



보도시점 2024. 11. 27.(수) 09:30
(2024. 11. 27.(수) 석간)

배포 2024. 11. 26.(화) 16:00

5월 윤 대통령 주재 ‘인공지능(AI) 서울정상회의’ 후속조치로

「한국 인공지능 안전연구소」 출범

– 국제 ‘인공지능 3대 협치체계(안전·혁신·포용)’를 실현하고,
인공지능 안전 확보를 위한 국제 관계망 강화

과학기술정보통신부(장관 유상임, 이하 ‘과기정통부’)는 11.27일(수) 판교 글로벌 R&D센터에서 ‘인공지능안전연구소’ 개소식을 개최하였다.

지난 5월 ‘인공지능(AI) 서울정상회의’에서 10개국 정상은 안전이 책임있는 인공지능 혁신을 위한 핵심요소임을 확인하고, 인공지능 안전연구소 설립과 안전한 인공지능에 대한 국제 협력을 강조하였다. 윤석열 대통령도 “대한민국 인공지능 안전연구소 설립을 추진해 국제 인공지능 안전성 강화를 위한 관계망에 동참할 것”이라는 의지를 밝혔으며, 과기정통부는 그간 연구소 조직·예산인력·기능에 대한 면밀한 사전준비를 거쳐 인공지능 안전연구소를 출범하게 되었다.

인공지능 안전연구소는 인공지능의 기술적 한계, 인간의 인공지능 기술 오용, 인공지능 통제력 상실 등으로 발생하는 다양한 인공지능 위협에 체계적·전문적으로 대응하기 위한 인공지능 안전연구 전담조직이다. 산·학·연 기관이 인공지능 안전분야 연구 협력과 정보 교류를 할 수 있는 대한민국 인공지능 안전연구 중심지이자, ‘국제 인공지능 안전연구소 관계망(네트워크)’ (10개국 참여, 11.21.~)의 일원으로 안전한 인공지능을 위한 세계 연대의 책임있는 역할을 수행해 나갈 계획이다. 이를 통해 인공지능 안전 분야의 경쟁력있는 기술·인력을 육성하고, 과학적 연구데이터에 근거한 인공지능 안전 관련정책의 개발·선진화를 추진한다.

개소식에는 류광준 과기정통부 과기혁신본부장, 염재호 국가 인공지능(AI)위원회 부위원장, 이경우 대통령실 인공지능(AI)·디지털 비서관 등 정부 주요인사와 오승필 KT 최고기술책임자(CTO), 오혜연 한국과학기술원(KAIST, 카이스트) 인공지능(AI) 연구원장, 이은주 서울대 인공지능(AI)신뢰성 연구센터장, 방승찬 한국전자통신연구원 원장을 비롯한 인공지능 분야 산·학·연 핵심관계자 40여명이 참석하였다.

이날 국가인공지능(AI)위원회 국제 자문위원인 인공지능 분야 세계적 석학 요슈아 벤지오 교수는 축사를 통해 “서울선언에 따라 한국정부가 인공지능 안전연구소를 설립하게 된 것을 축하하며, 인공지능 안전연구소의 역할로 ①업계와 협업을 통한 위험 평가방법 연구·발전, ②인공지능 안전성 요건 수립 지원, ③국제협력을 통한 인공지능 안전체계의 상호 조화가 중요하다”고 강조하였다.

아울러, 미국 Elizabeth Kelly 인공지능(AI) 안전연구소 소장은 “대한민국이 인공지능 안전 증진을 위해 보여준 세계적 지도력과 지원에 깊이 감사하다”며, “한-미 인공지능(AI)안전연구소가 전세계 동료들과 함께 인공지능 위험을 줄이고 혜택을 극대화하며, 혁신을 촉진하기 위해 공통된 과학적 기준을 구축해 나가자”고 강조하였다. 영국 Oliver Illot 인공지능(AI) 안전연구소 소장과 일본 아키코 무라카미 인공지능(AI) 안전연구소 소장도 “대한민국 인공지능 안전연구소 출범을 축하하며, 인공지능 기술을 안전하게 만들어 가기 위해 국경을 넘어선 협력을 확대해 나가자”고 전했다.

‘인공지능 안전연구소’ 초대소장으로 임명된 김명주 소장은 ‘연구소 이상 및 운영계획’ 발표를 통해 “인공지능 활용으로 발생할 수 있는 제반위험 평가와 이를 예방·극소화하기 위한 정책·기술을 개발·보급하고, 국내·외 연대를 강화해 나갈 것”이라며, “연구소는 규제기관의 성격이 아니라, 국내 인공지능 기업들이 국제 경쟁력을 확보하는데 장애가 되는 위험요소들을 최소화하도록 지원하는 협력기관으로 역할을 하겠다”고 강조하였다.

이어진 ‘대한민국 인공지능(AI) 안전 연합체(컨소시엄)’ 협약식에서는 인공지능 분야 국내 대표 산학연 기관 총 24곳이 참여한 가운데, 인공지능 안전에 관한 정책연구, 평가, 연구개발의 상호협력을 위한 업무협약(MOU)을 체결하였다. 인공지능 안전연구소와 연합체 참여기관이 함께 인공지능 안전 체계(프레임워크)(위험 식별·평가·완화) 연구·개발·실증, 인공지능 안전관련 국제규범 대응 등

정책연구, 인공지능 안전관련 기술 협력 등을 중점적으로 추진할 계획으로, 향후 연합체의 세부 연구주제와 운영방안을 구체화해 나갈 예정이다. 연합체 참여기관들도 인공지능 안전 연구와 관련된 전문성과 기여방안을 제시하며, 인공지능 안전연구소와 적극적으로 협력하겠다는 의지를 밝혔다.

< 「인공지능 안전 연합체(컨소시엄)」 참여기관 >

산업계	네이버(퓨처AI센터), KT(책임있는 AI센터), 카카오(AI Safety), LG AI연구원, SKT(AI 거버넌스 TF), 삼성전자, 코난테크놀로지, 뽀튼테크놀로지스, 이스트소프트, 포티투마루, 클라우드웍스, 트웰브랩스, 라이너
학계	서울대(AI신뢰성 연구센터), 카이스트(AI공정성 연구센터), 고려대(정보보호대학원), 성균관대(AI신뢰성 연구센터), 숭실대(AI안전성 연구센터), 연세대(AI임팩트연구소)
연구계	한국정보통신기술협회(AI신뢰성 센터), 한국지능정보사회진흥원(AI정책본부), (디지털사회 전략연구실), 정보통신정책연구원(AI·디지털융합단), 소프트웨어정책연구소(AI정책연구실)

* '인공지능 안전컨소시엄' 참여기관은 향후 추가로 모집할 예정

유상임 과기정통부 장관은 “인공지능 안전은 지속가능한 인공지능 발전의 전제로 인공지능 분야에 종사하는 우리 모두가 함께 해결해 나가야 할 가장 큰 도전과제이며, 특히, 지난 ‘영국 인공지능(AI)안전성 정상회의’(‘23.11.), ‘인공지능(AI) 서울정상회의’(‘24.5.)’ 이후, 불과 1년 만에 미국, 영국, 일본, 싱가포르, 캐나다 등 주요국이 모두 인공지능 안전연구소를 설립하고, 전례없이 기민하고 체계적인 국제 인공지능 안전 협력체계가 구축되고 있다.”고 밝히며,

“우리도 인공지능 안전연구소를 중심으로 산·학·연 연구역량을 결집해 인공지능 안전의 기술·정책적 역량을 조속히 확보하여, 대한민국 인공지능 산업 혁신을 촉진할 튼튼한 안전기반을 구축해 나가는 한편, 국제 인공지능 안전 연대의 주도적인 역할을 수행하여 아태지역을 대표하는 인공지능 안전 연구중심지로 발전해 나갈 수 있도록 적극 지원하겠다.”고 강조하였다.

담당 부서	과학기술정보통신부 인공지능기반정책과	책임자	과 장	남철기 (044-202-6270)
		담당자	사무관	이재호 (044-202-6262)

AISI 인공지능안전연구소

5월 'AI서울정상회의' 후속조치, 'AI 3대 거버넌스' 실현

AI서울
정상회의

"대한민국은 국제사회의 책임있는 일원으로서
AI안전, 혁신, 포용을 추진해 나가겠습니다."

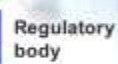
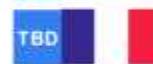
"대한민국도 AI안전연구소 설립을 추진해
글로벌 AI안전 강화를 위한 네트워크에 동참할 것입니다."

윤석열 대통령 ('24.5.21)

안전한 AI 위한 글로벌 연대 중심

국제 AI 안전연구 네트워크 출범('24.11~)

Evaluation of AI systems, foundational research, facilitation of information exchange



아태지역을 대표하는 글로벌 AI안전 거점연구소 구현

AISI 비전 및 핵심과제

Vision

안전한 AI로 만드는 지속가능한 미래

Creating a Sustainable Future with Safe AI

미션

Mission 1

AI안전에 대한
과학적 이해 증진

Mission 2

AI안전 정책 고도화
및 안전 제도 확립

Mission 3

국내 AI기업의
안전 확보 지원

핵심
가치

히말라야 등정을 돕는
셰르파(Sherpa)같은 연구소



AI안전에서 전문성을 갖춘
AI기업의 헬퍼

핵심
과제

핵심과제 1

AI안전 평가

핵심과제 2

AI안전 정책 연구

핵심과제 3

AI안전 글로벌 협력

핵심과제 4

AI안전 R&D

□ 행사 개요

- (일시/장소) '24.11.27.(수), 09:30~10:30 / AI안전연구소(판교 글로벌R&D센터)
* 약 330평 규모(개인 연구공간 33실, 대·중·소회의실, 세미나룸 등)
- (참석) [정부] 과기혁신본부장, 국가AI위원회 부위원장, 개보위 부위원장 등
[공공] AI안전연구소장, ETRI원장, IITP원장, NIA원장 등
[민간] AI대기업·스타트업, 학계, 연구소설립준비·자문위원 등
- (주요내용) ①도착·환담 → ②개회 → ③축사(과기혁신본부장, 국가AI위원회 부위원장, 글로벌 석학, 주요국 AI안전연구소장) → ④개소 선언, 비전 및 운영계획 소개 → ⑤'대한민국 AI안전컨소시엄' 협약 체결 → ⑥참석자 기념촬영 → ⑦현판 제막식 및 기념촬영 → ⑦연구소 투어 및 직원 격려

□ 세부 계획(안)

시간		주요내용		비고
~ 09:30	-	■ 과기혁신본부장 도착 및 사전환담		1층(VIP 대기실)
09:30~09:35	(5')	■ 개회, 내빈 소개		1층 (로비)
09:35~09:39	(4')	■ 축사	과기혁신본부장	
09:39~09:42	(3')		국가AI위원회 부위원장(태재대 염재호 총장)	
09:42~09:47	(5')		글로벌 석학, 주요국 AI안전연구소장(영상) ※ 요슈아 벤지오 교수, 美·英·日 AI안전연구소장	
09:47~09:57	(10')	■ 연구소 개소 선언, 비전 및 운영계획 소개 ※ AI안전연구소장		
09:57~10:04	(7')	■ '대한민국 AI안전컨소시엄' 협약 체결(MOU) ※ 본부장, 연구소장, 산·학·연 기관(24개) 대표 참여		
10:04~10:07	(3')	■ 참석자 기념촬영 ※ 본부장, 전체 참석자(40여명)		
10:07~10:10	(3')	■ 연구소 사무실(4층) 이동		1층→4층
10:10~10:15	(5')	■ 현판 제막식 및 기념촬영 ※ 본부장 포함 총 10명		4층
10:15~10:25	(10')	■ 연구소 투어 및 격려 ※ 연구소 직원 총 12명		(연구소)
10:25~10:30	(5')	■ 과기혁신본부장 환송		1층

참고3

참석자 명단

No	구분	성명	소속	직위	비고
1	정부(7)	류광준	과기정통부	본부장	과기혁신본부
2		최장혁	개인정보보호위	부위원장	-
3		송상훈	과기정통부 정보통신정책실	실장	-
4		이경우	대통령실 과기수석실	비서관	AI·디지털
5		김현욱	대통령실 국가안보실	비서관	경제안보
6		신용석	대통령실 국가안보실	비서관	사이버안보
7		김경만	과기정통부 AI기반정책관	국장	-
8	국가AI 위원회(5)	염재호	태재대	총장	부위원장
9		고진	한국무역정보통신	대표	분과장
10		강민구	법무법인 도울	대표	분과장
11		오혜연	카이스트 AI연구원	원장	분과장
12		임종인	대통령실	사이버특보	위원
13	산업계 (13)	손지윤	네이버 정책전략총괄	전무	'서울 AI기업 서약'(24.5.) 참여기업
14		한진규	삼성리서치 기술표준연구팀	상무	
15		이상호	카카오	성과리더	
16		임우형	LG AI연구원	선임랩장	
17		엄종환	SKT AI거버넌스 TF	부사장	
18		오승필	KT	CTO	AI중소· 스타트업
19		김영섬	코난테크놀로지	대표	
20		이세영	뤼튼테크놀로지스	대표	
21		정상원	이스트소프트	대표	
22		김동환	포티투마루	대표	
23		김우승	클라우드웍스	대표	
24		김윤	트웰브랩스	CSO	
25		우찬민	라이너	대표(韓 법인)	
26	학계 (6)	이은주	서울대 AI신뢰성 연구센터	센터장	연구협력
27		유창동	KAIST AI공정성 연구센터	센터장	
28		송경희	성균관대 AI신뢰성 연구센터	센터장	
29		최대선	숭실대 AI안전성 연구센터	센터장	
30		권헌영	고려대 정보보호대학원	원장	
31		박기영	연세대 AI Impact연구소	소장	
32	공공 (6)	방승찬	ETRI	원장	
33		홍진배	IITP	원장	
34		황종성	NIA	원장	
35		이재범	TTA	소장(대참)	
36		김정언	KISDI	원장대행	
37		김형철	SPRI	소장	
38	설립준비· 자문위원(2)	김경훈	카카오 AI Safety	이사	위원
39		박성필	카이스트 문술미래전략대학원	원장	