**[Workshop Update] International Cooperation on CBRN AI Security Risk in Livermore, USA**

From August 25 to 28, 2025, Korea Artificial Intelligence Safety Institute (Korea AISI) participated in the AI Security Risk International Workshop held in Livermore, California, USA. This workshop was supported by the U.S. Department of State and jointly brought together the Lawrence Livermore National Laboratory (LLNL), the Pacific Northwest National Laboratory (PNNL), and Korea AISI.

The central theme of the workshop was CBRN (Chemical, Biological, Radiological, and Nuclear) risks in the context of artificial intelligence, with a particular focus on preventing the misuse of advanced AI models in these sensitive areas. Participants discussed how large language models (LLMs) could potentially be exploited in biological or chemical threat scenarios, and how such risks can be effectively mitigated.

Key topics included ▲AI safety evaluation methods ▲red teaming practices ▲benchmarking frameworks—all highlighted not only as technical challenges but also as essential components of international security efforts. Sessions also addressed translation research, counter-censorship benchmarking, and hands-on red teaming and hacking exercises. Experts from the U.S. National Institute of Standards and Technology (NIST) Center for AI Standards and Innovation (CAISI) and MITRE contributed by sharing the latest threat analyses and mitigation strategies directly relevant to national and global security.

At the workshop, Korea AISI emphasized that "AI safety and security are shared challenges that transcend borders." In particular, we underscored the importance of building international cooperation to prevent the misuse of AI in the CBRN domain. The event also reaffirmed Korea's growing role as a key partner in shaping the global AI safety and security governance landscape.

Looking ahead, the 2nd workshop will be held in Korea from November 10 to 13, 2025, where discussions will be deepened and concrete avenues for cooperation will be further explored. Korea AISI will continue to work closely with international partners to help safeguard humanity through stronger technical and policy responses to AI-related risks.


STRENGTHENING NATIONAL SECURITY
Innovations in experiments, theories, and computation lead to enhanced predictive capabilities
that help ensure the safety, security, and effectiveness of our nation's nuclear deterrent.
Additive Manufacturing

**[워크숍 소식] 미국 리버모어에서 열린 CBRN AI 안보 리스크 국제 협력**

2025년 8월 25일부터 28일까지 인공지능안전연구소는 미국 캘리포니아주 리버모어에서 개최된 AI 안보 리스크 국제 워크숍에 참가했습니다. 이번 워크숍은 미국 국무부의 지원으로 진행되었으며, 로렌스 리버모어 국립연구소(LLNL), 퍼시픽 노스웨스트 국립연구소(PNNL), 그리고 인공지능안전연구소가 함께했습니다.

워크숍의 핵심 주제는 CBRN(Chemical, Biological, Radiological, and Nuclear: 화학·생물·방사능·핵) 분야에서의 인공지능 위험이었으며, 특히 첨단 AI 모델이 이러한 민감한 영역에서 오남용되는 것을 방지하는 데 초점이 맞추어졌습니다. 참가자들은 대형 언어모델(LLM)이 생물학적 또는 화학적 위협 시나리오에서 어떻게 악용될 수 있는지, 그리고 이러한 위험을 어떻게 효과적으로 완화할 수 있는지를 논의했습니다.

주요 논의 주제는 ▲AI 안전성 평가 방법 ▲레드팀(red teaming) 실습 ▲벤치마킹 프레임워크 등이었으며, 이는 단순한 기술적 과제가 아니라 국제 안보를 위한 핵심 요소임이 강조되었습니다. 또한 번역 연구, 검열 회피(counter-censorship) 벤치마킹, 레드팀 및 해킹 실습 등도 다뤄졌습니다. 이와 함께 미국 국립표준기술연구소(NIST) 산하 AI표준혁신센터(CAISI)와 MITRE 연구진이 참여하여 국가 및 글로벌 안보와 직결된 최신 위협 분석과 대응 전략을 공유했습니다.

이번 워크숍에서 인공지능안전연구소는 "AI 안전과 안보는 국경을 초월한 공동 과제"임을 강조했습니다. 특히 CBRN 분야에서의 AI 오남용을 막기 위해 국제 협력 체계를 구축하는 것이 중요하다는 점을 제시했으며, 한국이 글로벌 AI 안전·안보 거버넌스 형성에서 중요한 파트너임을 다시 한번 확인했습니다.

앞으로 제2차 워크숍은 2025년 11월 10일부터 13일까지 한국에서 개최될 예정이며, 이번 논의를 심화하고 보다 구체적인 협력 방안을 모색하게 될 것입니다. 한국 AISI는 국제 파트너들과 긴밀히 협력하여 AI 관련 위험에 대응하는 기술적·정책적 역량을 강화하고, 인류의 안전을 지키는 데 기여하겠습니다.