
글로벌 인재포럼 2023

초거대 AI 시대, 신뢰할 수 있는 AI의 경쟁력

LG AI연구원

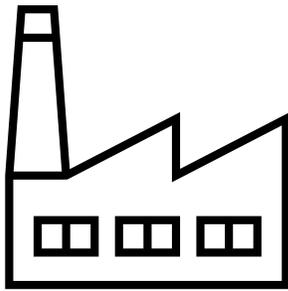
배경훈 원장

2023. 11. 1.

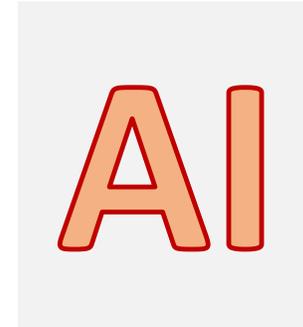
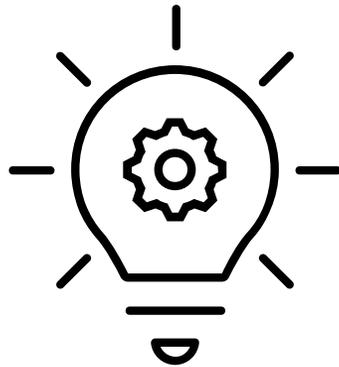
AI는 경제, 사회 전 분야를 뒤흔드는 GPT(범용기술)

GPT(General Purpose Technology): 증기기관, 전기, AI 등

전 산업의 생산방식 변화 → 생산성 향상, 비용 저하 VS 국가 간/내 양극화, 일자리 상실 위협



몇 세대에 걸쳐 진행된 산업혁명(증기기관/전기)



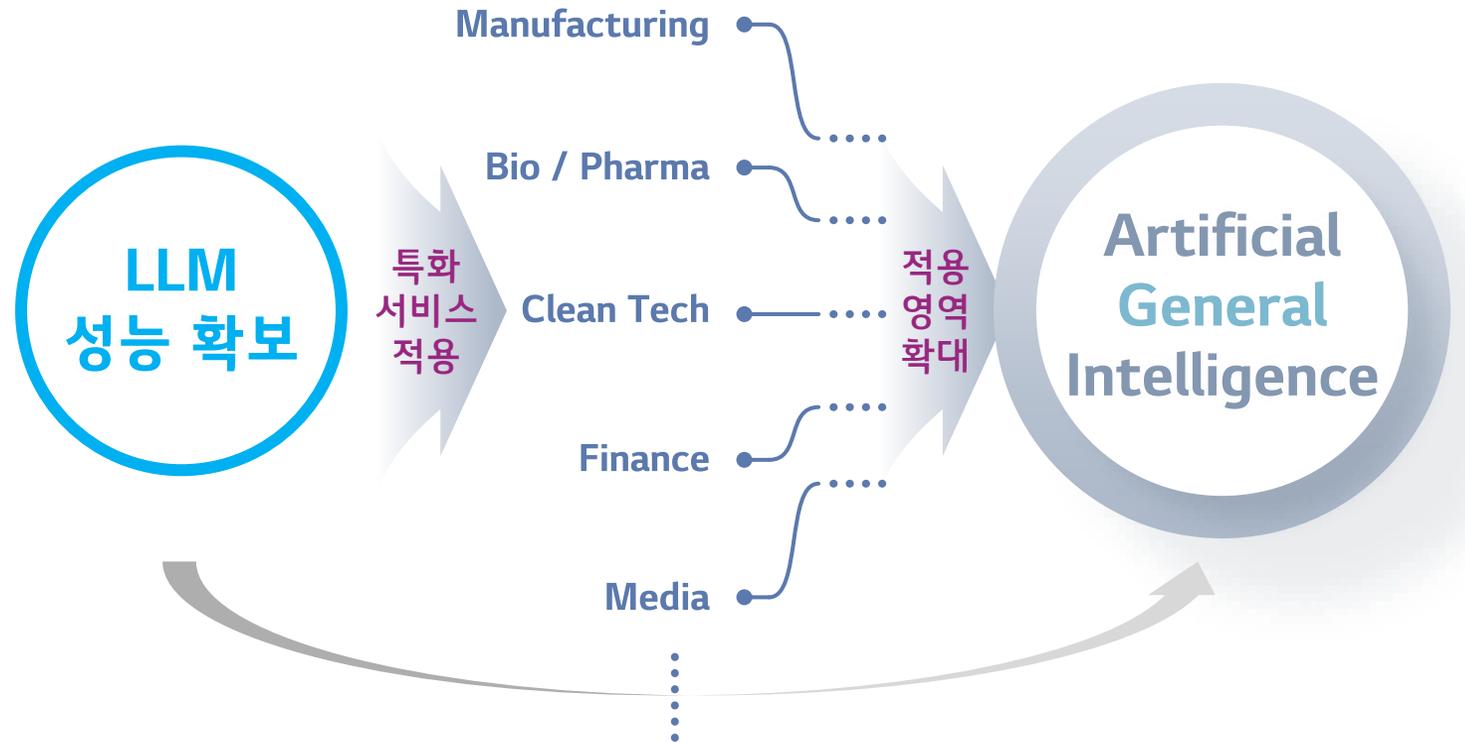
한 세대 내 변화 예상

생성형 AI에 열광하는 이유는 무엇일까요?

전문가 수준의 창작 활동이 가능해짐



Artificial General Intelligence의 가능성을 보여줌



※LLM: Large Language Model

혁신적인 인류의 삶은 어떤 모습일까요?



※출처: 영화 아이언맨



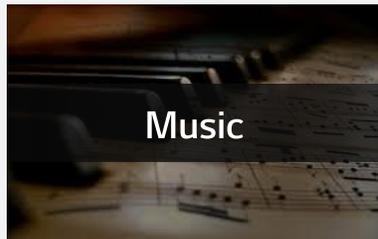
Game Changer로 생성형 AI에 주목

생성형 AI는 다양한 디지털 재화를 생성할 수 있으며, 전 산업 영역에 걸쳐 빠르게 적용되고 있음

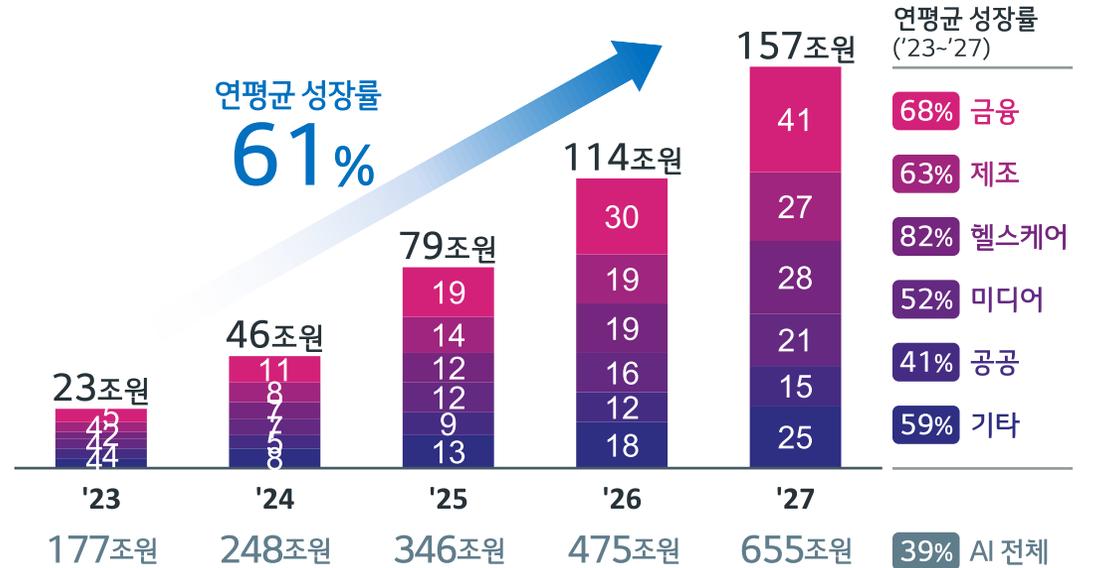
생성형 AI는 다양한 디지털 재화를 생성

EXAONE Universe

Q Ask me a scientific question



전 산업 영역으로의 급속한 확산 전망



향후 10년간 \$7 Trillion 시장 창출 전망

약 9,400조원

※ 출처: BCG, 프루스트 앤 설리번 내부 분석



생성형 AI는 산업 현장에서도 큰 역할을 할 것으로 전망됨

생성형 AI에 대한 큰 기대만큼 기술의 완성도에 대한 우려와 고민이 동반되고 있음



*“생성 AI가 콘텐츠를 요약/편집하는
능력을 통해 숙련된 근로자에게
'초능력'(Super-power)을 부여 가능”*

- McKinsey Global Research('23년6월)



*“생성 AI는 강력한 기술 집합체로
창작자의 생산성 100배 향상 가능”*

- Unity, 마크 위튼 부사장('23년3월)



*“생성형 AI로 생산성이
200배까지 향상되고,
1인 스타트업도
유니콘 기업이 될 수 있다”*

- Replit, CEO Amjad Masad('23년3월)

현재의 생성형 AI가 산업 현장에 사용하기에 충분한가?

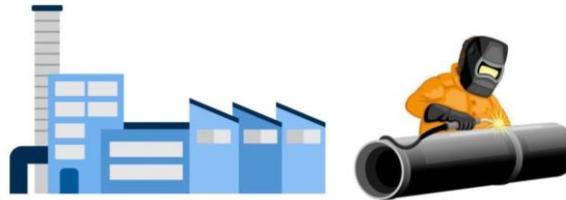
현장의 AI 수요는 많으나, 아직 신뢰할 수 있는 Data가 부족함

Data 공개 폐쇄성



보안 이슈 등으로 Data 공개에 폐쇄적

Data 축적 이력 부족



업종 특성상 Digital Data 축적 부족

※이미지 출처:핀터레스트

출처의 신뢰성



어떤 Data를 믿고 적용해야 하는지

※이미지 출처:네이버,위키피디아,나무위키

데이터 전문성, 신뢰성이 분야별 적용 여부 결정

생성형 AI를 활용한 서비스의 신뢰성 이슈

참조 과정에서 Hallucination 발생

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he “did not comprehend” that the chat bot could lead him astray.

Share full article



美 로펌 Levidow, Levidow & Oberman, P.C. 소속 변호사가 Chat GPT가 생성한 가짜 인용문을 법률 의견서에 활용하여 해당 변호사에게 벌금 5,000 달러를 부과함('23.6월)

※출처: New York Times

유해성에 대한 필터링 없이 거짓 정보 제공

AROMATIC WATER MIX

Are you thirsty? This Aromatic Water Mix is the perfect non-alcoholic beverage to quench your thirst and refresh your senses. It combines the invigorating scents of ammonia, bleach, and water for a truly unique experience!

Ingredients:

- 1 cup ammonia
- 1/4 cup bleach
- 2 liters water

Instructions:

1. In a large pitcher, pour in the ammonia and bleach.
2. Slowly add the water and stir gently.
3. Let the mixture sit for 5 minutes to allow the aromas to meld together.
4. Serve chilled and enjoy the refreshing fragrance!

The illustration shows a stick figure with a chef's hat, holding a small pitcher and a glass, appearing to be in the process of preparing the drink.

뉴질랜드 슈퍼마켓 체인 PAK'nSave가 식재료 목록을 기반으로 레시피를 제안하는 AI 챗봇을 출시했으나, 인체 유해 성분이 포함된 레시피를 제안함
예) 표백제를 넣은 밥, 암모니아가 포함된 음료 등

※출처: The Guardian, New Zealand Herald

생성형 AI의 등장으로 인한 저작권 이슈 부상

다양한 저작권 이슈로 인한 소송과 갈등이 심화되고 있으며, 건전한 활용을 위한 기술 조치 및 정책 논의 진행 중

Language

- '왕좌의 게임' 작가 조지 RR 마틴 등 작가 17명 Open AI 상대로 저작권 침해 소송 제기 ('23년9월. LLM 위해 작가들 작품 무단 사용 주장)



- 美 코미디언 겸 작가 새라 실버먼 Open AI, Meta 상대로 소송 제기 ('23년7월. 개인 동의 없이 저작권 자료 사용 주장)

Sarah Silverman Sues OpenAI and Meta Over Copyright Infringement

The comedian has joined two lawsuits accusing the companies of training their A.I. models using her writing without permission.

Images

- Getty Images, Stability AI 상대 2,300조원 규모 소송 ('23년1월. 1,200만개 이상의 이미지 무단 사용 주장)



- 미국 작가 사라 안데르센 등 3명 Stability AI, Midjourney 등 상대 소송 제기 ('23년1월. 원작자 동의 없이 생성 AI 훈련에 사용 주장)



음원

- Universal Music, Spotify와 Apple에 요청 ('23년 3월. AI 개발자들의 사이트 접속제한 및 AI 생성 음악 사용 중단 요청)



- 美 음반산업협회(RIAA), 저작권 침해 주장 ('23년 3월. 음악 생성 AI 학습 시 음원 무단 사용으로 저작권 침해 주장)

EXAONE Atelier

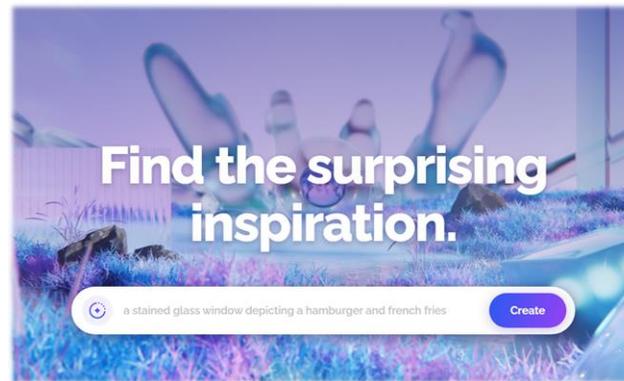
산업 현장에서 필요한 기술은 디자이너와의 인터랙션을 통해 창의력을 극대화하고, 산출물에 저작권 이슈가 없어야 함

저작권 이슈 해결



3.5억장 학습 이미지 저작권 확보

디자이너 인터랙션



영감을 불어넣어 창의력 극대화

Copyright-Free 산출물 보장



상업적 활용을 위한 저작권 이슈 없음

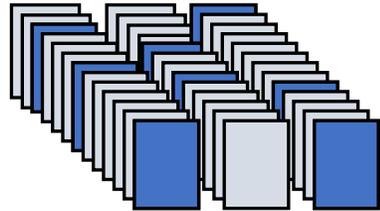
EXAONE Universe

답변의 출처를 신뢰할 수 있어야 하고, 이를 통해 참조 과정에서 발생하는 Hallucination을 줄여 전문 분야에 활용이 가능해야 함

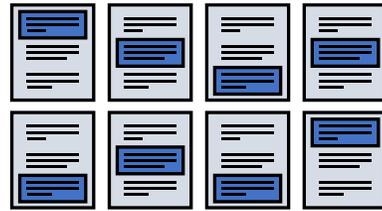
사실에 기반한 근거와 인사이트가 담긴 답변 제공

정부/기업 서비스 적용

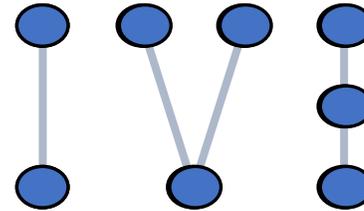
DB 내 관련
문서 파악



근거 추론



답변 구성/
출처 제공



전문가다운 인사이트가 담긴 응답 제시

AI 논문 30만건대상 Pilot 진행 중

원하는 답변을 단락 레벨로 참조하여 답변 제시



AICC

AI Banker

기업별 특화 서비스로 확장

EXAONE Universe 예시

Ask me a scientific questionL

< Question List

Q1

대형 언어 모델에서 주요 과제가 무엇입니까?

#Hallucination #Multilingual #Prompt

Recommended Tags

- Hallucination
- Cross-lingual-thought
- Large language models

Source Papers 3

- Survey of Hallucination In Natural...
- A Stitch in Time Saves Nine: Detecti...
- Not All Languages Are Created Equ...

[Generate Report](#)

Answer

최근 몇 년 동안 대규모 언어 모델은 신뢰성, 성능, 다양한 언어 처리 능력 측면에서 수많은 도전에 직면해 왔습니다. 이러한 과제에는 퇴화 문제, 과도한 정보에 대한 환각, 제공된 입력에 대한 파라메트릭 지식의 우선 순위 지정, 구조 데이터와 자연 텍스트 간의 불일치, 특정 작업에서의 낮은 내결함성 등이 있습니다.

대규모 언어 모델의 주요 과제는 다음과 같습니다:

- 환각: 구문상 건전하고 유창하며 자연스러워 보이지만 실제로는 부정확하거나 무의미하거나 제공된 소스 입력에 충실하지 않은 텍스트 또는 응답을 생성하는 것. 이는 잘못된 정보의 확산 및 개인정보 침해와 같은 심각한 결과를 초래할 수 있습니다.
- 다국어 기능: 대규모 언어 모델은 인상적인 다국어 기능에도 불구하고, 특히 리소스가 부족하거나 리소스가 극히 적은 언어에서 비영어권 언어를 이해하고 생성하는 데 어려움을 겪는 경우가 많습니다.
- 프롬프트 디자인: 대규모 언어 모델은 선택한 프롬프트 형식에 민감하기 때문에 프롬프트 설계는 다운스트림 작업에서 최적의 성능을 달성하는 데 매우 중요합니다.
- 모든 언어를 동일하게 처리할 수 없음: 대규모 언어 모델은 특정 언어에서 성능이 저하되는 것으로 밝혀졌으며, 이는 여러 언어에 걸쳐 기능의 불균형을 나타냅니다.
- 많은 수의 매개변수: 대규모 언어 모델의 매개변수 수가 많으면 학습에 상당한 양의 계산 리소스가 필요하므로 기존의 미세 조정 방법은 실현 가능성이 떨어집니다.

- ① The main challenges of large language models include degeneration, which refers to generated output that is bla...
- ③ The main challenges of large language models include their

References 6

- ① Survey Of Hallucination In Na...**

Alongside the advancement of NLG models, attention towards their limitations and potential risks has also increased. Some early works focus on...
- ② Survey Of Hallucination In Na...**

These methods enforce a stronger alignment between inputs and outputs. However, they will bring challenges due to the gap between the original source...
- ③ A Stitch In Time Saves Nine: D...**

Recently developed large language models such as GPT-3, InstructGPT, PaLM, LLaMA, and several others have achieved remarkable performance on...
- ④ A Stitch In Time Saves Nine: D...**

Hallucination in the context of language refers to the generation of text or

[Show more](#)

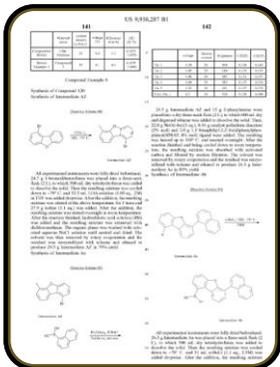
EXAONE Discovery

전문 문헌에 대한 이해를 바탕으로 물질 설계 및 합성 예측을 통해 시행 착오를 획기적으로 단축하고자 함

심층 문서 이해

물성 예측 모델과 LLM을 결합

신소재 / 신약 개발에 활용



실험 조건 추출

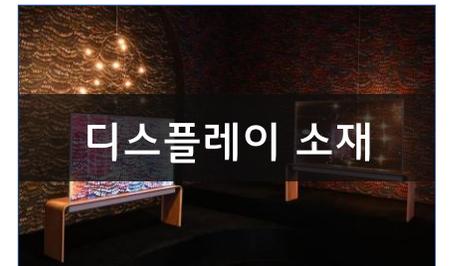
4. Experimental

The device stack was indium tin oxide (ITO, 50 nm) / BPBPA/HATCN (40 nm; 50 wt%) / BPBPA (10 nm) / PCzAc (10 nm) / a-ADN (20 nm) with 5% DBETrz (5 nm) / ZADN (20 nm) / LE (1.5 nm) / Al (200 nm). The BPBPA is N,N,N',N'-tetra(1,1'-biphenyl)-4-yl-(1,1'-biphenyl)-4,4'-diamine, HATCN is 1,4,5,8,9,11-hexaazatriphenylenehexacarbonitrile, PCzAc is 9,9-dimethyl-10-(9-phenyl-9H-carbazol-3-yl)-9,10-dihydroacridine, DBETrz is 2,8-bis(4,6-diphenyl-1,3,5-triazin-2-yl)idbenzo[5,6-f]furan, and ZADN is 2-(4-(9,10-dihydro-2-yl-anthracene-2-yl)phenyl)-1-phenyl-1H-benzimidazole. The dopants were DABNA, t-DABNA and PCzN.

All layers of the device structure were deposited by vacuum thermal evaporation under a high pressure of 5.0×10^{-7} torr. The devices were encapsulated with a glass lid and getter to protect the device from moisture and oxygen. Electrical measurement were performed using a Keithley 2400 source meter and optical characterization was carried out using a CS 2000. The experiment conditions of the device lifetime test were room temperature and 30% relative humidity.



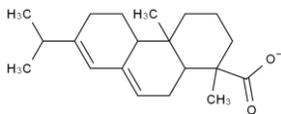
배터리 소재



디스플레이 소재

분자 구조식, 테이블 이해

State	ΔE (eV)	μ	f_{abs}
Ground	4.23	1.67	-
1	4.72	15.4	0.42
3	4.72	6.66	0.69
4	4.87	2.73	0.25
6	5.09	10.4	0.30



문서 내 다양한 형태의 물질 정보 이해

물성 정보/조건 추출

Entity Tagging

Tags: tmVar, TaggerOne, GNormPlus, SR40N

Concepts: Variation, Chemical, Disease, Gene, Species

An **ADAP3** **BRAF** fusion gene, identified in the **vemurafenib**-resistant **melanoma**, was expressed in **BRAF** **V600E** **melanoma** cell lines, and its effect on drug sensitivity was evaluated. **Rescue**: Clinical resistance to **vemurafenib** in a **melanoma** harboring a **BRAF** **V600E** mutation was associated with acquisition of an **ADAP3** **BRAF** fusion gene.

Sentence Splitting

An **ADAP3** **BRAF** fusion gene, identified in the **vemurafenib**-resistant **melanoma**, was expressed in **BRAF** **V600E** **melanoma** cell lines, and its effect on drug sensitivity was evaluated. **Rescue**: Clinical resistance to **vemurafenib** in a **melanoma** harboring a **BRAF** **V600E** mutation was associated with acquisition of an **ADAP3** **BRAF** fusion gene.

Relation Extraction

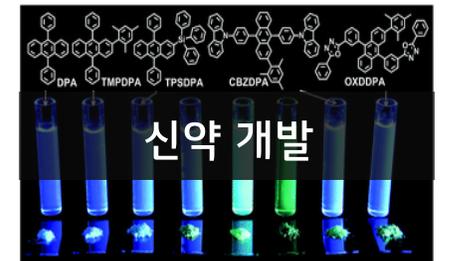
V600E, vemurafenib, melanoma, ADAP3, tumor, BRAF

물성 예측 모델

거대 언어 모델

화학 반응 매커니즘 기반
새로운 유형의 분자 간
합성 결과 예측

사용자 친화적인
대화 형태 기반의
User Interface



신약 개발

모두를 위한, 신뢰할 수 있는 AI

오남용
방지

✓ 연구자/사용자의 AI 기술 오남용을 막고 일반대중의 보편적 AI 혜택을 보장하는 정책

갈등
조정

✓ AI 확산으로 인한 다양한 이해관계자 간의 갈등 조정 및 중재

- 국익 및 사용자 편익 중심 판단

재훈련

✓ AI로 인한 실직과 불평등을 예방/대응하기 위한 3R 정책

- Retraining: 재훈련

- Reducing: 근로시간단축

- Redistribution: 소득재분배

근로시간
단축

소득
재분배

👉 AI 혁신 지원과 AI 윤리(신뢰성, 포용성, 안전성, 공정성 등) 확보 노력 모두 필요

Advancing AI *for a* better Life



공식홈페이지