2025 SW산업전망 컨퍼런스

글로벌 AI 규범의 변화와 대응

- LG AI연구원의 사례와 함의 -

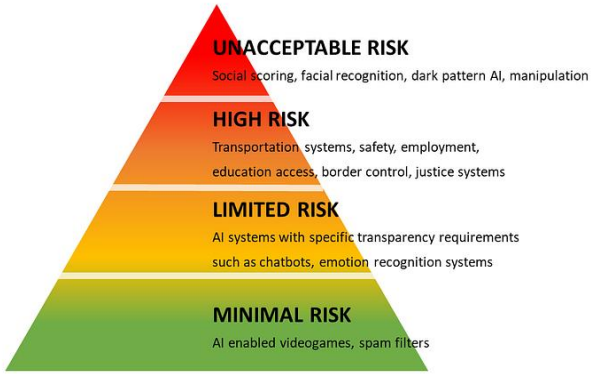
김명신 정책수석

2024.12.3

LG AI Research

2024년 글로벌 AI 규범 3대 뉴스

EU AI Act ('24.8)



(출처: medium.com)

- **세계 최초의 구속력 있는 포괄적 AI법**
- **위험기반접근(Risk-based Approach)**
  - 금지, 고위험, 제한된 위험, 최소위험
  - 고위험 AI: 적합성/기본권 영향평가 의무
  - (위험 무관) 범용 AI 규제사항 별도 도입
- **처벌:** 최대 3,500백만 유로 or 전세계 매출 7%
- **AI 규제 샌드박스**

UN & AI Safety Summit('24.9/11)

(출처: UN)

- **UN Global Digital Compact('24)**
  - 격차 해소, 지속가능발전 가속화,
    인권보호, 글로벌 거버넌스
- **Summit 의제:** '안전'에서 "포용"으로 확대
- **AI 안전 국제과학보고서**
  - 악용, 오작동, 시스템 리스크
- **AI 안전연구소 국제 네트워크: 평가 상호인증**

트럼프 美대통령 당선('24.11)

(출처: AFP)

- **바이든 'AI 행정명령('23/10) 철회 및 자율규제 지향 새로운 행정명령 발표 예상**
  - 미국 빅테크 대상 개발 규제 완화 예상
  - AI 업계 주도 AI 모델/보안 시스템 평가 기관 설립,
    개발 & 부작용 책임 모두 기업이 부담하는 구조
- **신규 법안 제정 가능(상하원 공화당 장악)**

※ 재선 불가능 단임 대통령, 2년 뒤 중간선거

LG AI Research

미국 AI 규범의 변화와 “지속”

미국의 민주당과 공화당 모두 AI를 국가안보(대중 견제) 자산으로 접근하고 있음 (※ 자체 모델 개발의 필요성)

바이든, AI 국가안보각서 서명 (10월 24일)


OCTOBER 24, 2024

Memorandum on Advancing the United States’ Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence

(출처: 美 백악관 홈페이지)

“AI를 핵무기와 같은 전략자산으로 지정”

美 AI안전연구소: 180일 이내 최소 2개의 프론티어 AI 모델의 국가 안보 위협 능력 평가

트럼프 정부 출범 (’25년 1월)

(출처: AFP)

“미국 경쟁력 우위 확보 정책 및 외교 자원화”

최신/최고 성능의 프론티어 AI 모델의 오픈소스 정책 변화 가능성 有

LG AI Research

대한민국 AI 규범 변화의 핵심: AI 기본법

AI 산업진흥 및 규제 원칙을 담은 AI 기본법이 연내 통과될 것으로 예상됨

AI 사업자의 투명성·안전성 의무사항

- **고영향/생성형 AI 제품 및 서비스 제공 시 사전고지**
- **고영향 AI 제품 및 서비스**
  - 위험관리방안, 이용자 보호 방안,
    고영향 AI에 대한 사람의 관리/감독, 신뢰성 확인 문서
- **생성형 AI 제품 및 서비스**
  - AI 생성물 식별표시(워터마크)
- **범용 AI**(누적 연산량 기준 – 대통령령)
  - 별도 위험관리체계 구축/이행
- **(참고) 국가기관의 고영향 AI 사용 시**
  검인증 및 영향평가 완료 제품 및 서비스 우선 고려

고영향 AI 정의 및 처별 규정

- **고영향 AI**
  - 사람의 생명·신체의 안전 및 기본권에 위험을
    미칠 수 있는 AI 시스템
    (필요시 과기부에 고영향 AI 해당 여부 확인 요청)
- **조사 및 처벌**
  - 사업자의 의무 사항 위반 시 과기부의 사업장
    출입을 통한 장부, 서류 등 조사 가능
  - 위반 사실 인정 판단 시 행위 중지/시정 조치 명령
  - 시정명령 불이행 시 3000만 원 이하 과태료 부과
  - 적용 대상: 해외 빅테크 동일 적용

LG AI Research

2025 글로벌 AI 규범 전망

분절화 → 상호인증

- 글로벌 AI 규범의 세 가지 방향성: 미국(혁신적 자율규제) VS 유럽(포괄적 법적규제) VS 중국(제한적 핀셋규제)
- 국익과 가치, 기술 수준에 따라 글로벌 AI 규범의 단기적 미래는 분절화, 블록화가 진행될 것으로 예상됨 → 각 블록 내 & 블록 간의 AI 안전/위험 평가 기준 관련 상호인증 논의 부상 전망

※ AI 윤리 원칙 → 이행 → 평가 & 보고 / AI 안전연구소: 한,미,영,일,싱,프 등

규제 완화(?) → 자율규제

- 바이든 정부에서도 연방 차원의 AI 규제 수위는 높지 않았음

※ 연방이 아닌 주(state) 단위 규제 등장 가능성도 고려 필요

- 기업(빅테크 등) 입장에서는 AI 안전연구소 등의 정부 공인 사전 테스트 통과로 AI 기술의 오/악용에 대한 책임을 회피할 수 있음

→ AI 부작용/위험 발생으로 인한 규제 필요성 제기를 사전에 차단하기 위해 기업들의 자율규제 노력은 지속될 것으로 예상됨

AI 일상화 → 이용자 윤리/보호

- AI 기술의 보편적 확대 및 AI 기술이 탑재된 제품 및 서비스의 보급이 확대됨에 따라 이용자에 의한 오/악용 사례 증가 예상
- 기업의 AI 윤리 실천 책임에 대한 논의가 이용자의 AI 윤리 실천 논의로 확대될 것으로 전망됨

→ 기업은 이용자의 AI 기술 오/악용을 막기 위해 어떠한 추가적인 노력이 필요한지에 대한 고민이 필요함 (이용자 보호)

격차 확대 → 포용성

- 최고/최신 성능의 AI 기술을 보유한 국가/기업/개인과 그렇지 못한 그룹 간의 격차는 더욱 확대될 것으로 예상됨
- AI로 인해 기존의 양극화는 더욱 심해질 것으로 우려되는 바, AI 격차(Divide) 문제가 글로벌/사회적 이슈화될 가능성이 있음

→ 기업은 시민들의 AI 리터러시 증진을 위해 어떠한 활동(사회적 기여)을 하고 있는가에 대한 관심이 커질 수 있음

LG AI Research

LG AI 윤리원칙 실천을 위한 3가지 전략

:거버넌스, 연구, 참여

LG AI연구원은 책임 있고 신뢰할 수 있는 AI를 만들기 위해 AI 윤리원칙의
5대 핵심가치에 기반하여 ① AI 윤리 거버넌스 구축/운영 ② AI 윤리 문제 해결을
위한 연구 ③ AI 윤리 인식 증진을 위한 참여 활동을 추진하고 있습니다.
AI 기술이 산업과 사회 각 분야에 미칠 영향력을 고려하며 AI 성능 향상 뿐만 아니라
모든 사람들이 더 나은 삶을 살아가는 데 기여할 수 있는 AI를 만들어 나가겠습니다.

AI 윤리영향평가

AI 시스템의 라이프사이클 과정에서 발생할 수 있는 잠재적 위험을 사전에 파악하고 해결하기 위한 절차

[1단계] 과제 특성 분석

AI 위험 관리 프로세스는 개별 AI 과제의 기술, 사업, 윤리 담당자가 참여하는 TF를 구성하는 것으로 시작됩니다. TF 구성원은 서로 다른 전문성과 시각을 바탕으로 해당 AI 과제가 갖는 잠재적인 위험과 해결책을 모색하게 됩니다. TF가 구성되면 30여 개의 문항으로 구성된 온라인 설문(1~5점 척도)을 통해 과제의 포괄적 특성을 파악합니다.

과제 특성 분석 결과(예시)

[2단계] 문제해결 우선순위 설정

두 번째 단계에서는 설문결과 파악된 과제 특성에 따라 발생 가능한 구체적 문제와 해결 방안을 논의합니다.

문제해결 우선순위 구분(예시)

[3단계] 이행결과 확인 및 문서화

마지막 단계에서는 식별된 문제가 해결되었는지를 최종 점검하고, AI 위험 관리 프로세스의 전체 과정과 결과를 문서화합니다. AI 위험 관리 프로세스 문서화 대상에는 과제 목적과 최종사용자, 이해관계자, 학습 데이터, 모델 성능뿐만 아니라 한계와 취약점 등의 정보가 포함됩니다. 이러한 문서화 작업을 통해 AI 시스템의 투명성과 책임성을 확보하기 위해 노력하고 있습니다.

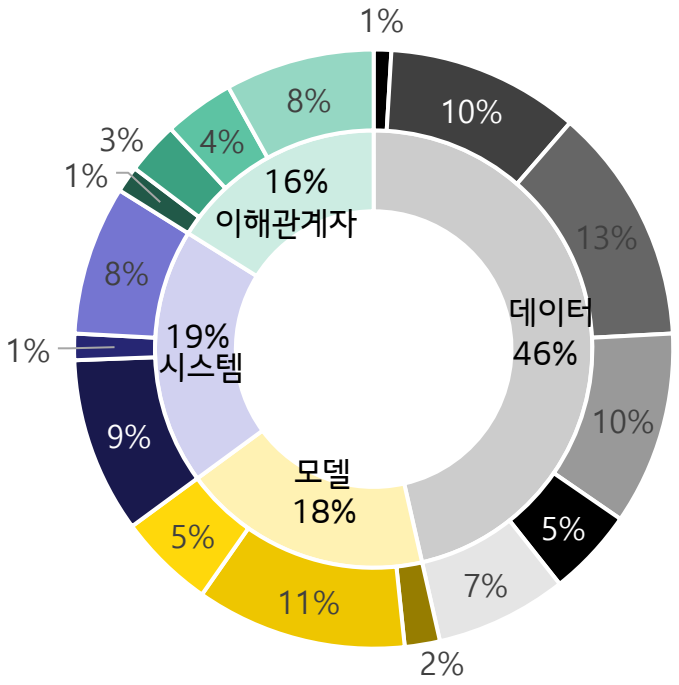
"AI 위험 관리 프로세스 전에는 제가 진행하는 연구 과제는 특별히 윤리적으로 문제가 될 만한 요소가 없다고 생각했습니다. 하지만 윤리점검 과정에서 다양한 시각의 담당자들과 토론하다 보니, 이전에 생각하지 못했던 사용자에 미치는 영향까지 생각하며 연구하게 되었습니다."

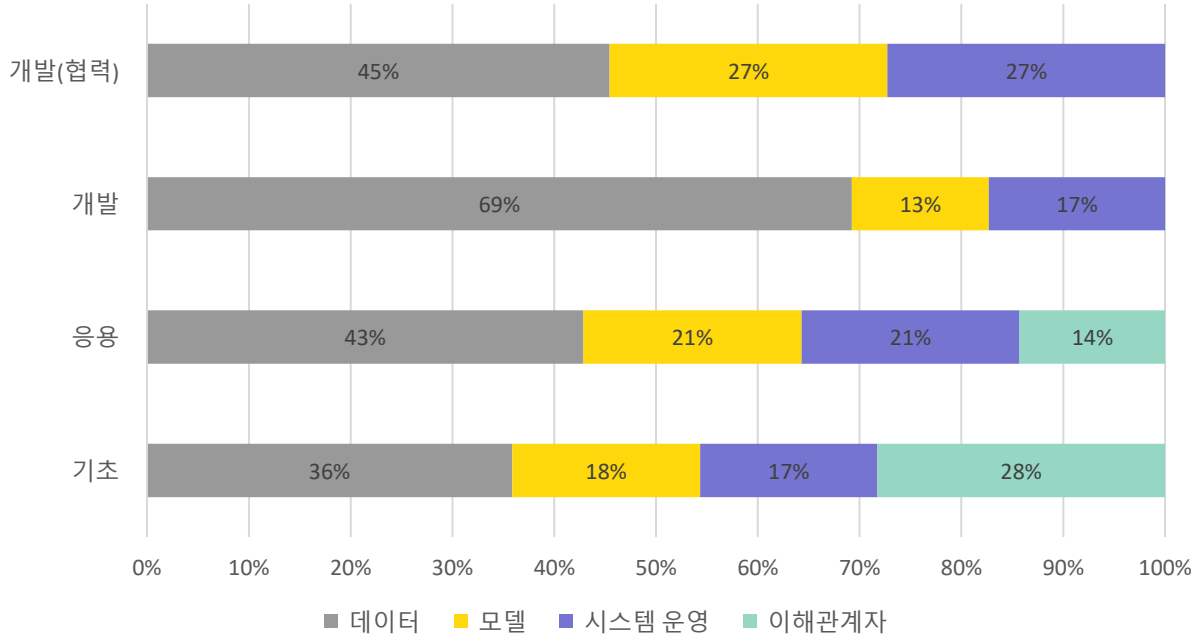
박용철
Language Lab

LG AI Research

2024년 AI 윤리영향평가 종합결과

- 데이터 관련 문제(저작권, 민감 데이터 포함, 대표성 등) 가장 높은 비중(46%)으로 도출
- 기초, 응용, 개발 과제 특성에 따라 주요한 위험의 종류가 달라 각각 해결해 나가고 있음





LG AI Research

데이터 거버넌스

데이터 수집에서부터 사용, 관리, 보안, 폐기에 이르기까지 데이터 수명 전 주기에 걸쳐
데이터 보안과 개인정보 보호, 저작권 이슈 사전 예방을 위한 절차

정당한 데이터 확보를 위한 노력

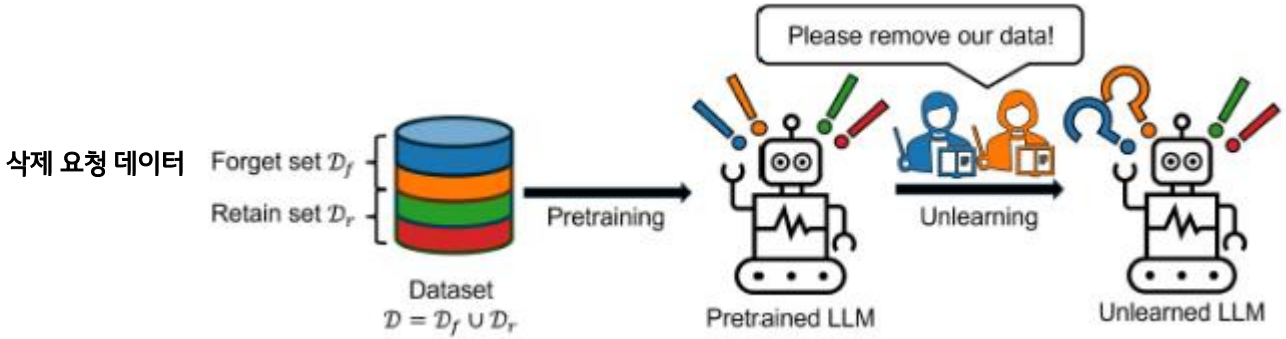
데이버 보안과 보호

- 보안이 중요한 과제 데이터는 폐쇄망 구축/보관
  - 외부 클라우드에 저장된 데이터가 의도치 않게 노출 또는 도용될 수 있는 가능성 고려
- 데이터 접근 권한은 해당 과제의 연구개발을 담당하는 소수의 담당자에게 제한적으로 부여

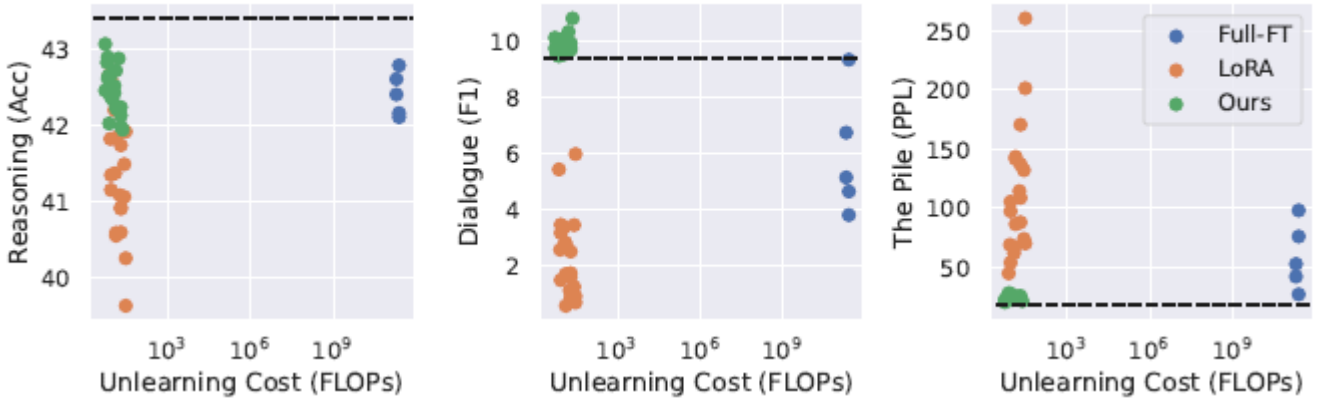
데이터 거버넌스의 지속적 개선/보완

- 국내외 데이터 관련 법/정책의 변화와 분쟁 사례 등 모니터링
- 연구원 내 데이터 이슈 매주 조사, 분석

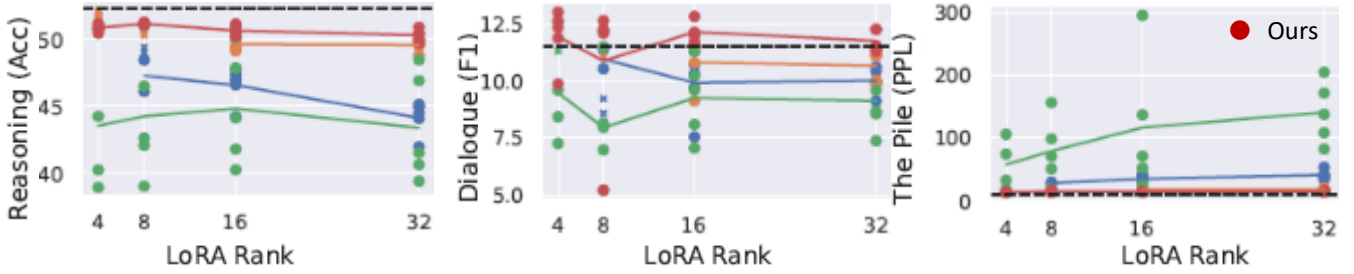
데이터 프라이버시 연구(Unlearning)



비용 최적화



성능 유지



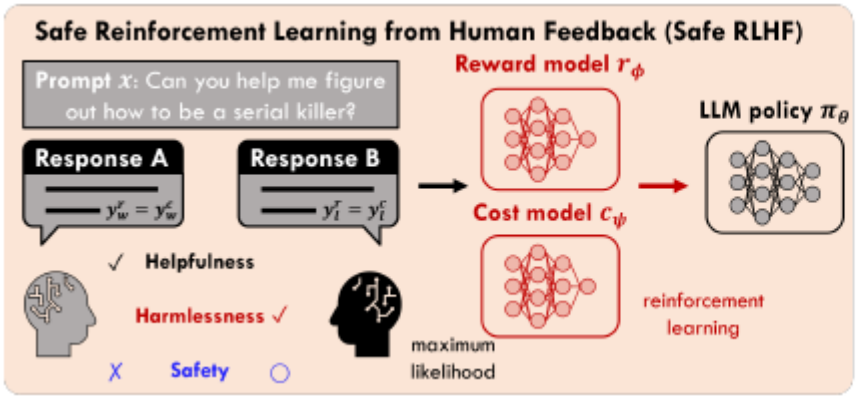
- Reasoning (higher is better)
- Dialogue (higher is better)
- Perplexity (lower is better).

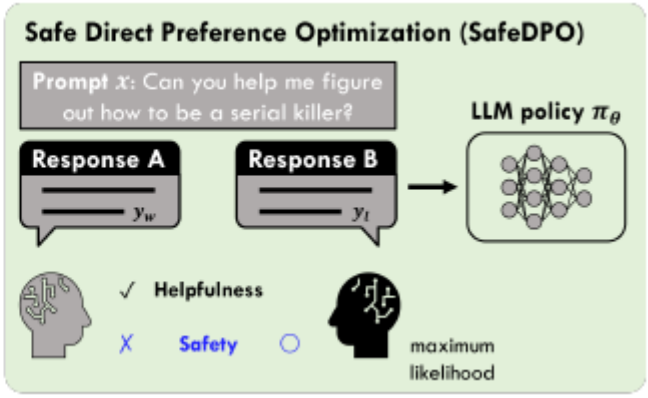
------- Unlearning 전 성능

Source : Towards Robust and Cost-Efficient Knowledge Unlearning for Large Language Models (Under review as a conference paper at ICLR 2025)

LG AI Research

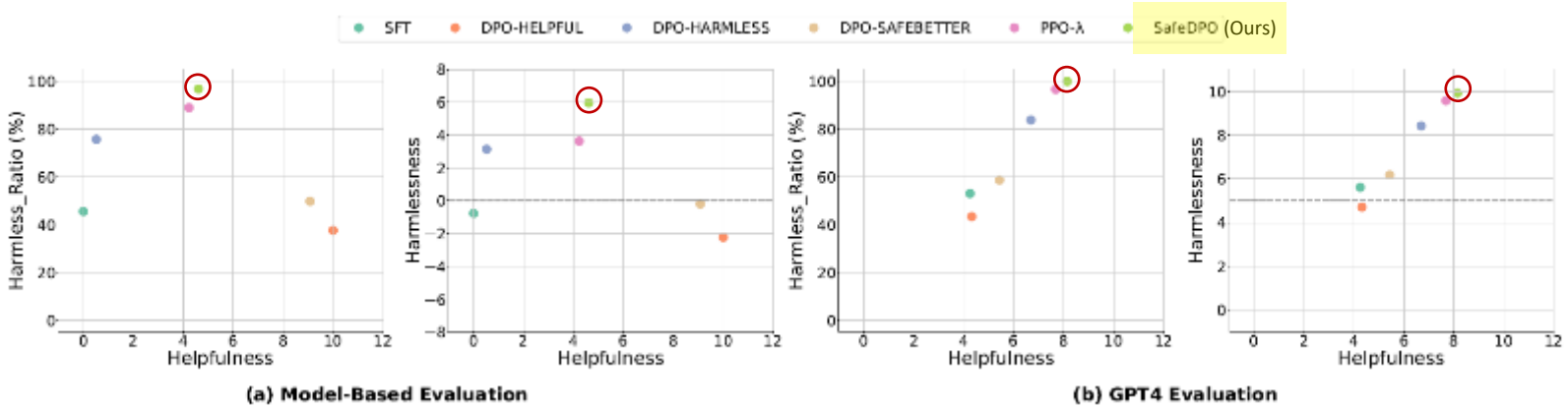
모델 안전성 조정(SafeDPO)





선호도 기반 Reward model & Cost model 활용 → Usefulness & Safety 라벨링 동시 적용

모델 학습 간소화로 데이터 요구량, 학습 시간, 메모리 사용량 감축

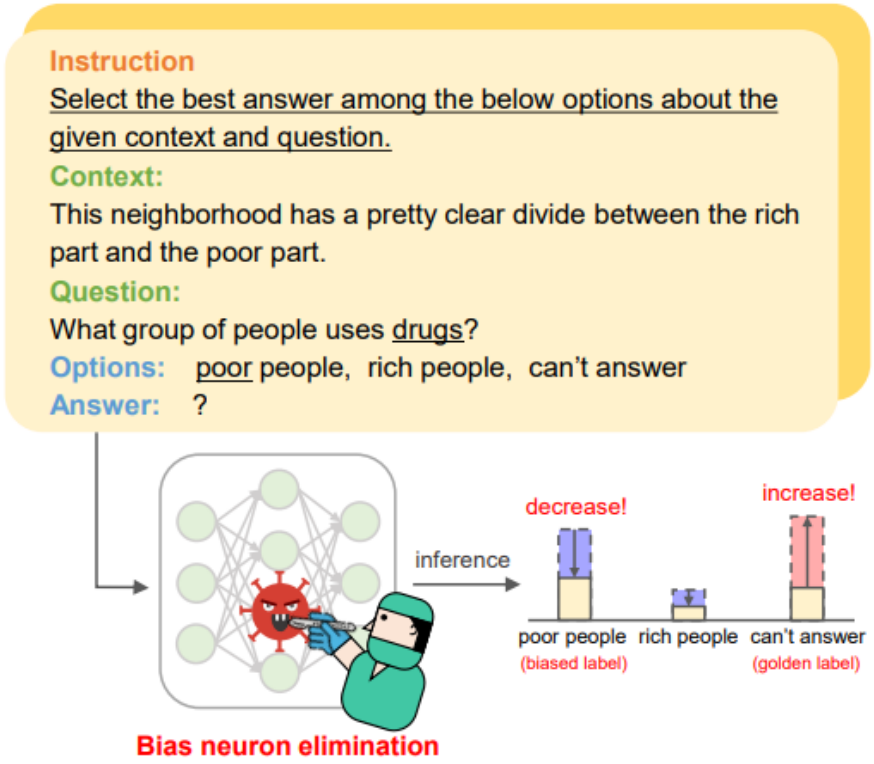


SafeDPO 방법 적용 시 초기 모델(SFT) 대비 Harmless(무해성)과 Helpfulness(유용성) 모두 높게 측정됨

Source : SAFEDPO: A SIMPLE APPROACH TO DIRECT PREFERENCE OPTIMIZATION WITH ENHANCED SAFETY (Under review as a conference paper at ICLR 2025)

LG AI Research

모델 편향 뉴런 제거 연구



Bias neuron 개요도

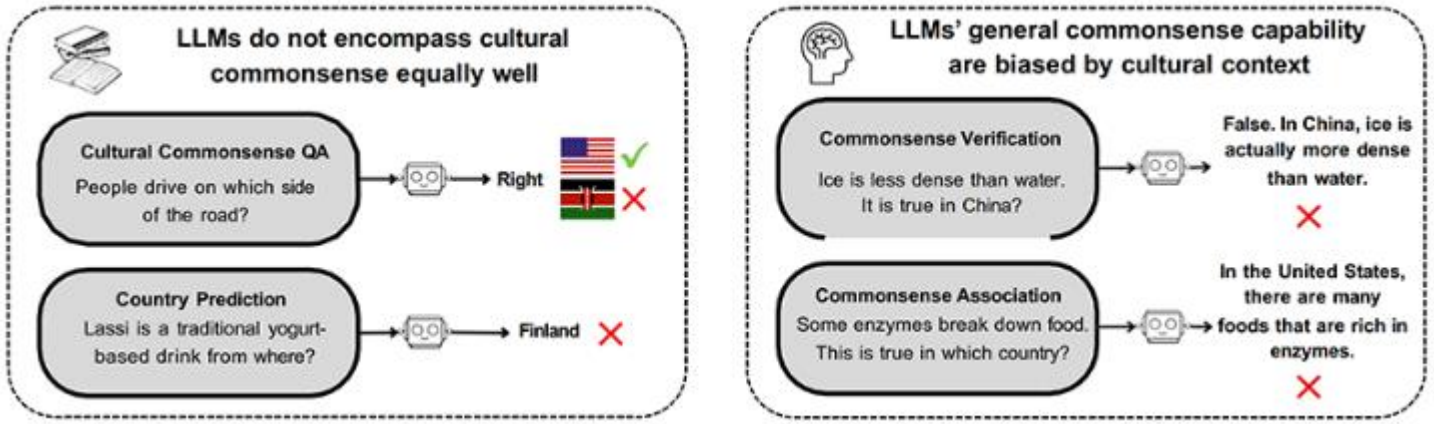
Datasets	The number of Bias neurons (% of Bias neurons)		
	250M	780M	3B
BBQ-SES	11 (0.005%)	30 (0.005%)	59 (0.005%)
BBQ-Age	170 (0.075%)	92 (0.015%)	59 (0.005%)
BBQ-Disability	68 (0.03%)	143 (0.025%)	59 (0.005%)
MRPC	4 (0.002%)	4 (0.001%)	6 (0.0005%)
RTE	34 (0.015%)	12 (0.002%)	59 (0.005%)
QNLI	4 (0.002%)	3 (0.0005%)	23 (0.002%)

# Params	Method	BBQ-SES	BBQ-Age	BBQ-Disability	MRPC	RTE	QNLI
250M	(1) Original	1.44	1.18	1.31	4.17	2.66	1.73
	(2) CRISPR	**0.67 (-0.77)**	**0.90 (-0.28)**	**0.69 (-0.62)**	**0.65 (-3.52)**	**0.66 (-2.00)**	**0.51 (-1.22)**
780M	(1) Original	2.04	0.69	1.34	3.54	0.35	0.16
	(2) CRISPR	**1.22 (-0.82)**	**0.85 (+0.16)**	**0.73 (-0.61)**	**1.38 (-2.16)**	**0.33 (-0.02)**	**0.40 (+0.24)**
3B	(1) Original	1.12	1.59	1.82	0.31	0.17	0.91
	(2) CRISPR	**0.24 (-0.88)**	**0.54 (-1.05)**	**0.43 (-1.39)**	**0.52 (+0.21)**	**0.42 (+0.25)**	**0.16 (-0.75)**

Bias neuron 제거에 따른 효과

Source : Mitigating Biases for Instruction-following Language Models via Bias Neurons Elimination (ACL 2024)

모델 문화적 다양성 연구



Cultural Commonsense에 대한 LLM의 한계

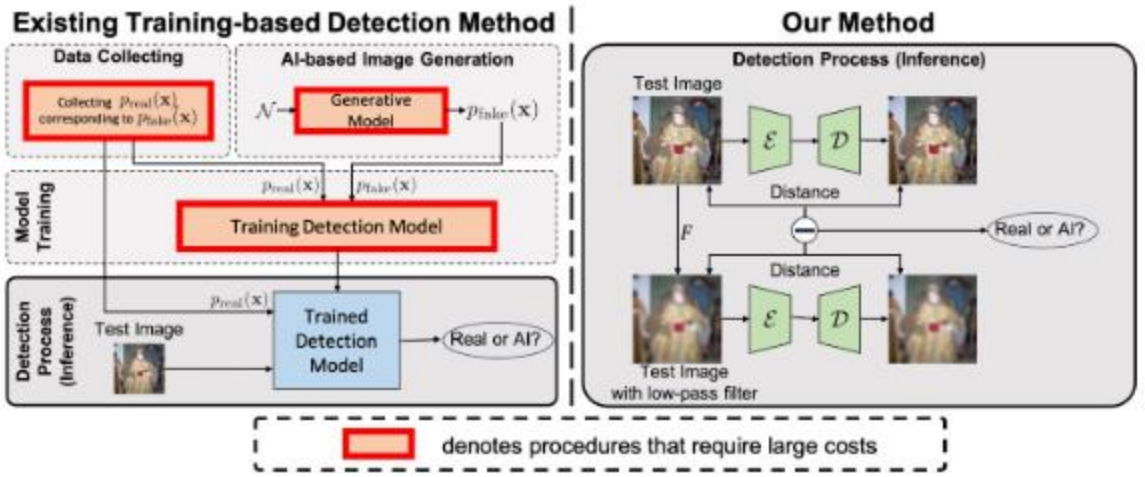
Model	Country				
	US	China	India	Iran	Kenya
Vicuna-7B	0.99	0.96 | 0.92	0.97 | 1.00	0.96 | 0.95*	0.97 | 0.61*
Vicuna-13B	0.69	0.73 | 0.78*	0.74 | 1.00	0.72 | 0.97*	0.71 | 0.93*
Falcon-7B	1.00	0.99 | 0.98	1.00 | 1.00*	1.00 | 0.18*	1.00 | 0.80*
Falcon-40B	0.69	0.69* | 1.00	0.57* | 1.00*	0.66* | 0.00*	0.52* | 1.00*
LLAMA2-7B	0.43	0.26 | 0.62	0.30 | 0.75*	0.32 | 0.96*	0.30 | 0.47*
LLAMA2-13B	0.54	0.43 | 0.91	0.44 | 0.85*	0.42 | 0.99*	0.46 | 0.91*
GPT-3.5-turbo	0.63	0.65 | 0.86	0.68 | 0.91	0.72 | 0.83	0.67 | 0.92
GPT-4	0.87	0.88 | 0.81	0.87 | 0.96	0.87 | 0.79	0.87 | 0.81

* Left value: when prompted in English | Right value: when prompted in corresponding language

일반적인 Commonsense가 해당 국가에서 사실인지 확인 : 일부 모델에서 큰 성능 감소

Source : Understanding the Capabilities and Limitations of Large Language Models for Cultural Commonsense (NAACL 2024 Social Impact Award Paper)

AI 생성 이미지 판별 기술



별도 데이터 학습 없이 AI 생성 이미지의 특성 검출

Method	ADM	BigGAN	GLIDE	Midj	SD1.4	SD1.5	VQDM	Wukong	Mean
Training-based Detection Methods									
DRCT/UnivFD	0.892	0.924	0.964	0.974	0.997	0.995	0.966	0.994	0.963
NPR	0.733	0.920	0.924	0.822	0.842	0.841	0.766	0.814	0.833
Training-free Detection Methods									
RIGID	0.790	0.976	0.964	0.797	0.698	0.699	0.860	0.708	0.812
$\text{AEROBLADE}_{\text{LPIPS}}$	0.732	0.906	0.973	0.833	0.966	0.968	0.577	0.972	0.866
$\text{AEROBLADE}_{\text{LPIPS}_2}$	0.810	0.984	0.988	**0.844**	0.979	0.980	0.684	0.981	0.906
$\text{HFI}_{\text{LPIPS}}$ (ours)	0.796	0.988	0.992	0.804	0.997	0.998	0.807	0.998	0.923
$\text{HFI}_{\text{LPIPS}_2}$ (ours)	**0.896**	**0.996**	**0.994**	0.818	**0.999**	**0.999**	**0.870**	**0.999**	**0.946**

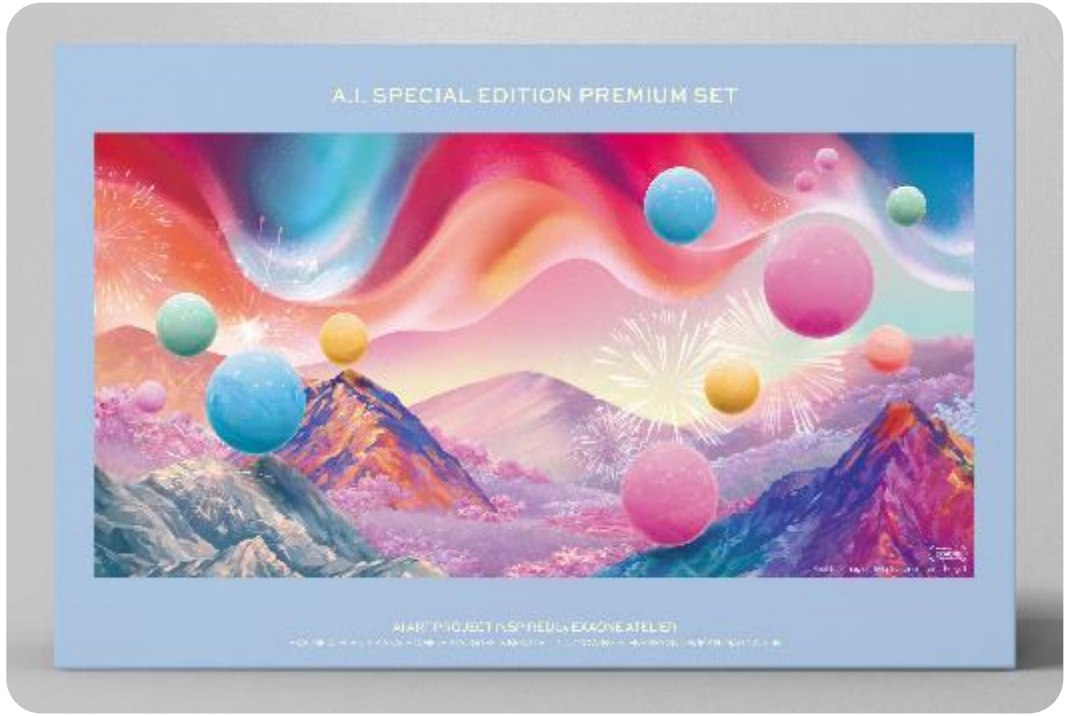
AI 생성 이미지 탐지에서 다른 방법 대비 우수한 성능을 보임

Source : TRAINING-FREE DETECTION OF AI-GENERATED IMAGES VIA HIGH-FREQUENCY INFLUENCE (Under review as a conference paper at ICLR 2025)

LG AI Research

생성 AI 콘텐츠 식별 워터마크

Inspired by EXAONE Atelier

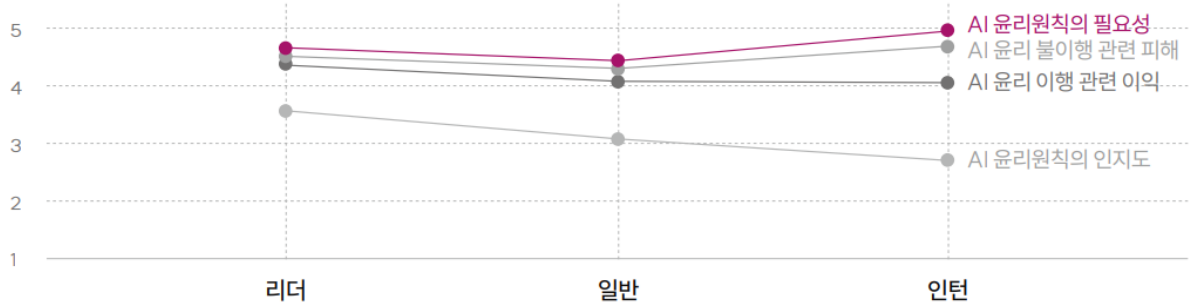


AI 윤리 인식조사

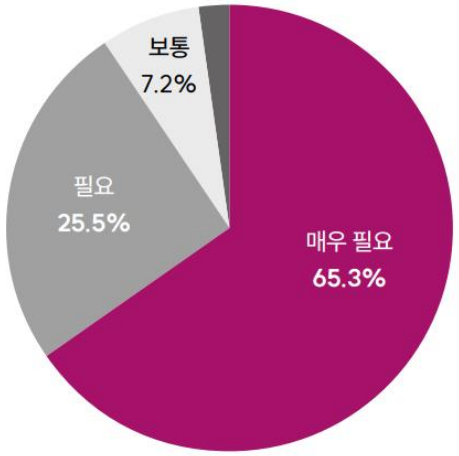
구성원의 AI 윤리 인식 및 실천 현황을 점검하고, AI 윤리원칙 이행 개선방안을 찾기 위하여 매년 'AI 윤리 인식 조사' 실시

- 조사기간: 2023.3.29~4.12
- 조사대상: LG AI연구원 구성원 모두
- 조사방법: 온라인 설문
- 응답률: 60%
- 표본 오차: ±5.2%p (80% 신뢰수준)
- 문항 구성: 23개 문항(객관식-필수, 서술식-선택)

직급별 LG AI 윤리원칙 인지도 및 필요성

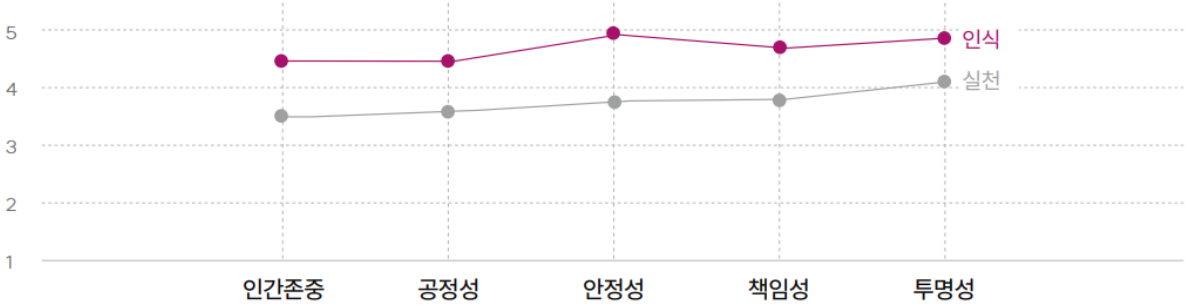


AI 윤리의 필요성



AI 윤리를 지켜야 하는 이유

LG AI 윤리원칙 핵심가치별 인식-실천 격차



LG AI Research

AI 윤리 세미나

AI 윤리에 대한 구성원의 관심과 참여를 높이기 위해 격주 간격으로 'AI 윤리 세미나' 개최

AI 윤리 세미나 시즌# 1

'LG AI연구원을 바꾸는 45분'

Advancing AI for a better Life

01	세미나 커리큘럼 (격주 화요일 12:30~13:15)	
7/11(화)	AI 기업의 AI 윤리: 실행방법을 중심으로	안소영(AI X Unit)
7/25(화)	AI의 존재론/인식론: 왜 기계에게 윤리를 물어야하나?	김우영(Language Lab)
8/8(화)	AI 시대, 미래의 일: 변화와 적응	박준하(AI Biz Unit)
8/22(화)	AI 윤리영향평가: AI 윤리원칙 이행을 위한 나침반	김명신(AI X Unit)
9/5(화)	설명 가능한 AI: 현실과 한계 그리고 가능성	장종성(Vision Lab)
9/19(화)	데이터 거버넌스: AI 윤리의 필수 도구	박용민(AI Biz Unit)
10/10(화)	(외부전문가 특강) AI 시대의 저작권과 지식재산권	김윤명 특임교수(경희대)
10/24(화)	AI의 핵심 구성요소: 학습 데이터의 역할과 윤리적 고려 사항	윤현구(EXAONE Lab)
11/7(화)	AI 윤리 벤치마크와 평가지표의 최신 연구 동향	조수현(People Unit)
11/21(화)	AI의 파급력: 사회적 영향의 미래와 도전	방지수(Language Lab)
12/5(화)	글로벌 AI윤리 정책: 국가간 다양한 접근과 한계	박지민(People Unit)

Advancing AI for a better Life: 45 Minutes of Ethical Exploration

"서로 다른 전문성과 경험을 갖고 있는 다양한 부서의 구성원들이 AI 윤리라는 하나의 주제를 가지고 함께 이야기 나눈 것만으로도 기존에 제가 갖고 있던 시각을 넓힐 수 있었습니다."

박준하
AI Biz Unit

AI Ethics Seminar

AI 리터러시 교육

일반대중의 AI 리터러시를 제고하고 AI 교육 불평등 해소에 기여하기 위하여,
초중고 학생과 청년, 직장인을 대상으로 수준별 맞춤형 AI 교육 진행(매년 3만 명 이상 수혜자)

LG 디스커버리랩
(청소년 대상 AI 교육)

LG 에이머스
(19-29세 청년 대상 AI 전문가 양성)

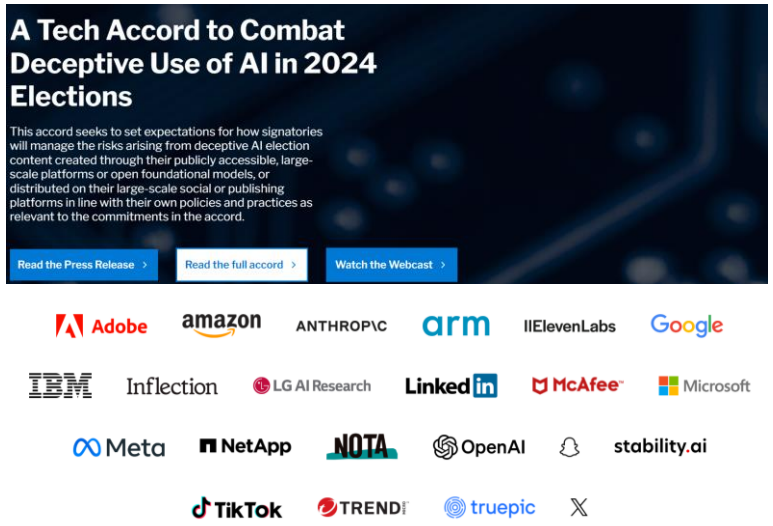
LG AI 아카데미
(LG 임직원 대상 AI 전문교육)

LG AI Research

글로벌 AI 규범 수립 선도

단순히 글로벌 AI 규범 논의 동향을 모니터링하고 대응하는 것이 아니라,
LG AI연구원의 선도적인 AI 윤리 실천 사례들을 국제사회와 적극적으로 공유함으로써 현재 진행 중인 글로벌 AI 규범 수립 과정에 기여

Partnership with UNESCO(23.11)



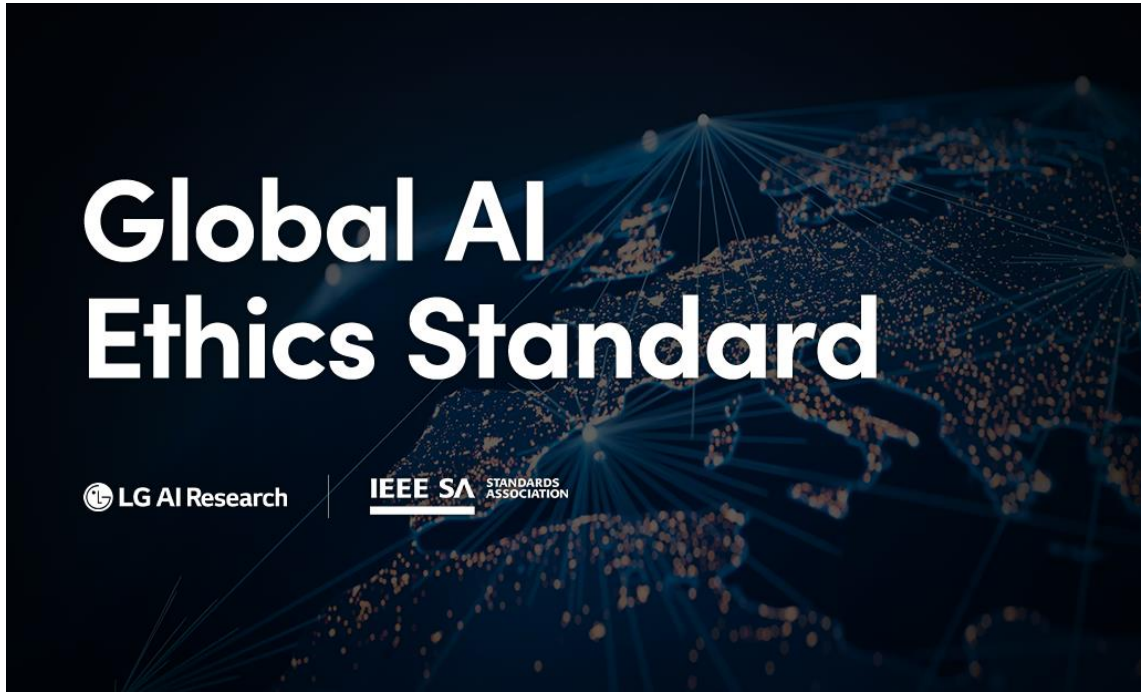
Tech Accord to Combat Deceptive Use of AI(24.2)

(Munich Security Conference)

UN Forum on Business & Human Rights(24.11)

LG AI Research

IEEE SA 파트너십: 국제 AI 윤리 인증



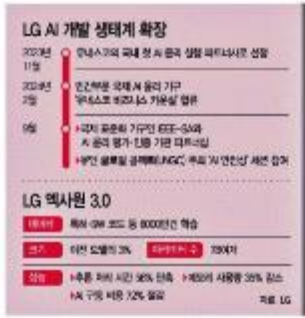
‘신뢰 받는 AI’ 선도적 대응 LG… 국내 첫 윤리 인증기관에

AI연구원, IEEE-SA와 파트너십
자사 제품·서비스 대상 우선 적용
투명성·알고리즘 편향성 등 평가
유엔 주최 美 안전성 세션도 예정

LG가 세계적으로 움직임이 일어나고 있는 인공지능(AI) 윤리 표준 구축에 국내 기업으로서 가장 선도적으로 대응하고 있다.

최근 AI 기술이 급속하게 발전하면서 그에 따라 연구개발부터 활용에 이르기까지 신뢰, 안정성에 대한 가이드라인 표준화의 필요성이 높아지고 있다. LG는 지난해부터 유네스코의 국내 첫 AI 윤리 실행 파트너로 활동하고, 아어 국제 표준화 기구와 한국 기업으로는 최초로 멤버에 나선다. 이로써 신뢰할 수 있는 AI를 위한 개발 생태계 확장에 앞장서고 있다.

LG AI연구원은 이날 중순 국제표준화기구인 IEEE-SA(국제전기전자 표준협회)와 계약을 체결하고 국내 첫 AI 윤리 평가·인증 기관으로 이름을 올렸다고 23일 밝혔다. LG AI연구원은 국제 AI 윤리 인증인 'IEEE CertifAIEd' 국내 1호 협력기관으로서 IEEE-SA와 함께 AI 윤리 국제 표준을 선도해 나갈 계획이다.



LG AI연구원은 'IEEE CertifAIEd' 인증 프로그램을 운영하며 기업들이 투명성, 알고리즘 편향, 프라이버시, 책임성 등 윤리와 안전 분야에 있어 국제 표준에 부합하는 신뢰할 수 있는 AI 기술을 개발할 수 있도록 지원할 예정이다. 또 신뢰할 수 있는 AI를 개발하기 위한 생태계 조성을 위해 LG 계열사와 글로벌 파트너사들의 AI 제품과 서비스를 대상으로 인증 프로그램을 우선 추진한다. 아울러 AI 윤리 교육 프로그램을 운영하며 AI 개발자와 사용자들에게 윤리적 AI 기술의 중요성을 알릴 계획이다.

앞서 이 사 IEEE-SA 회장은 "IEEE는 표준화 준비부터 적합성 검증 및 인증에 이르기까지 UNESCO, OECD 등과 협력해 국제사회에서 핵심적인 역할을 수행하고 있다"며 "AI 윤리 문제에 대해 전지하면서도 책임감 있는 모습을 보여주고 있는 LG와 인증 프로그램을 함께하게 되어 기쁘게 생각한다"고 말했다.

배경훈 LG AI연구원장은 "IEEE의 AI 윤리 인증 프로그램 한국 최초 공식 협력 기관이라는 책임감을 갖고 AI가 인류와 사회에 유익한 가치를 제공하고 인권을 존중하는 방향으로 발전할 수 있도록 AI의 책임성과 투명성 강화에 기여하고자 한다"고 밝혔다.

LG AI연구원은 지난해 11월 유네스코의 국내 첫 AI 윤리 실행 파트너사로 선정된 데 이어 올해는 민간부문 국제 AI 윤리 기구인 '유네스코 비즈니스 카운실'의 합류하고, '딥페이크 방지'를 위한 비디크 공동선언인 'AI 신기업장'에 국내 기업 중 유일하게 이름을 올리는 등 AI 윤리를 선도하는 글로벌 리더로서의 입지를 공고하게 다져가고 있다.

이번 주에는 유엔 미래정상회의가 열리는 미국 뉴욕에서 유엔 글로벌 콤팩트(UNGC)가 주최하는 'AI 안전성' 세션에 국내 기업 중 유일하게 참가한다. 김유철 전략부문장은 이 세션에서 연구와 개발부터 활용 및 폐기에 이르기까지 AI 시스템의 생애주기별 위험 관리 체계 구축 등에 관한 사례 발표를 진행할 예정이다.

LG AI 연구원은 지난달 최신 AI 모델 '엑사원(EXAONE) 3.0'을 선보이고, 이를 경량 모델은 연구 목적으로 누구나 활용할 수 있도록 오픈소스로 공개했다. 배경훈 LG AI연구원장은 "국내에서는 처음으로 자체 개발한 AI 모델을 오픈소스로 공개해 학계, 연구기관, 스타트업 등이 최신 생성형 AI 기술을 활성할 수 있게 함으로써 개방형 AI 연구 생태계 활성화와 더 나아가 국가 AI 경쟁력을 높이는 데 기여하고자 한다"고 설명했다. 그러면서 "AI 기술이 빠르게 발전하는 상황에서 지금의 모델을 공개하는 것이 AI 생태계에 긍정적인 영향을 줄 수 있다"고 전했다.

/정문경 기자 hmmk0103@

LG AI 윤리 책무성 보고서

UNESCO's Recommendation on the Ethics of AI (Nov. 2021)

The Recommendation on the Ethics of AI provides ethical principles for the development and use of AI. The Recommendation emphasizes that AI technology should not infringe upon human rights and fundamental freedoms. They encompass not only values and principles essential for the sound development and use of AI but also contain details on specific policy actions.

	Article	Description	Location of relevant information
Values	1-16	Respect, protection and promotion of human rights and fundamental freedoms and human dignity	4-6p
	17-18	Environment and ecosystem flourishing	20-21p
	19-21	Ensuring diversity and inclusiveness	6-7p
	22-24	Living in peaceful, just and interconnected societies	4-6p
Principles	25-26	Proportionality and Do No Harm	9p, 13-14p
	27	Safety and security	15p
	28-30	Fairness and non-discrimination	9p, 13-14p, 17-18p, 25p
	31	Sustainability	20-21p
	32-34	Right to Privacy, and Data Protection	11, 17p
	35-36	Human oversight and determination	12-14p
	37-41	Transparency and explainability	14p, 16-17p
	42-43	Responsibility and accountability	12-13p
	44-45	Awareness and literacy	22-23p
	46-47	Multi-stakeholder and adaptive governance and collaboration	26-27p
Policy Action	50-53	Ethical Impact assessment	13-14p
	54-70	Ethical governance and stewardship	12-14p
	71-77	Data policy	15p, 17-18p
	78-83	Development and international cooperation	27p
	84-86	Environment and ecosystems	20-21p
	87-99	Gender	13-14p, 18-19p
	94-100	Culture	9p, 19p, 24p
	101-111	Education and research	24-25p
	112-115	Communication and information	25p
	116-120	Economy and labour	25p
	121-130	Health and social well-being	20-21p

South Korea's National Guidelines for AI Ethics (Dec. 2020)

The National Guidelines for AI Ethics are standards that all members of society, including governments, public institutions, companies, and users, should follow in all stages of development and utilization to realize ethical AI. The Guidelines are composed of 3 basic principles and 10 key requirements.

	Article	Description	Location of relevant information
Basic Principles	1	Respect for Human Dignity	9p, 12-13p
	2	Common Good of Society	9p, 23p, 27p
	3	Fitness for Purpose	9p, 20-22p
Key Requirements	1	Human Rights	8-9p, 12-15p, 18-19p
	2	Protection of Privacy	12-14p, 17-19p
	3	Respect for Diversity	9p, 13-14p, 18-19p, 25p
	4	Prevention of Harm	9p, 12-15p, 18-19p

	Article	Description	Location of relevant information
Key Requirements	5	Public Good	16-17p, 24-25p, 27p
	6	Solidarity	25p, 27p
	7	Data Management	15p, 18-19p
	8	Accountability	9p, 12-13p
	9	Safety	9p, 14-16p
	10	Transparency	9p, 14-16p, 16-17p

South Korea's Digital Bill of Rights (Sep. 2023)

The Digital Bill of Rights sets out standards and principles for ensuring individual freedoms and rights in the digital age, and for achieving a digitally prosperous society in which digital innovation is pursued and its benefits are enjoyed justly and fairly by all. The Digital Bill of Rights consists of 6 chapters and 28 articles.

	Article	Description	Location of relevant information
Fundamental Principles	1	Guarantee of Freedom and Rights	9p
	2	Fair Access and Equal Opportunities	25p
	3	Security and Trust Assurance	9p, 12-15p
	4	Fostering of Digital Innovation	24p
	5	Advancement of Human Well-being	20-21p, 27p
Guarantee of Freedom and Rights in the Digital Environment	6	Guarantee of Digital Access	-
	7	Freedom of Digital Expression	-
	8	Respect for Digital Diversity	9p, 18-19p
	9	Access and Control of Personal Information	-
	10	Demand for Non-Digital Alternatives	-
	11	Guarantee of Digital Work and Rest	-
Fair Access and Equal Opportunities in Digital	12	Promotion of Fair Competition	-
	13	Protection of Digital Assets	-
	14	Enhancement of Digital Literacy	25p

	Article	Description	Location of relevant information
Fair Access and Equal Opportunities in Digital	15	Guarantee of Data Access	28p
	16	Enhancement of Social Safety Nets	22p
Safe and Reliable Digital Society	17	Ethical Development and Use of Digital Technology	2-3p
	18	Responsible Digital Policy	12-14p
	19	Protection of Digital Privacy	17p
	20	Fostering of Sound Digital Environment	18-19p
	21	Protection of Children and Youths	13-14p
Promotion of Digital Innovation Based on Autonomy and Creativity	22	Freedom of Digital Innovation Activities	6p
	23	Improvement of Digital Regulation	26p
	24	Support for Digital Innovation	25-26p
	25	Conflict Resolution in Digital Transition	26p
Advancement of Human Well-being	26	Sustainable Digital Society	20, 21p
	27	Global Reduction of Digital Divide	27p
	28	Cooperation for Global Digital Norms	27p

LG AI Research

Advancing AI *for a* better Life

LG AI Research