

보도시점 2023. 12. 12.(화) 10:00
(2023. 12. 12.(화) 석간)

배포 2023. 12. 11.(월) 16:00

안전한 인공지능 확산을 위한 '한·영 공동 인공지능 안전·신뢰성 세미나' 개최

- ▶ 지난 11월 「인공지능 안전성 정상회의」 후속조치 일환으로 양국 인공지능 전문가들이 모여 인공지능 안전성 확보를 위한 기술교류 및 협력방안 모색

과학기술정보통신부(장관 이종호, 이하 '과기정통부')와 한국정보통신기술협회(회장 손승현, 이하 'TTA')는 '한-영 공동 인공지능 안전·신뢰성 세미나'를 12월12일(화)부터 이틀간 경기도 판교 더블트리 바이 힐튼호텔에서 개최한다.

※ 행사 1일차는 공개세미나, 2일차는 비공개 토론행사로 진행

이번 행사는 지난 11월 영국 블레츨리 파크에서 개최된 「인공지능(이하 'AI') 안전성 정상회의」 논의의 연장선 상에서, AI 안전·신뢰성을 확보하기 위해 양국 간 기술 교류와 협력을 강화하는 차원에서 개최되었다.

세계 최고 AI 연구소로 불리는 영국 앨런튜링연구소를 포함하여 딥마인드, 삼성전자, 네이버, 서울대 등 양국을 대표하는 AI 기업·기관이 참석한 가운데, AI 안전·신뢰성 관련 표준화R&D 동향을 공유하고 부문별 협력 방안을 모색하였다.

< '한-영 공동 AI 안전·신뢰성 세미나' 개요 >

- (일자/장소) 12.12(화)~12.13(수) / 경기도 판교 더블트리 바이 힐튼호텔
- (참석자) 영국 앨런튜링연구소 Ardi Janjeva 수석연구원, 삼성전자 글로벌 AI 센터 이주형 부사장, 네이버클라우드 하정우 센터장, 서울대 이경무 교수, TTA AI 융합 기획단 이강해 단장, 딥마인드 국제정책담당 Sebastien Krier 등 전문가 100여명
- 주요내용
 - ① (기술 세미나) 영국 AI 안전테스트 프레임워크 등 전문가 강연 및 네트워킹
 - ② (협력방안 논의) AI 안전성 확보를 위한 양국 기업·기관 간 협력 방안 논의
 - ③ (워크숍 발족) 민간 기술 교류 및 협력체계 구축, 양국 간 교류 정례화 선언

먼저, 영국 앨런튜링연구소 Ardi Janjeva 수석연구원은 기조강연을 통해 ‘생성형 AI의 급속한 발전에 따른 AI 안전성 및 국가안보에 대한 위험 평가 방안’과 ‘국가 안보를 위한 제3자 AI 보증방안’을 소개하며, AI 안전성을 확보하기 위한 인프라 및 기술 연구 관점에서의 제언과 모범사례 구축을 강조하였다.

이어진 기조강연에서 삼성전자 글로벌 AI 센터 이주형 부사장은 챗 GPT 시대의 책임있는 AI 확장을 위한 기술적 도전과제와 윤리적·법적 고려사항에 대해 발표하였다.

전문세션에서는 ①세계를 선도하는 AI 전략과 ②신뢰와 윤리를 향한 AI 혁신을 주제로 각 기업·기관*에서 추진전략을 소개하고, AI의 혁신과 안전·신뢰성의 조화로운 발전에 대한 공감대 위에서, 기업·기관 간 협력방안을 모색하였다.

- * (TTA) AI 안전·신뢰성 글로벌 선도·혁신을 위한 TTA 전략
- * (Advai) 예측 불가능한 AI 시스템 보안의 필요성
- * (네이버) 네이버의 책임있는 프론티어 AI 전략
- * (업스테이지) 윤리적이고 신뢰할 수 있는 초거대언어모델을 위한 업스테이지의 비전
- * (딥마인드) 구글 딥마인드의 AI 안전 및 윤리 접근방식
- * (에이아이트릭스) 의료분야에서 신뢰할 수 있는 AI

과기정통부 엄열 AI기반정책관은 “오늘 행사는 지난 11월 AI 안전 정상 회의에서 합의된 안전한 AI를 향한 국제 공조를 이행하고, 양국의 기술과 모범사례를 교류한다는 점에서 의미가 깊다” 고 밝히며, “최근 빠르게 발전하는 생성형 AI의 안전성과 신뢰성 확보를 위해 정부는 기업과 국제사회와 함께 해결방안을 강구하고, 내년 5월 영국과 공동 개최예정인 AI 안전성 정상 회의의 논의를 주도해 나가겠다” 고 말했다.

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

韓-英 공동 인공지능(AI) 안전·신뢰성 세미나

2023. 12. 12. 화 10:00 - 16:30

더블트리 바이 힐튼 서울 판교, 그랜드볼룸

지난 2023년 11월 영국에서 개최된

AI Safety Summit(AI 안전성 정상회의) 후속 이행

조치의 일환으로, 인공지능 신뢰성 및 안전성 분야

기술 교류 및 협력 체계를 구축하기 위하여

‘인공지능 안전·신뢰성 세미나’를 준비하였습니다.

여러분들의 관심과 참여를 부탁드립니다.



과학기술정보통신부
Ministry of Science and ICT



British Embassy
Seoul



TTA 인공지능정보통신기술원

프로그램

09:30~10:00 등록 및 비표수령

10:00~10:30 개회식

개회사 한국정보통신기술협회 송승현 회장

축사 과학기술정보통신부 웹 별 인공지능기반정책관

축사 주한영국대사관 Tony Clemson 참사관

10:30~11:00 해외 기조강연

1 The Alan Turing Institute / Ardi Janjeva Research Associate
The Rapid Rise of Generative AI: Assessing the Risk to AI Safety and National Security

(생성 AI의 급속한 부상 : 인공지능 안전성 및 국가 안보에 대한 위험 평가)

2 The Alan Turing Institute / Rosamund Powell Research Associate
Assurance of Third-Party AI for UK National Security

(영국 국가 안보를 위한 제3자 인공지능 보증)

11:00~11:30 국내 기조강연

삼성전자 / 이주형 Global AI센터 부사장

ChatGPT 시대의 책임있는 인공지능 확산 방안

(How to scale responsibly in the era of ChatGPT)

11:30~13:00 점심식사

13:00~14:30 전문세션① 세계를 선도하는 AI 전략

1 한국정보통신기술협회(TTA) / 이강해 AI융합기획단 단장
인공지능 안전·신뢰성 글로벌 선도·혁신을 위한 TTA의 전략
(TTA's Strategy for Global Leadership and Innovation in AI Safety and Trustworthiness)

2 Advai / David Sully CEO and Co-Founder
The need for Securing Unpredictable AI Systems
(예측 불가능한 인공지능 시스템 보안의 필요성)

3 네이버 / 하정우 센터장
네이버의 책임 있는 프론티어 인공지능 전략
(NAVER's Strategy for Responsible Frontier AI)

공동 Q&A 세션

14:30~15:00 티 브레이크 및 네트워킹

15:00~16:30 전문세션② 신뢰와 윤리를 향한 AI 혁신

1 업스테이지 / 박찬준 LLM Tech Leader
윤리적이고 신뢰할 수 있는 LLM을 위한 업스테이지의 비전
(Upstage's Vision for Ethical and Trustworthy LLM)

2 Google DeepMind / Sebastien Krier International Policy Manager
Google DeepMind's approach to safety and ethics of AI
(구글 딥마인드의 인공지능 안전 및 윤리 접근 방식)

3 에이아이트릭스 / 한병은 CSO
의료 분야에서 신뢰할 수 있는 인공지능
(Trustworthy AI in medical field)

공동 Q&A 세션