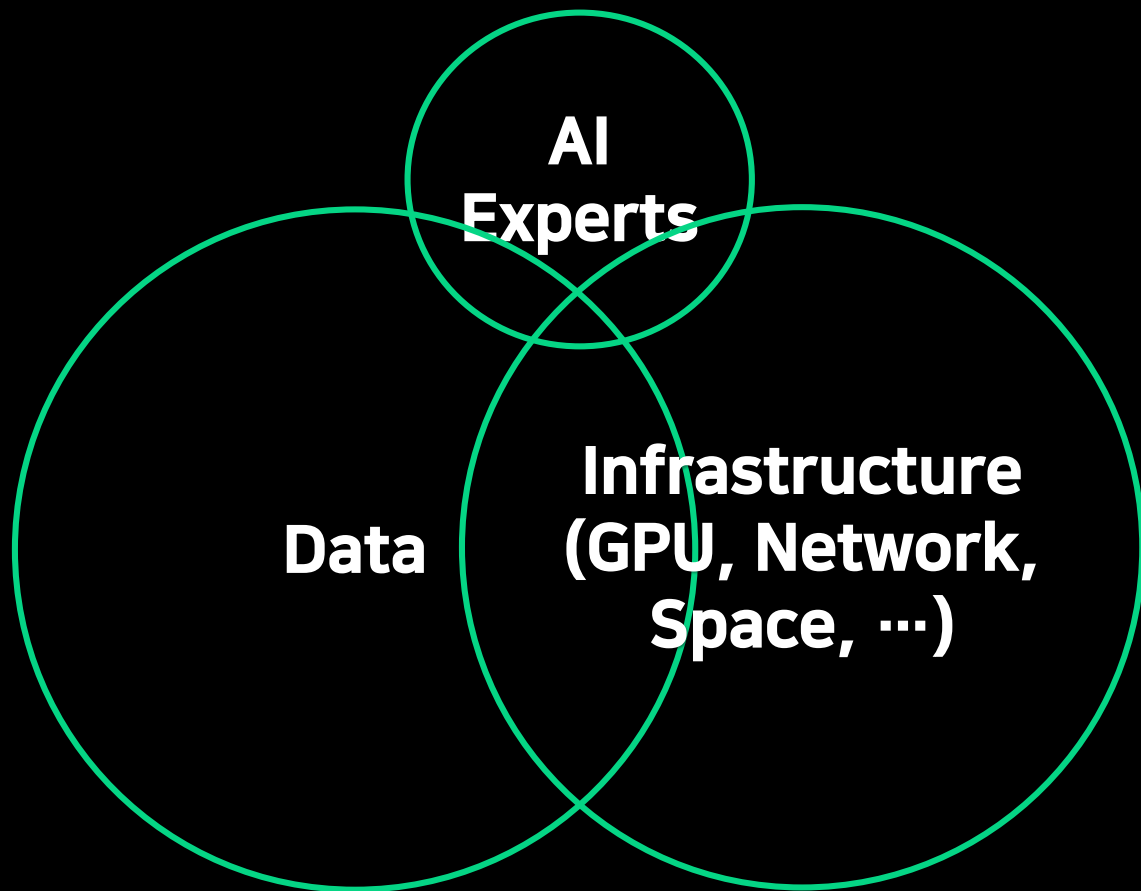
합리적인 설명과 편리성의 조화

안전을 고려한 서비스 설계

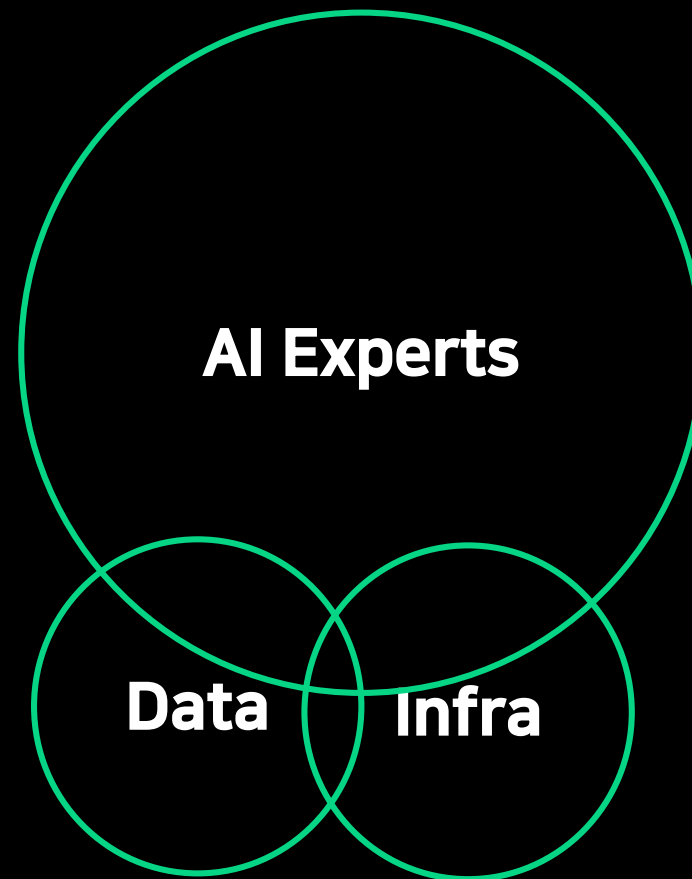
프라이버시 보호와 정보 보안

새로운 산합협력 연구 필요성

산업계



대학/연구소



서울대, KAIST AI 연구센터

- 공개된 데이터 문제를 넘어선 실세계의 AI 문제 발굴 / 해결
- 데이터, 인프라, 연구자 공유협력 → NAVER 연구원 겸임교수 및 공동지도
- 글로벌 초대규모 AI 기술 생태계 선도

HyperCLOVA

Hyperscale



Hypercreative



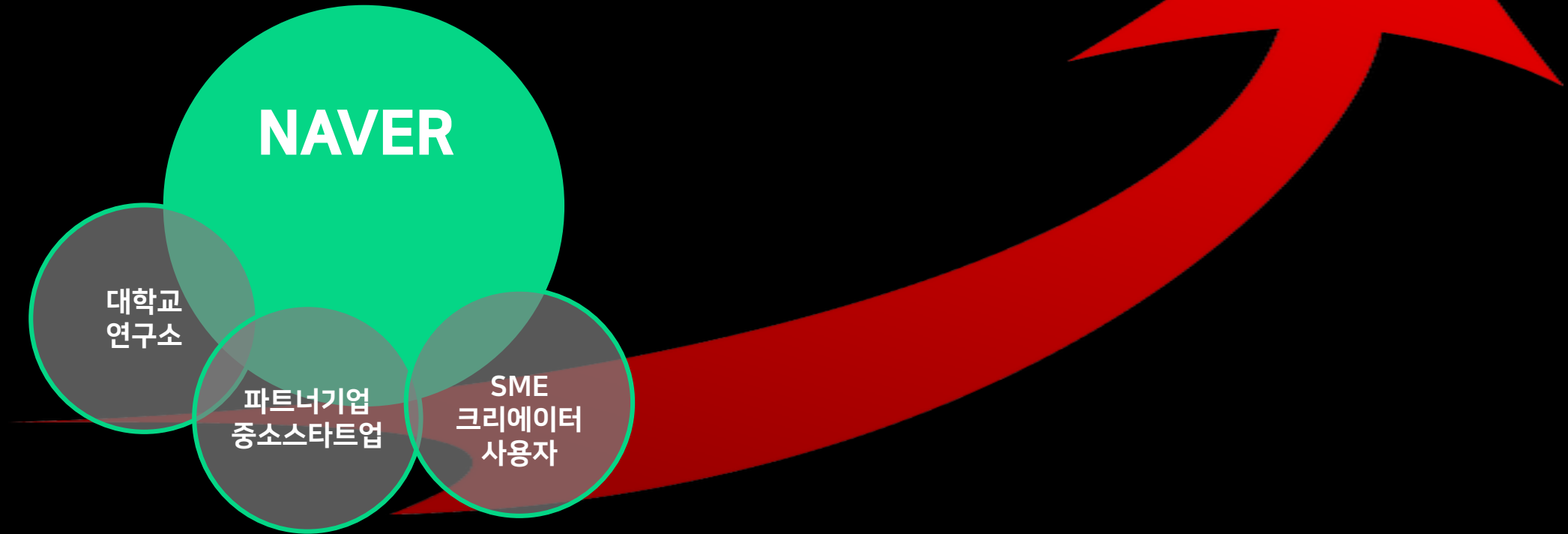
Global AI R&D 협력체계



제 3의 글로벌 AI 리더십을 향해

New AI R&D ecosystem → Beyond US and China

The 3rd Global AI Leadership



감사합니다

HyperCLOVA: 초대규모 AI의 현재와 미래

