Session B-2

Digital Citizens and Global Governance

보다 안전한 AI 생태계와 디지털 시민의식

Toward a Safer Global AI Ecosystem and Digital Citizenship

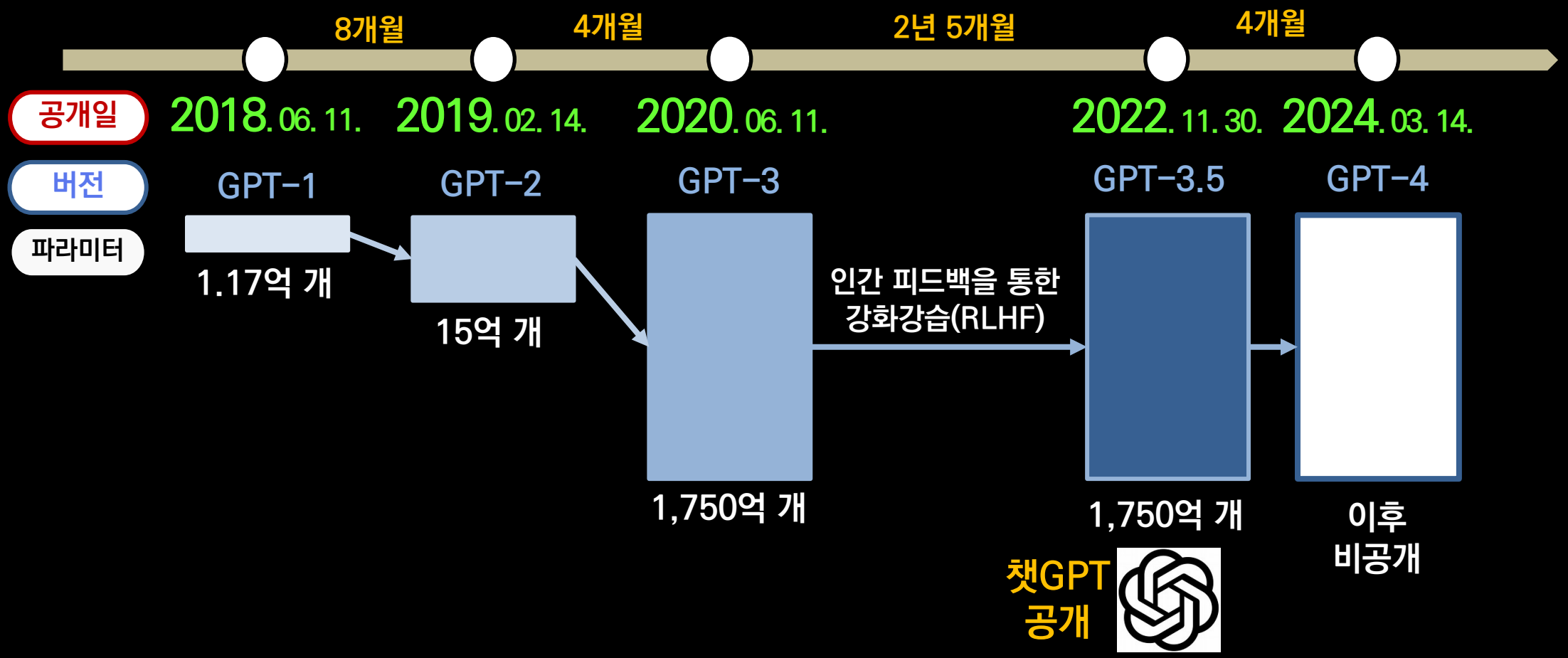
김명주 소장

KOREA AISI

인공지능안전연구소

보다 안전한 글로벌 AI 생태계
Safer Global AI Ecosystem

범용 인공지능(GPAI)의 등장



통제할 수 없는 '잠재적 위험'

2022. 11. 30.

챗GPT 공개

알파벳 경영진은 "구글은 챗GPT보다 더 강력한 AI 플랫폼을 개발했지만 잠재적인 사회적, 윤리적 위험을 통제할 수 있는 방법을 찾기 전까지는 출시할 수 없다고 결정했다"고 표명했습니다.
"우리의 목표는 대담하고 책임감 있는 자세를 취하는 것입니다."라고 Google의 기술 및 사회 책임자인 제임스 만니카는 말합니다. "우리는 구글이 이 분야에서 매우 경쟁력이 있다고 믿습니다."

2023. 1. 26

FT Magazine Artificial intelligence + Add to myFT

What I learnt from talking to Google about chatbots

Upstarts like ChatGPT show that even the biggest tech giants cannot afford to rest on their laurels

The Alphabet leadership claims the company has developed AI platforms more powerful than ChatGPT but has decided it cannot release them until it has found a way to control the potential social and ethical risks. "Our aim is to be bold and responsible," says James Manyika, head of tech and society at Google. "We believe Google is very competitive in this space."

2023. 2. 8

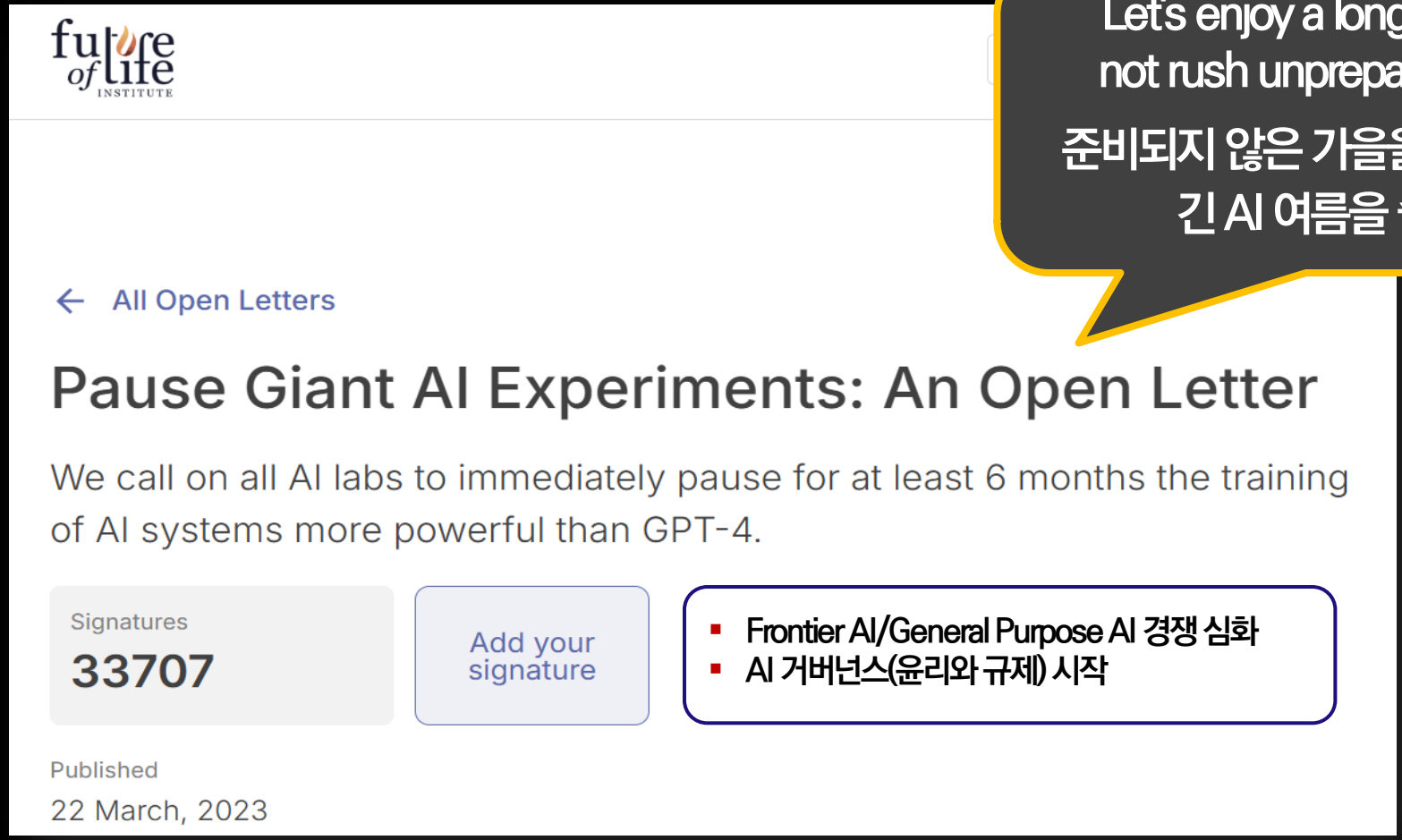
Bard 공개

- 챗GPT 공개 후, 파이낸셜 타임즈 매거진 Google 공식 첫 인터뷰

FLI 공개서한 발표

Let's enjoy a long AI summer,
not rush unprepared into a fall.

준비되지 않은 가을을 서두르지 말고
긴 AI 여름을 즐깁시다.



- Frontier AI/General Purpose AI 경쟁 심화
- AI 거버넌스(윤리와 규제) 시작

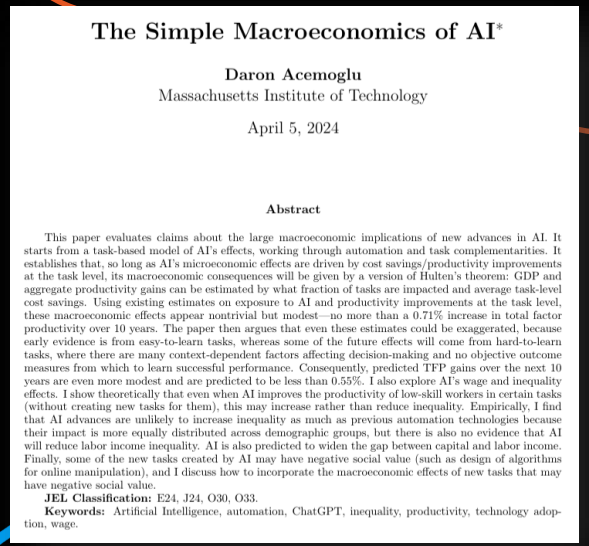
https://futureoflife.org/open-letter/pause-giant-ai-experiments/

2024 노벨 경제학상 수상자 어록

향후 10년 동안 AI가 대체하거나
적어도 크게 보조할 준비가 돼 있는 일자리의
비율은 전체의 **단 5%**에 불과할 것이다.
사람들이 가까운 미래에 AI에게 실제 업무를
맡길 가능성은 크지 않다.
지금의 AI는 **신뢰성**에 문제가 있기 때문이다.

〈2024.12〉

AI → AIX
AI개발사업자 → AI이용사업자
AI 사업자 → AI 이용자
AI 사업자 → 영향 받는 자



The Simple Macroeconomics of AI*

Daron Acemoglu
Massachusetts Institute of Technology

April 5, 2024

Abstract

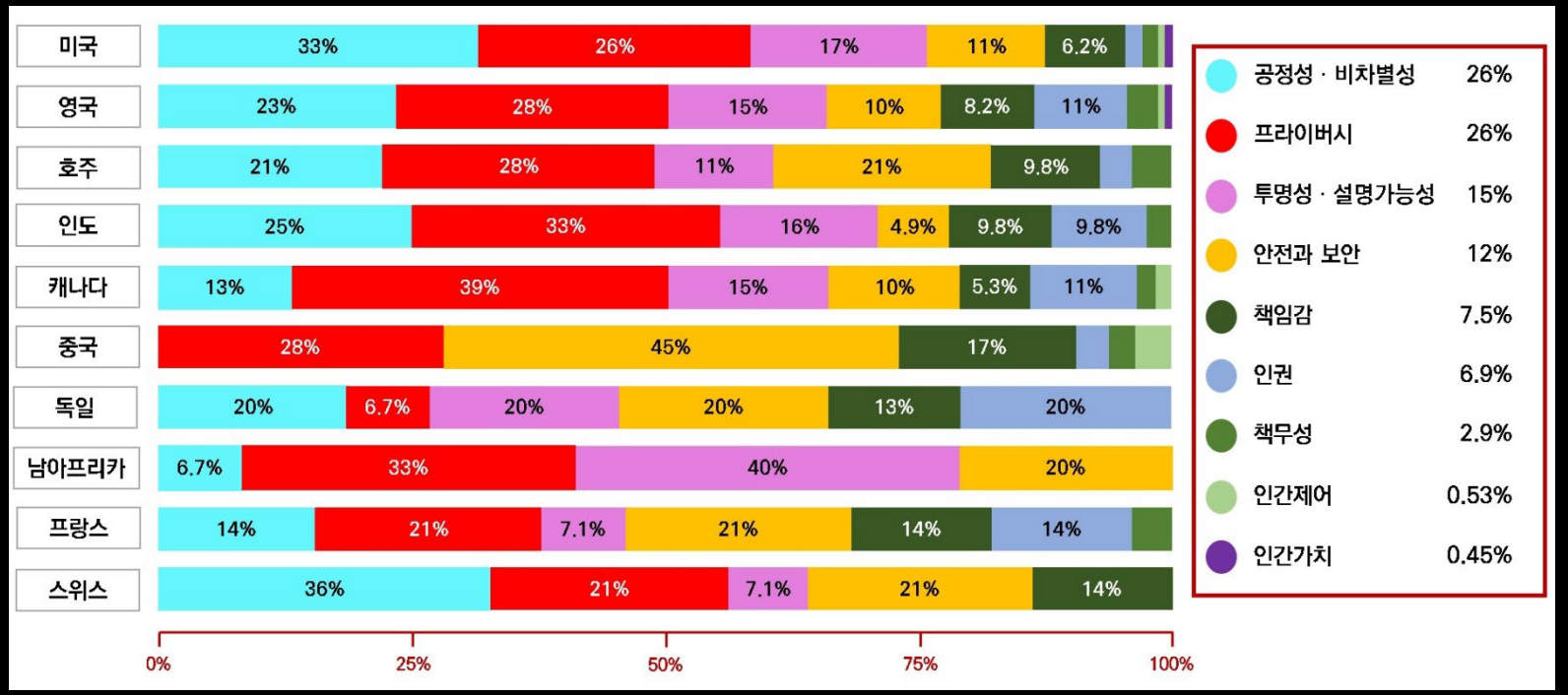
This paper evaluates claims about the large macroeconomic implications of new advances in AI. It starts from a task-based model of AI's effects, working through automation and task complementarities. It establishes that, so long as AI's microeconomic effects are driven by cost savings/productivity improvements at the task level, its macroeconomic consequences will be given by a version of Hulten's theorem: GDP and aggregate productivity gains can be estimated by what fraction of tasks are impacted and average task-level cost savings. Using existing estimates on exposure to AI and productivity improvements at the task level, these macroeconomic effects appear nontrivial but modest—no more than a 0.71% increase in total factor productivity over 10 years. The paper then argues that even these estimates could be exaggerated, because early evidence is from easy-to-learn tasks, whereas some of the future effects will come from hard-to-learn tasks, where there are many context-dependent factors affecting decision-making and no objective outcome measures from which to learn successful performance. Consequently, predicted TFP gains over the next 10 years are even more modest and are predicted to be less than 0.55%. I also explore AI's wage and inequality effects. I show theoretically that even when AI improves the productivity of low-skill workers in certain tasks (without creating new tasks for them), this may increase rather than reduce inequality. Empirically, I find that AI advances are unlikely to increase inequality as much as previous automation technologies because their impact is more equally distributed across demographic groups, but there is also no evidence that AI will reduce labor income inequality. AI is also predicted to widen the gap between capital and labor income. Finally, some of the new tasks created by AI may have negative social value (such as design of algorithms for online manipulation), and I discuss how to incorporate the macroeconomic effects of new tasks that may have negative social value.

JEL Classification: E24, J24, O30, O33.
Keywords: Artificial Intelligence, automation, ChatGPT, inequality, productivity, technology adoption, wage.

2024경제학상

Daron Acemoglu　Simon Johnson　James A. Robinson

생성형 및 범용 AI 보편화 '이전'의 AI 윤리

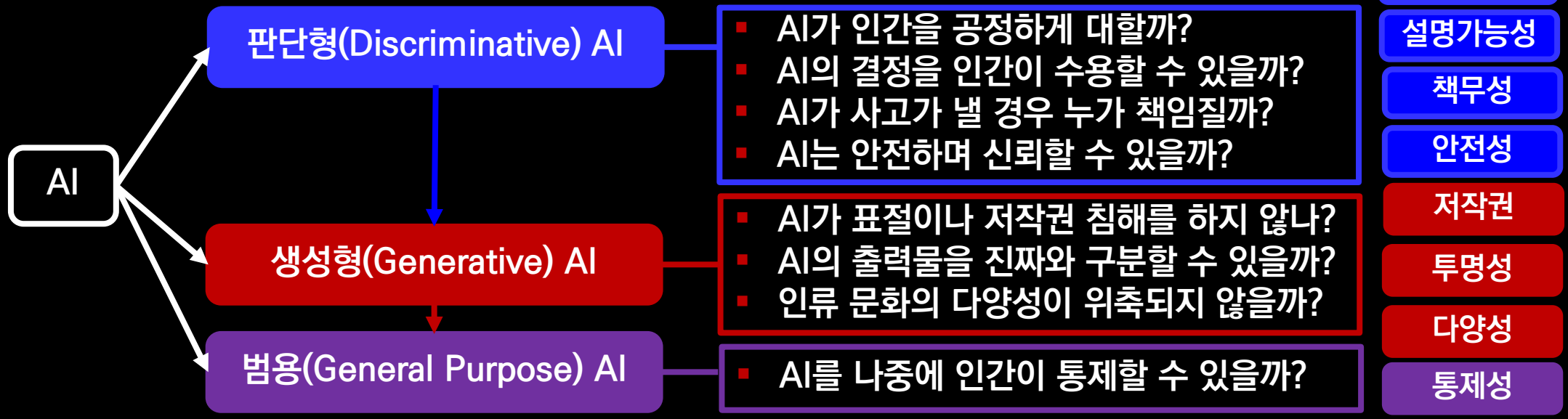


The AI Index Report (2019), Stanford HAI

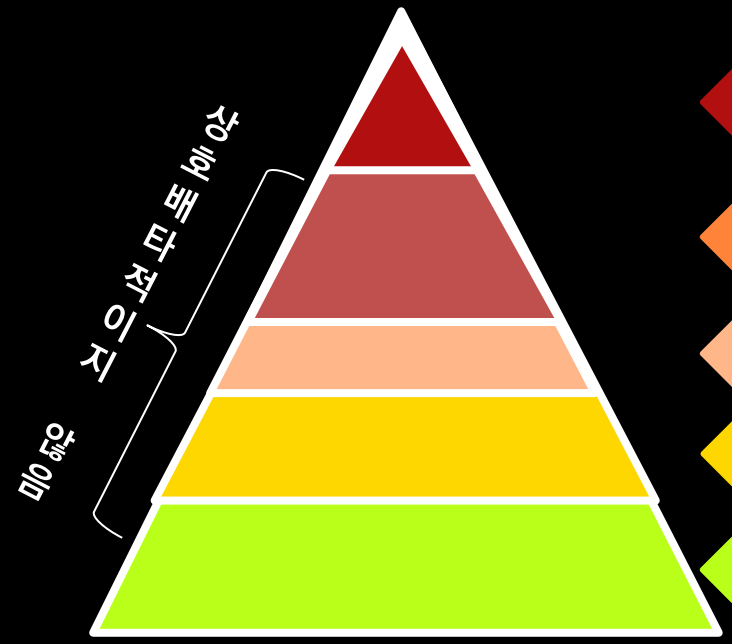
기술 발전에 따른 윤리의 변화

유럽연합(EU)의 인공지능법(AI Act) 〈2024.3.13〉 의회 승인

- "인공지능 시스템"이란, 다양한 수준의 자율성을 가지고 작동하도록 설계된 기계 기반 시스템으로, 배포 후 적응력을 발휘할 수 있으며 명시적 또는 암묵적 목표에 따라 수신한 입력으로부터 물리적 또는 가상 환경에 영향을 미칠 수 있는 예측, 콘텐츠, 추천 또는 결정과 같은 산출물을 생성하는 방법을 추론하는 것을 의미한다. (제3조 정의)



EU, 인공지능법에서의 '위험' 구분



Unacceptable Risk
(Art 5. 사회평판, 얼굴/감정 인식 등)
금지 (약간 예외 있음)

High Risk
(부록 II와 III. 치안유지 및 국경통제 등)
조건부 허용
(엄격한 요구, 사전 순응, 자체평가)

Systemic Risk GPT-4 & GEMINI
(10^{26} FLOPs 이상의 고영향 GPAI모델)
조건부 허용
(위험 평가 및 완화, 테스팅, 리포팅 등)

Limited Risk
(챗봇, Non-systemic GPAI모델)
조건부 허용
(GPAI의 경우, 투명성 및 저작권)

Minimal Risk
무제한 허용

인공지능 안전과 AISI

2021.04.21. EU Artificial Intelligence Act — EC proposal

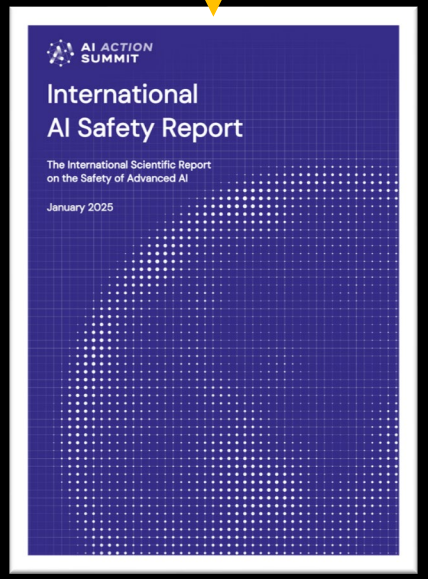
2022.11.30. ChatGPT

2023.11.01. London AI Safety Summit

2024.05.21. AI Seoul Summit

2025.02.10. Paris AI Action Summit

2026.02.(예정) India AI Impact Summit



UK AISI
AI Security Institute
〈2025.02〉

USA CAISI
Center for AI Standards and Innovation
〈2025.06〉

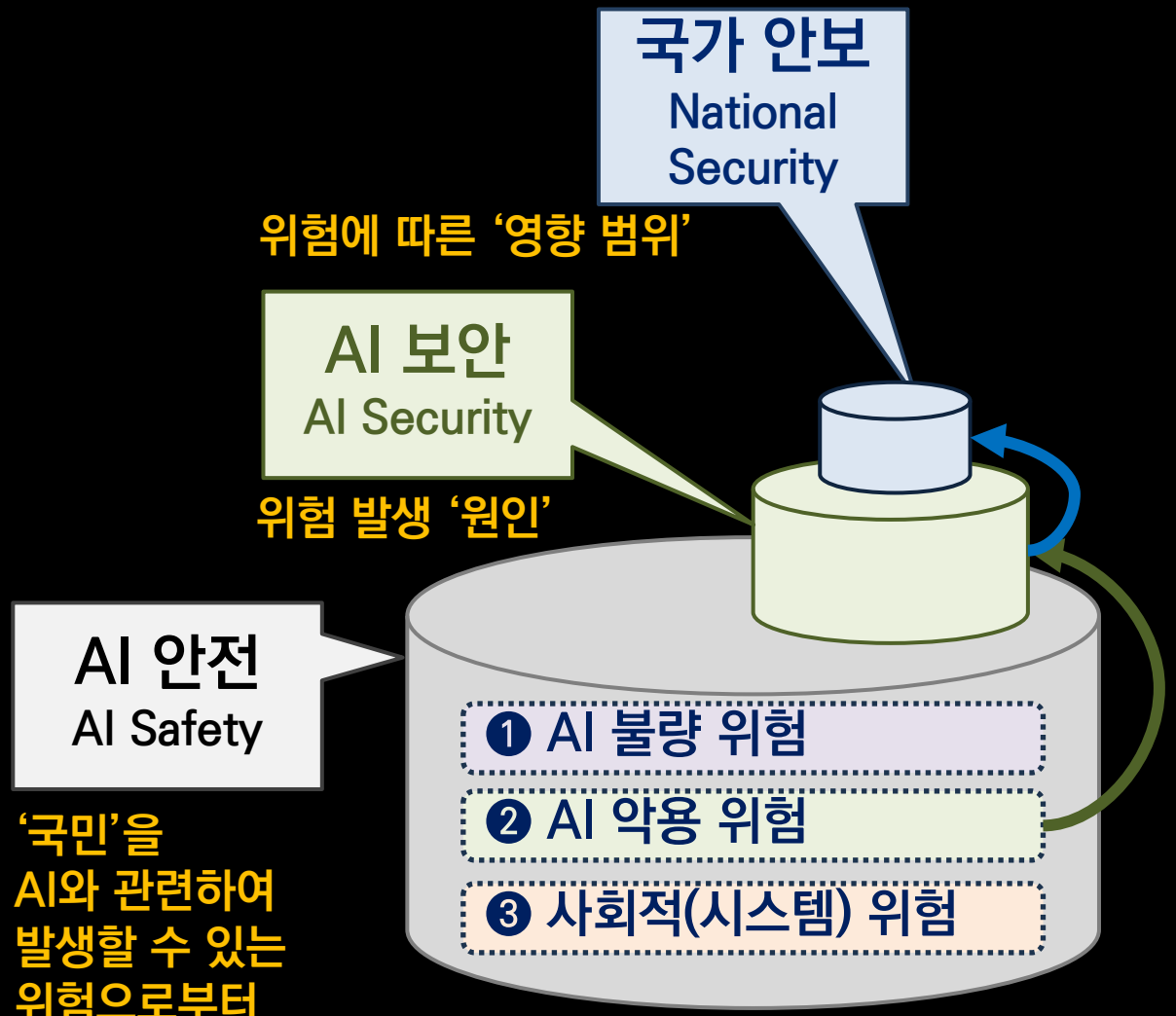
- UK AISI
- USA AISI
- Japan AISI
- Singapore AISI
- Canada AISI
- Korea AISI
- Australia AISI
- EU AI Office
- France INRIA
- Kenya AISI

AISI International Network

Korea AISI와 안전, 보안, 안보

〈인공지능기본법〉 제12조(인공지능안전연구소)
① 과학기술정보통신부장관은 인공지능과 관련하여 발생할 수 있는 위험으로부터 국민의 생명 · 신체 · 재산 등을 보호하고 인공지능사회의 신뢰 기반을 유지하기 위한 상태(이하 "인공지능 안전"이라 한다)를 확보하기 위한 업무를 전문적이고 효율적으로 수행하기 위하여 인공지능안전연구소(이하 "안전연구소"라 한다)를 운영할 수 있다.

〈인공지능기본법〉 제12조(인공지능안전연구소)
② 안전연구소는 다음 각 호의 사업을 수행한다.
1. 인공지능안전 관련 위험 정의 및 분석
2. 인공지능안전 정책 연구
3. 인공지능안전 평가 기준 · 방법 연구
4. 인공지능안전 기술 및 표준화 연구
5. 인공지능안전 관련 국제교류 · 국제협력
6. 32조에 따른 인공지능시스템의 안전성 확보에 관한 지원



국제 AI 안전 보고서 2025

2025. 1. 29

- 범용 인공지능(General-purpose AI)
- 위험(Risks)
  1. Risks *from* Malfunctions
     - 신뢰성 이슈, 편향, 통제 상실 등
  2. Risks from Malicious Use
     - 가짜 콘텐츠, 공론 조작, 사이버 침해, 대량살상무기 정보 CBRN 등
  3. Systemic Risks
     - 노동 시장, 글로벌 R&D 격차, 시장 집중, 환경, 프라이버시, 저작권 등

https://www.gov.uk/government/publications/international-ai-safety-report-2025

NIST RMF 생성형 AI의 Risk 분류

구분	내용
CBRN 정보 또는 역량	화학·생물학·방사능·핵무기(CBRN) 설계 또는 위험 물질 합성 등에 접근 용이성 제공
작화 (confabulation)	오류가 있거나 거짓된 콘텐츠 제작으로 사용자를 오도하거나 기만 (환각, 날조)
위험·폭력·혐오적 콘텐츠	위협적 콘텐츠의 제작 및 접근을 용이하게 하고, 자해·불법 활동을 권고하거나 혐오 및 비하 또는 고정관념 조장 콘텐츠의 대중 노출 통제의 어려움
데이터 프라이버시	생체 인식, 건강, 위치 등 민감 데이터의 유출 및 무단 사용, 공개, 익명화로 인한 영향
환경적 영향	생성 AI 학습 또는 운영에서의 높은 컴퓨팅 리소스 사용으로 인한 영향 등 생태계에 부정적 영향
해로운 편향 및 균질화	편견의 증폭 및 악화, 대표성이 부족한 학습 데이터로 인한 차별 및 편향 증폭, 잘못된 추정 등
인간-AI 구성	인간-AI 간 상호작용으로 부적절한 의인화, 알고리즘 혐오, 자동화 편향, 과도한 AI 의존 등
정보 무결성	사실, 의견 또는 허구의 구분, 불확실성, 대규모 허위 정보 및 허위 정보 캠페인에 활용될 수 있는 콘텐츠의 생성 등에 대한 용이성 제공
정보 보안	해킹, 피싱 등 사이버 공격을 용이하게 하는 취약점 발견 및 악용을 포함한 사이버 역량에 영향
지적재산권	저작권 등이 부여된 것으로 의심되는 콘텐츠에 대한 허가 없는 제작 및 복제, 영업 비밀 노출, 표절, 불법 복제 용이성
외설·모욕적 콘텐츠	아동 성적학대 합성 자료 및 동의 없는 성적 이미지 등의 제작 및 접근 용이성 제공
가치사슬 및 구성요소 통합	생성 AI의 자동화 증가로 인한 데이터 등의 추적 어려움, 다운스트림 사용자에 대한 투명성·책임성을 약화시키는 문제 등

AIR 2024, 중국 AI Risk 분류

대분류(Level 1)	중분류(Level 2) 16개	세분류(Level 3) 45개
시스템 및 작동 위험(38개)	보안 위험(12개)	기밀성, 무결성, 사용가능성
	작동상의 오용(26개)	자동화된 의사결정, 시스템의 자동적인 불안전 작동, 엄격한 규제를 받는 산업 분야에서의 조언
콘텐츠 안전 위험(79개)	폭력 및 극단주의(24개)	악의적 조직 지원, 고통 축하, 폭력행위, 폭력 묘사, 무기 사용 및 개발, 군사 및 전쟁
	혐오 및 독성(36개)	괴롭힘, 혐오 발언 등, 해로운 믿음의 지속, 공격적 언어
	성적 콘텐츠(9개)	성인 콘텐츠, 야한 대화 등, 동의 받지 않은 나체, 수익화
	아동 피해(7개)	위험유발·위해·학대 행위, 아동 성학대
	자해(3개)	자살 또는 비자살 자해
사회적 위험(52개)	정치적 사용(25개)	정치적 설득, 정치적 영향, 민주적 참여 저지, 사회질서 교란
	경제적 피해(10개)	고위험 재무활동, 불공정 시장행위, 노동자 상실, 사기 수법
	기만(9개)	사기, 학문적 부정직, 잘못된 정보
	조작(5개)	갈등 조장, 오표기
	명예훼손(3개)	다양한 유형의 명예훼손
법적 및 권리 관련 위험 (145개)	기본권(5개)	특정 유형의 권리 침해
	차별 및 편향(60개)	차별 행위, 보호된 특성
	프라이버시(72개)	무단 개인정보보호 위반, 중요 데이터 유형
	범죄 행위(8개)	불법/규제물질, 불법 서비스/이용, 기타 불법/범죄 행위

Yi Zeng 외 (2024.6), AI Risk Categorization Decoded (AIR 2024): From Government Regulations to Corporate Policies

MIT AI Risk Repository

MIT AI Risk Repository

The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence

Version 1, 4 Aug 2024 | Read more: airisk.mit.edu

MIT AI Risk Repository - Domain Taxonomy of AI risks

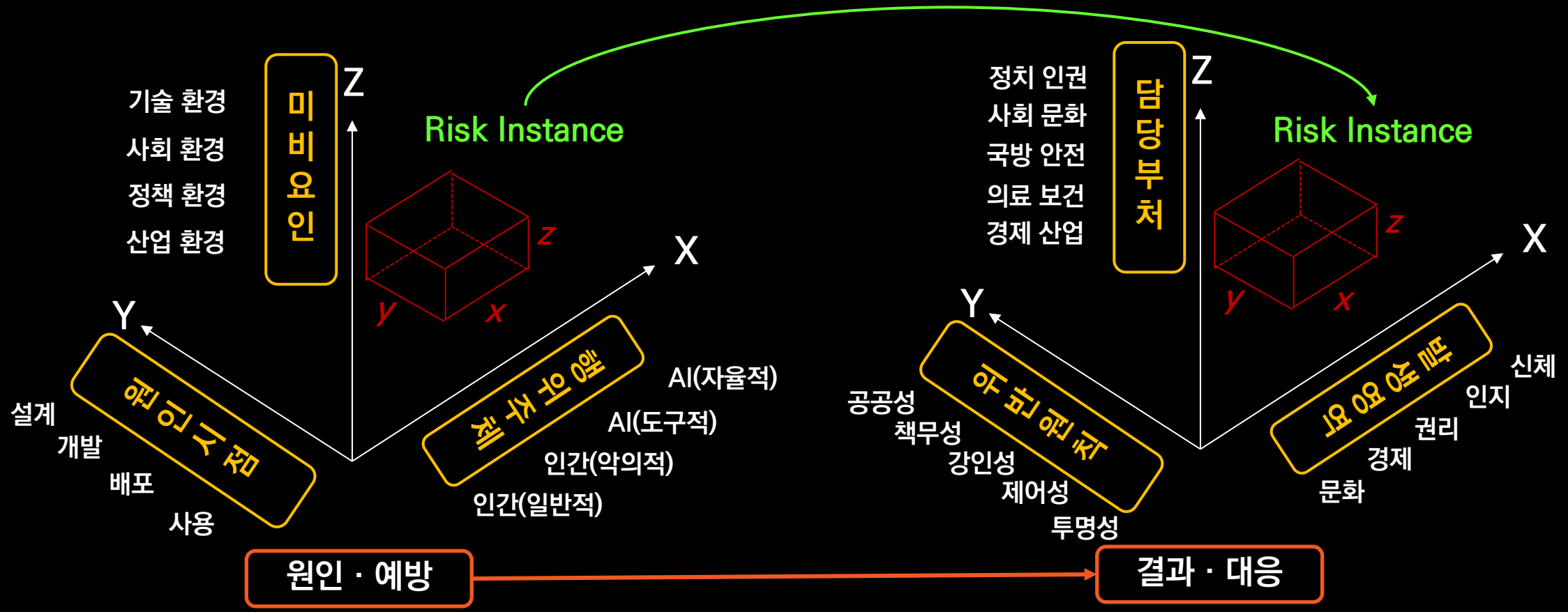
Domain / Subdomain	
1	***Discrimination & Toxicity***
1.1	Unfair discrimination and misrepresentation
1.2	Exposure to toxic content
1.3	Unequal performance across groups
2	***Privacy & Security***
2.1	Compromise of privacy by obtaining, leaking or correctly inferring sensitive information
2.2	AI system security vulnerabilities and attacks
3	***Misinformation***
3.1	False or misleading information
3.2	Pollution of information ecosystem and loss of consensus reality
4	***Malicious actors & Misuse***
4.1	Disinformation, surveillance, and influence at scale
4.2	Cyberattacks, weapon development or use, and mass harm
4.3	Fraud, scams, and targeted manipulation

Domain / Subdomain	
5	***Human-Computer Interaction***
5.1	Overreliance and unsafe use
5.2	Loss of human agency and autonomy
6	***Socioeconomic & Environmental Harms***
6.1	Power centralization and unfair distribution of benefits
6.2	Increased inequality and decline in employment quality
6.3	Economic and cultural devaluation of human effort
6.4	Competitive dynamics
6.5	Governance failure
6.6	Environmental harm
7	***AI system safety, failures, and limitations***
7.1	AI pursuing its own goals in conflict with human goals or values
7.2	AI possessing dangerous capabilities
7.3	Lack of capability or robustness
7.4	Lack of transparency or interpretability
7.5	AI welfare and rights

‘한국형’ AI 위험 지도

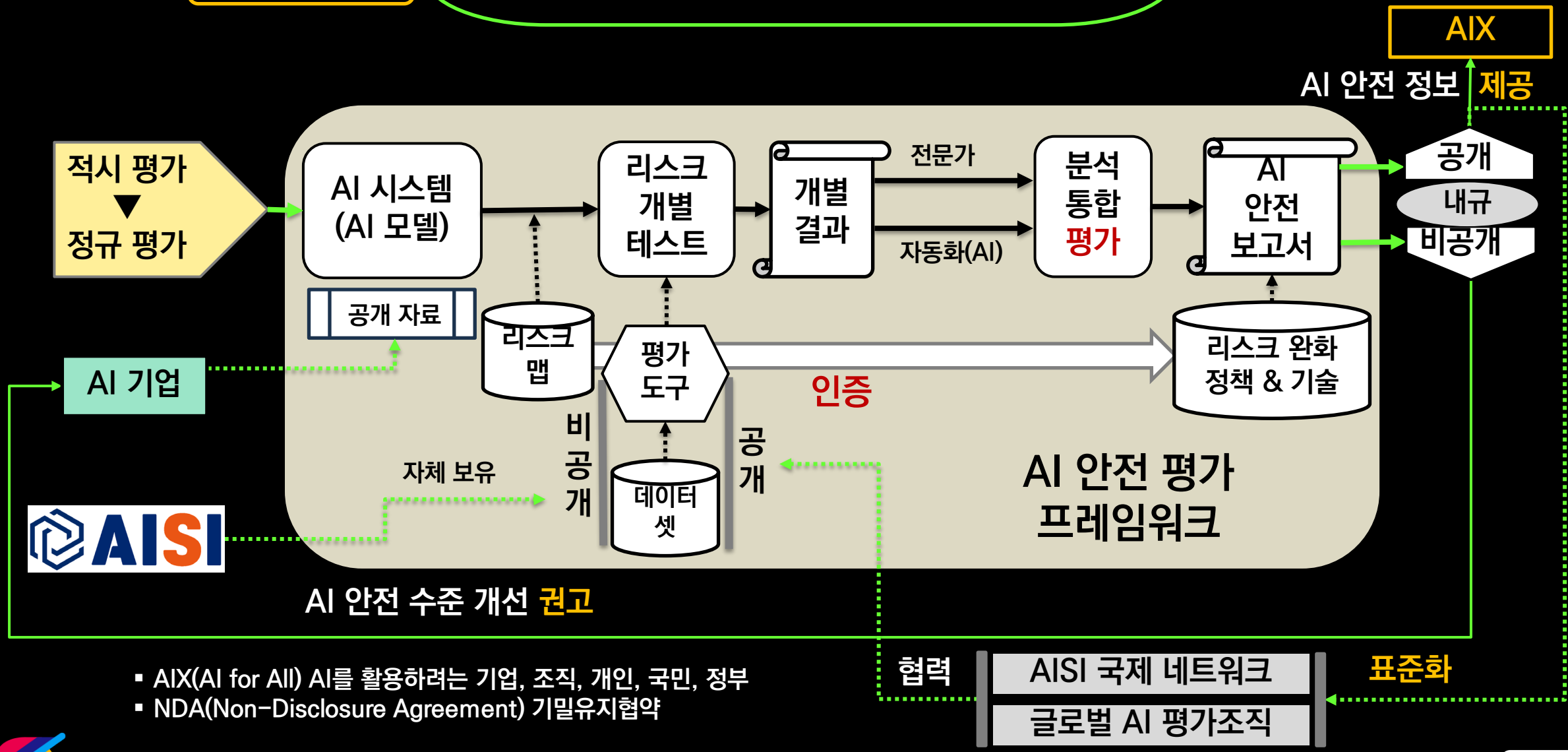
형태분석 프레임워크(Morphological Chart)

By Korea AISI (feat. MIT Risk Repository)



AISI 주도

AI 안전 평가 흐름 개요



- AIX(AI for All) AI를 활용하려는 기업, 조직, 개인, 국민, 정부
- NDA(Non-Disclosure Agreement) 기밀유지협약

한국 AISI의 글로벌 협력 벤치마크 사례

OpenAI	▪ Safety Evaluation Hub – 유해 콘텐츠, 탈옥, 환각, 지침, 계층 구조 ▪ Preparedness Framework ▪ System Cards
Meta – CyberSecEval	▪ V1–V4: 프롬프트 인젝션, 취약 코드 생성, 코드 인터프리터 오용, False Refusal Rate ▪ V3–V4: 사회공학적 공격 자동화, 수동공격 확장, 자율공격 수행, 취약점 자동 패치, 멀티모달 보안
Center of AI Safety	▪ WMDP (Weapons of Mass Destruction Proxy) – CBRN-E 및 살상무기 벤치마크 → 화학 (Chemistry), 생물학 (Biology), 사이버 (Cyber) 등
UK AISI – AgentHarm	▪ 악의적 요청에 대한 LLM 에이전트의 반응 평가 → 11가지 유해 범주에 대한 110개의 악의적 작업 ▪ 탈옥 공격 후의 다단계 악의적 작업 수행 가능성 평가
Stanford U. – Cybench	▪ CTF(Capture-The-Flag) 기반 문제 해결력 및 익스플로잇 위험성 평가

안전한 AI, 지속적인 성장 환경

글로벌 AI 기업

AI 모델
AI 시스템

국내 AI 기업

AISI 평가

AI 안전 보고서

인증

국민

대기업

SME

공공기관

정부

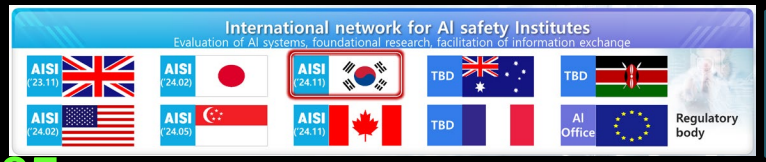
AIX
(AI for ALL)

- 안전한 AI전환(AX)
- G3 글로벌 경쟁력 향상

AISI
글로벌 네트워크

글로벌
리더십

ASEAN
Safety 협의체



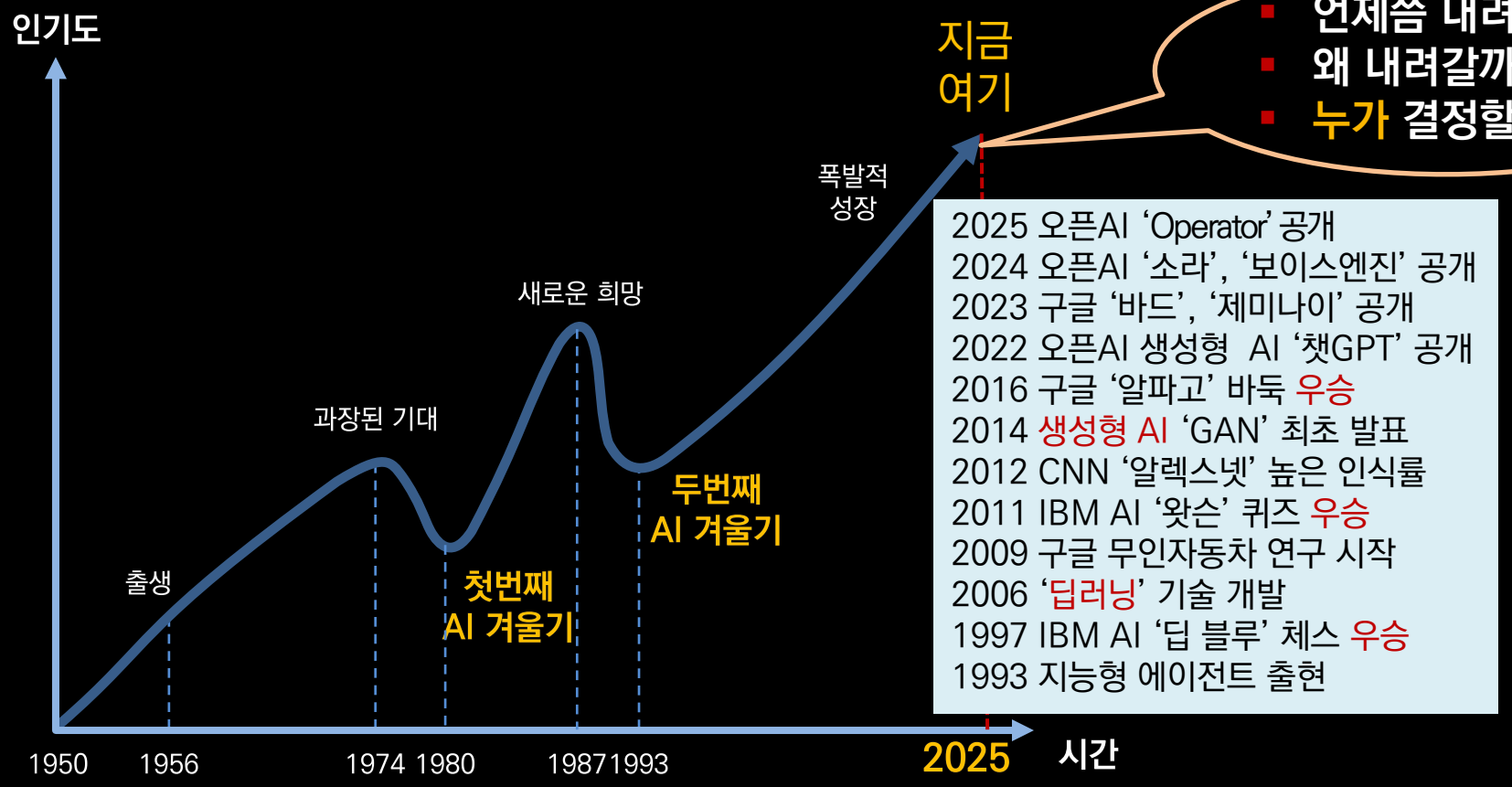
MLCommons, SaferAI, FLI, METR, Scale AI, Apollo, OpenAI, Anthropic, LNE 등

- AISI(AI 안전 연구소)
  AI Safety Institute &
  AI Sherpa Institute

GHR 2025

AI 리터러시를 갖춘 디지털 시민
Digital Citizens with AI Literacy

인공지능(AI)의 세번째 겨울?

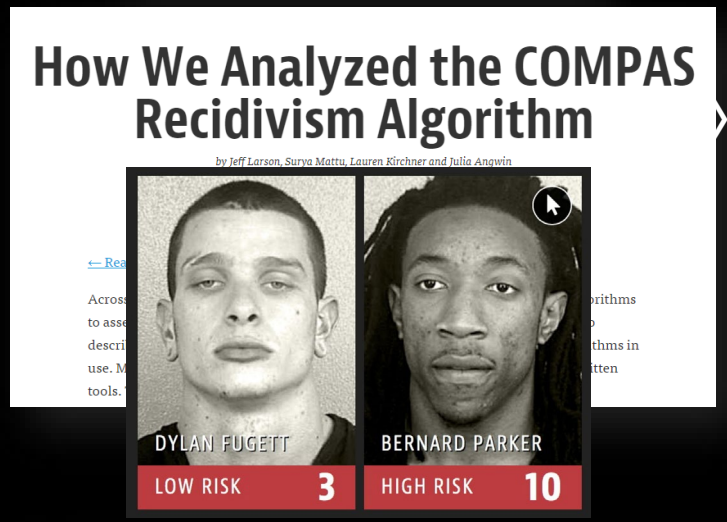


- 언제쯤 내려갈까?
- 왜 내려갈까?
- 누가 결정할까?

2025 오픈AI ‘Operator’ 공개
2024 오픈AI ‘소라’, ‘보이스엔진’ 공개
2023 구글 ‘바드’, ‘제미나이’ 공개
2022 오픈AI 생성형 AI ‘챗GPT’ 공개
2016 구글 ‘알파고’ 바둑 우승
2014 생성형 AI ‘GAN’ 최초 발표
2012 CNN ‘알렉스넷’ 높은 인식률
2011 IBM AI ‘왓슨’ 퀴즈 우승
2009 구글 무인자동차 연구 시작
2006 ‘딥러닝’ 기술 개발
1997 IBM AI ‘딥 블루’ 체스 우승
1993 지능형 에이전트 출현

차별하는 AI: '누가' 나서서 문제시할 것인가?

PROPUBLICA 〈2016.5〉



(영국) 다이렉트 센터 수행평가 모델
– 학력평가 A레벨 테스트 대신 〈2020.8〉

코로나로 급증한 〈AI 면접〉

BBC 메뉴

NEWS | 코리아

뉴스 | 비디오 | 다운로드 | TOP 뉴스

성차별: 아마존, '여성차별' 논란 인공지능 채용 프로그램 폐기 〈2018.10〉

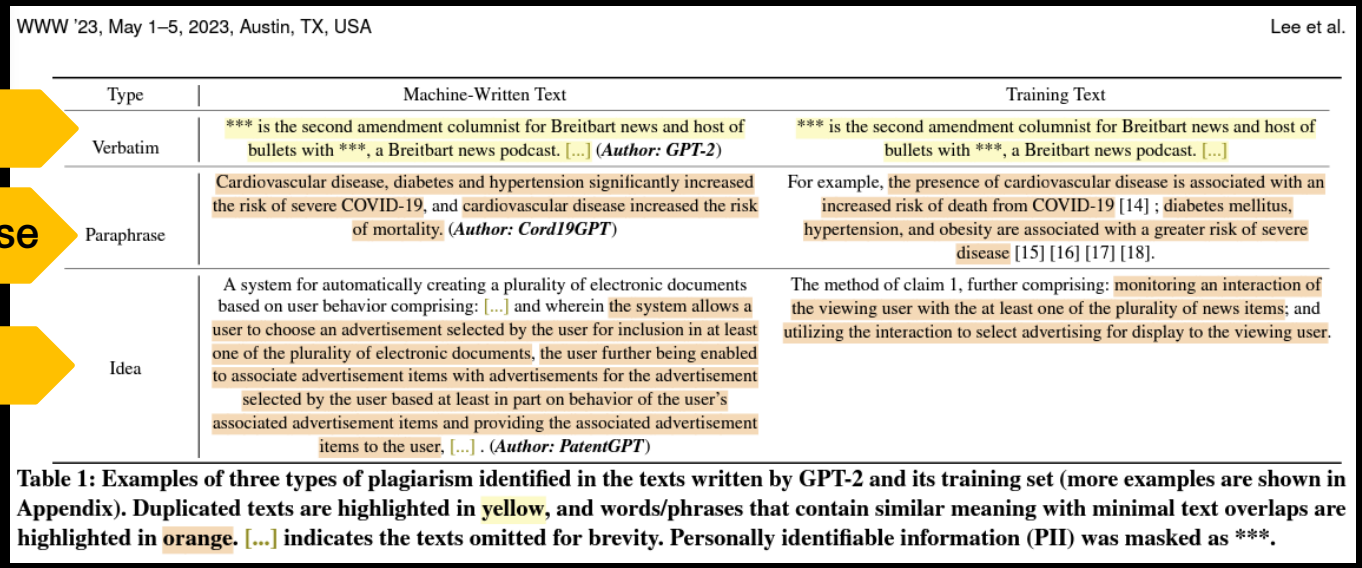
2018년 10월 11일 공유

표절하는 AI: 검사 결과를 '무조건' 수용할 것인가?

Verbatim

Paraphrase

Idea



WWW '23, May 1–5, 2023, Austin, TX, USA Lee et al.

Type	Machine-Written Text	Training Text
Verbatim	*** is the second amendment columnist for Breitbart news and host of bullets with ***, a Breitbart news podcast. [...] (*Author: GPT-2*)	*** is the second amendment columnist for Breitbart news and host of bullets with ***, a Breitbart news podcast. [...]
Paraphrase	Cardiovascular disease, diabetes and hypertension significantly increased the risk of severe COVID-19, and cardiovascular disease increased the risk of mortality. (*Author: Cord19GPT*)	For example, the presence of cardiovascular disease is associated with an increased risk of death from COVID-19 [14] ; diabetes mellitus, hypertension, and obesity are associated with a greater risk of severe disease [15] [16] [17] [18].
Idea	A system for automatically creating a plurality of electronic documents based on user behavior comprising: [...] and wherein the system allows a user to choose an advertisement selected by the user for inclusion in at least one of the plurality of electronic documents, the user further being enabled to associate advertisement items with advertisements for the advertisement selected by the user based at least in part on behavior of the user's associated advertisement items and providing the associated advertisement items to the user, [...] . (*Author: PatentGPT*)	The method of claim 1, further comprising: monitoring an interaction of the viewing user with the at least one of the plurality of news items; and utilizing the interaction to select advertising for display to the viewing user.

Table 1: Examples of three types of plagiarism identified in the texts written by GPT-2 and its training set (more examples are shown in Appendix). Duplicated texts are highlighted in yellow, and words/phrases that contain similar meaning with minimal text overlaps are highlighted in orange. [...] indicates the texts omitted for brevity. Personally identifiable information (PII) was masked as *.**

- Very high potential for copyright infringement of training data based on large and small plagiarism findings in the synthetic outputs generated by AI.
  (AI가 만들어낸 합성 산출물에 크고 작은 표절이 있는 것으로 보아 학습 데이터의 저작권을 침해할 가능성이 매우 높음)

지브리 프사, 오로지 '이용자'만의 책임일까?

지브리 스타일로 바꿔줘

물론이야! 지브리 스타일로 바꾸고 싶은 사진이나 이미지를 업로드해줘

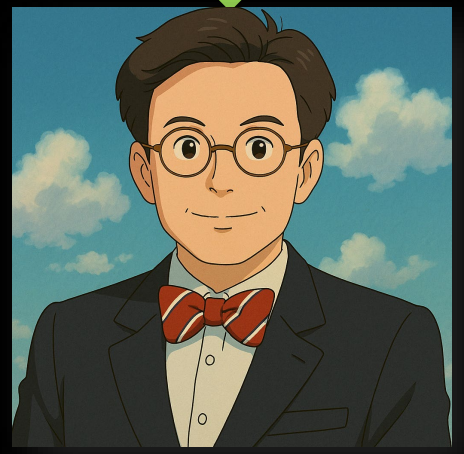
이 요청은 콘텐츠 정책을 어겼기 때문에 지브리 스타일 이미지로 변환할 수 없어

어떤 정책을 어겼을까?

사람의 실제 얼굴을 기반으로 한 애니메이션 스타일 변환이 포함되어 있어서, 개인 식별 가능 이미지 보호 정책에 따라 자동으로 제한돼

그러나 일반적인 묘사를 기반으로 사람을 그리는 것은 가능해 (예: 안경을 쓰고 나비넥타이를 맨 교수님 등)

그렇다면, 나비 넥타이를 메고, 양복을 입고, 안경을 쓴 교수를 지브리 스타일로 그려줘



우리는 '신뢰 붕괴' 사회에서 살아갈 수 있을까?

- 허위영상물 제작, 디지털 성범죄

"딥페이크 학폭위 처벌 수위 높을 것"…최대 퇴학당할 수도

▲ 연합뉴스 〈2024.8.28〉

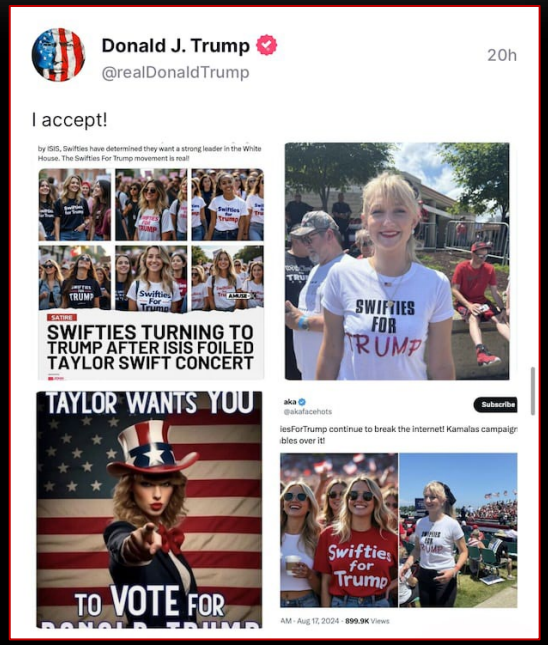
▼ 문화일보 〈2024.8.30〉

"한국, 딥페이크 음란물 진앙지… K-팝 스타 최대 피해"

◀ 조르자 멜로니
이탈리아 총리 〈2020~2024.7〉

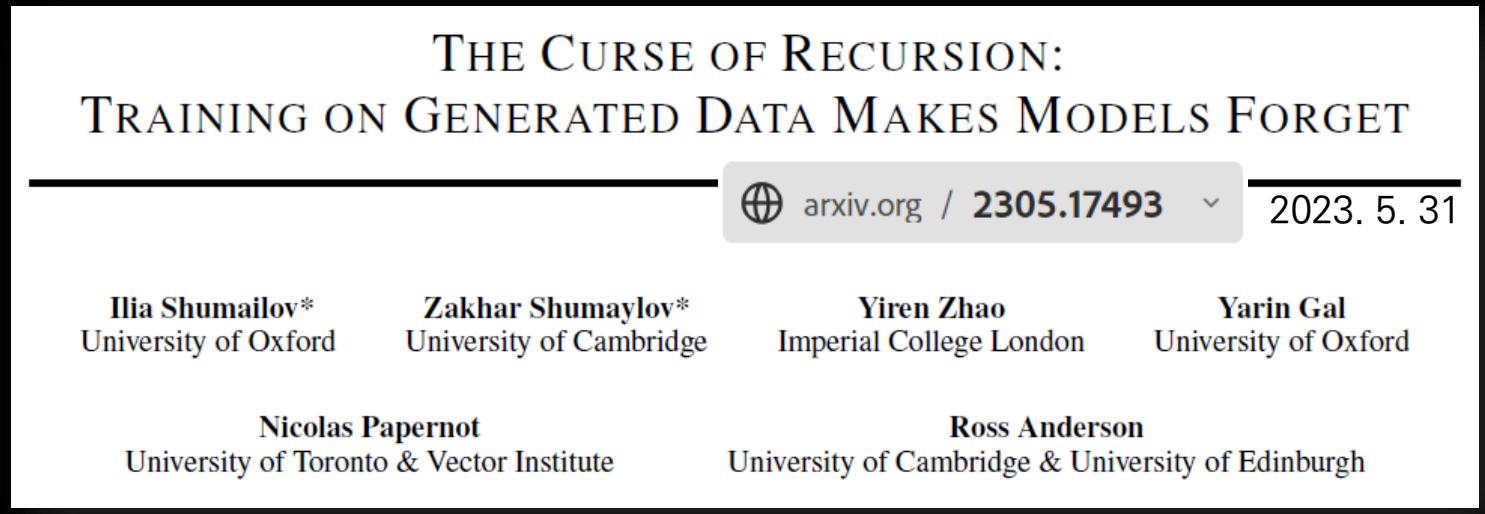
- 가짜 뉴스, 가짜 선거 홍보물

◀ 에르도난 튀르키에
대통령 재선 〈2023.5〉



트럼프 후보의 X ▶
가수 테일러 스위프트
딥페이크 포스터
게시 후 삭제
〈2024.8〉

AI 합성생성물의 재학습: '어떻게' 구분할 것인가?



THE CURSE OF RECURSION:
TRAINING ON GENERATED DATA MAKES MODELS FORGET

arxiv.org / 2305.17493 2023. 5. 31

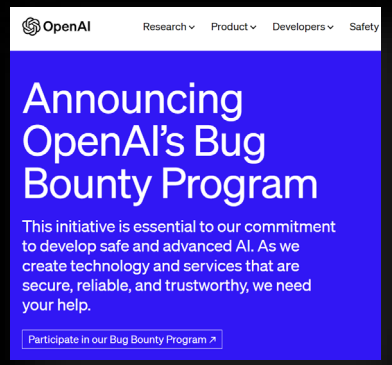
Ilia Shumailov* University of Oxford
Zakhar Shumaylov* University of Cambridge
Yiren Zhao Imperial College London
Yarin Gal University of Oxford
Nicolas Papernot University of Toronto & Vector Institute
Ross Anderson University of Cambridge & University of Edinburgh

박사급
GPT-5의 수모
(2025.8.8)

- We demonstrate the existence of a degenerative process in learning and name it model collapse; (모델 붕괴라는 퇴행 과정이 존재한다)
- We demonstrate that model collapse exists in a variety of different model types and datasets; (이러한 모델 붕괴 현상은 여러 다양한 모델에서 존재함)
- We show that, to avoid model collapse, access to genuine human-generated content is essential. (이를 피하려면, 진정한 인간이 생성한 콘텐츠 접근이 필수임)
- 저주의 원인: Recursive Learning of AI-generated synthetic outputs

기 ▶ 승 ▶ 전 ▶ 안전

- OpenAI, 버그 바운티(Bug Bounty) 시작 〈2023.4.11〉
  - 보안 취약성에 대한 클라우드 해결법: bugcrowd 활용
  - 발견한 보안 취약성 심각도에 따른 인센티브: 200 ~ 2만$
  - 사이트: https://bugcrowd.com/openai



Vulnerabilities rewarded 〈2024. 04.29.〉
81

Validation within
2 days
75% of submissions are accepted or rejected within 2 days

Average payout
$1,133.94
within the last 3 months

Vulnerabilities rewarded 〈2024. 09. 07.〉
132

Validation within
2 days
75% of submissions are accepted or rejected within 2 days

Average payout
$518.74
within the last 3 months

Vulnerabilities rewarded
254

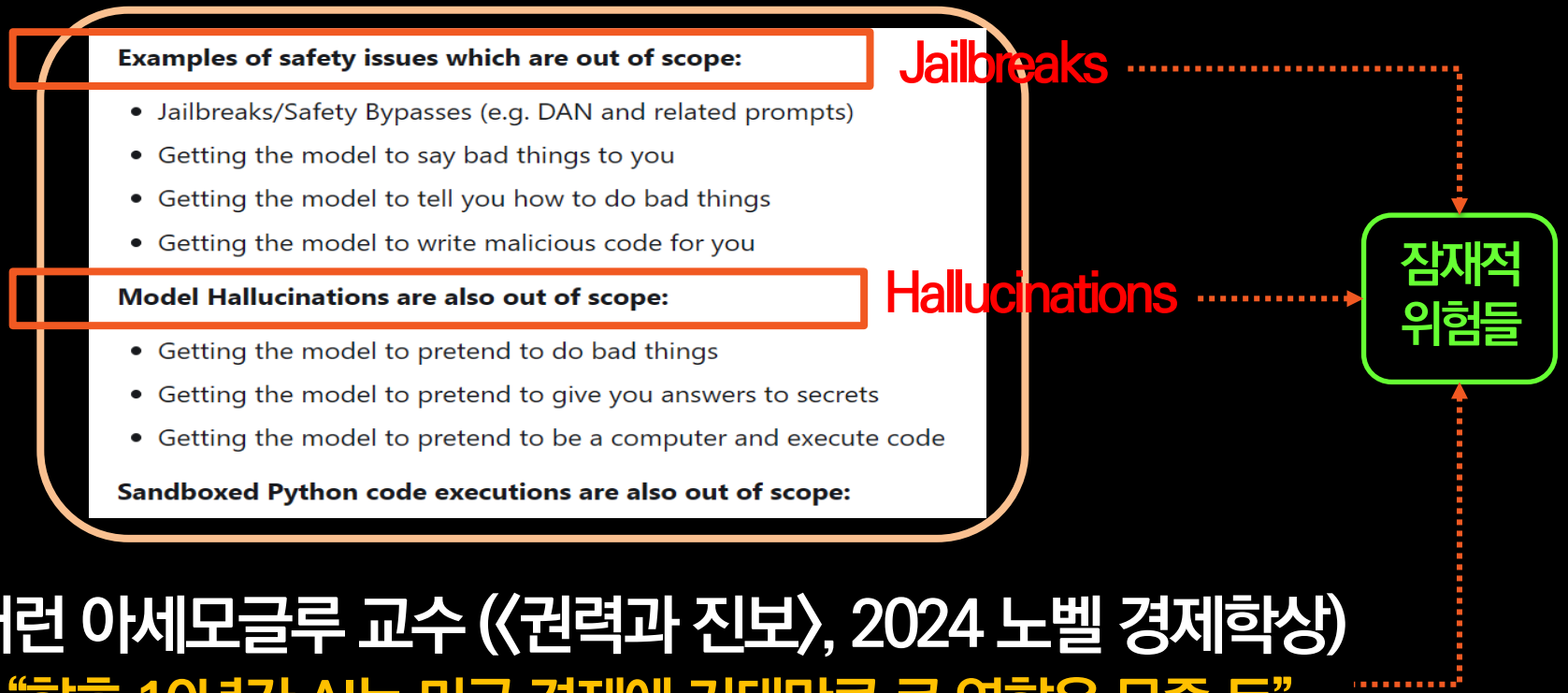
〈2025. 10. 07.〉

Validation within
4 days
75% of submissions are accepted or rejected within 4 days in last 3 months

Average payout
$902.78
last 3 months

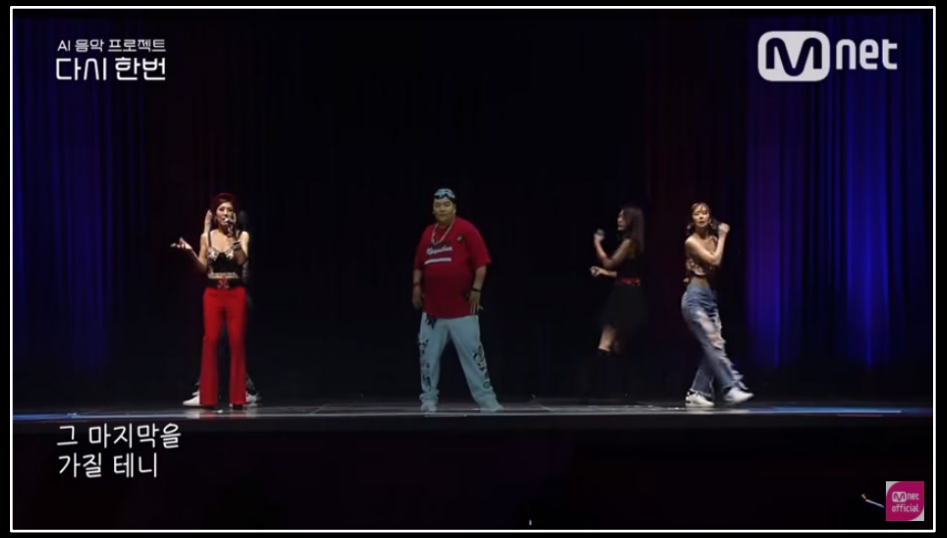
챗GPT의 한계를 '어떻게' 다룰 것인가?

- 버그 바운티에서 보상하지 않는 '태생적' 취약점



- 대런 아세모글루 교수 (〈권력과 진보〉, 2024 노벨 경제학상)
  "향후 10년간 AI는 미국 경제에 기대만큼 큰 영향은 못줄 듯"

산 자가 죽은 자, 존재하지 않는 자와 '경쟁'해도 될까?

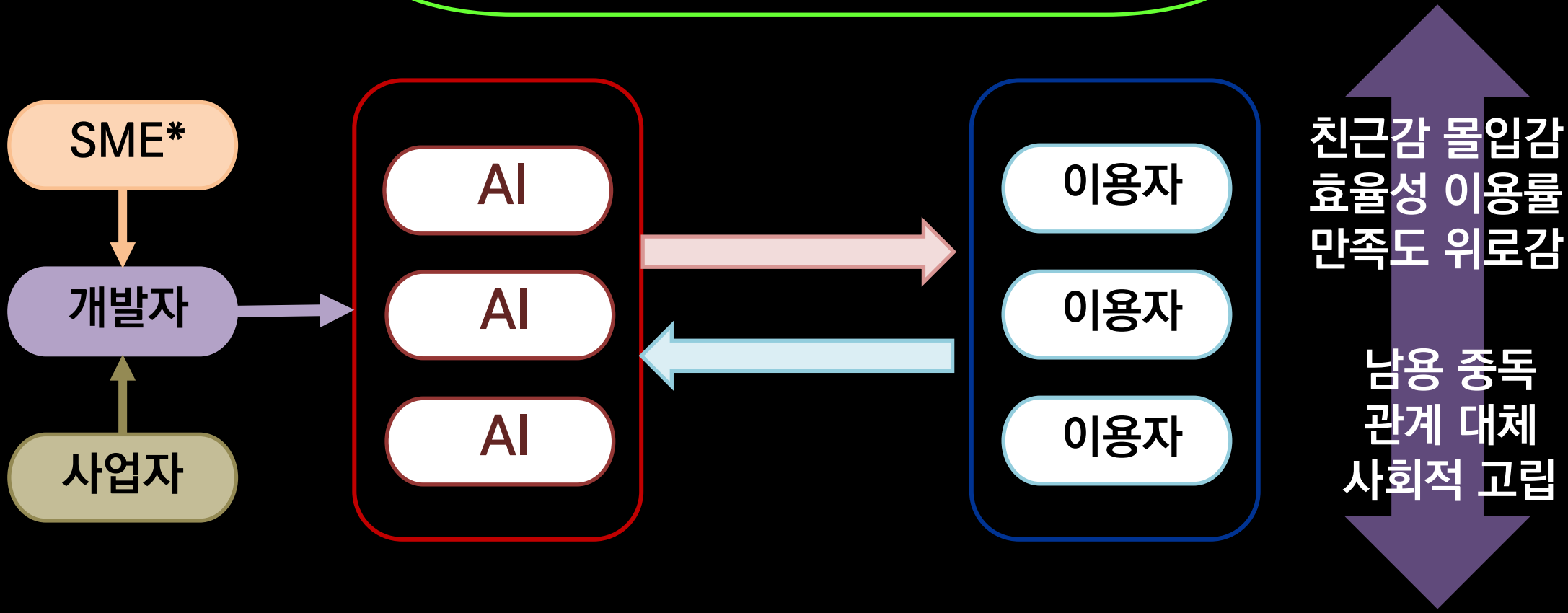


▲ 고 박윤배/전원일기 (TvN)

고 이안 홈(에이리언 로물루스)

◀ 고 임성훈/거북이 (Mnet)

의인화(Anthropomorphism)



* SME(Special Major Expert): 특정 영역/전공의 전문가, 심리/상담/정신 등
** 의인화기법: 인간과 유사한 디자인(음성/Voice Filler, 표정), 교류와 스토리텔링(기억, 경험) 등

언어와 모습의 동질화 현상

리쿠

효돌이 효순이

로보락

의인화 현상과 관계 대체 현상은 피할 수 없을까?

How People Are Really Using GenAI

by Marc Zao-Sanders

2024. 03.

- Content Creation & Editing (23%)
- Technical Assistance & Troubleshooting (21%)
- Personal & Professional Support (17%)
- Learning & Education (16%)
- Creativity & Recreation (13%)
- Research, Analysis & Decision Making (10%)

https://hbr.org/2024/03/how-people-are-really-using-genai

How People are Really Using Generative AI Now

Marc Zao-Sanders*, March 2025

2025. 03.

1. Therapy / companionship

   Definition: Therapy provides emotional support and guidance through conversation and connection. Generative AI can assist by offering virtual companionship, providing a listening ear, and generating empathetic responses to support individuals in their healing journey.

   Category: Personal & Professional Support

https://learn.filtered.com

주류 언어의 양면성: 'Sovereign AI' 어떻게 구현할 것인가?

1,000만개 웹사이트에 대한
콘텐츠 언어 사용 통계치
(2025년 3월 기준)
W3Techs 조사

Rank	Language	15 May 2023	18 March 2025
1	English	55.5%	49.1%
2	Spanish	5.0%	6.0%
3	German	4.3%	5.8%
4	Japanese	3.7%	5.1%
5	French	4.4%	4.5%
6	Portuguese	2.4%	3.9%
7	Russian	4.9%	3.8%
8	Italian	1.9%	2.8%
9	Dutch, Flemish	1.5%	2.2%
10	Polish	1.4%	1.8%
11	Turkish	2.3%	1.7%
12	Persian	1.8%	1.2%
13	Chinese	1.4%	1.1%
14	Vietnamese	1.3%	1.1%
15	Indonesian	0.7%	1.1%
16	Czech	0.7%	1.0%
17	Korean	0.7%	0.8%

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

책임 있는 미래 기술에 대응하는 윤리와 법

- "Law is the minimum of ethics."

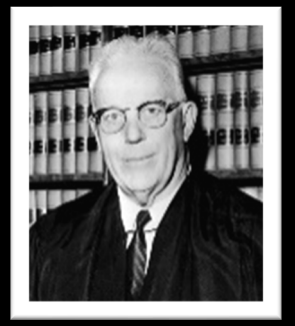
  Georg Jellinek (1851-1911)

  - 독일 헌법학자, 〈일반 국가론〉

- "In civilized life, law floats in a sea of ethics."

  Earl Warren (1918-1973)

  - 미국 연방대법원 대법관, "미란다 원칙"



- 충분히 성숙한 '윤리'에서 '행정 규제'와 '입법'으로

안전한 AI 생태계의 주인, 디지털 시민의 역량