ETRI 인공지능안전연구소, OpenAI 와 AI 안보·보안 공동 워크숍 개최

국방·외교·보안 분야 관계자 등 60 여 명 참석…AI 국가안보 대응방안 논의
CBRN·사이버·국가연계 위협 등 첨단 AI 가 가져올 새로운 안보위협 집중 점검

ETRI 인공지능안전연구소(K-AISI)는 10 월 19 일부터 20 일까지 이틀간 서울 양재 L 타워에서 OpenAI 와 'AI 안보·보안 분야 공동 워크숍'을 개최했다. 이번 워크숍은 인공지능(AI)의 빠른 발전이 국가안보 환경에 미치는 영향을 살펴보고, 새롭게 등장하는 위협에 대한 기술적·정책적 대응방안을 논의하기 위해 마련됐다.

이번 행사에는 국방·외교·보안 분야 관계자 등 약 60 명이 참석해 AI 기술 발전에 따른 주요 안보 현안과 대응 방향을 논의했다. 참석자들은 AI 국가안보를 비롯해 화학·생물·방사능·핵(CBRN), 사이버 위협, 국가연계 행위자(State-linked actors), 테러와 초국가범죄, 데이터 유출 등 다양한 안보 현안을 폭넓게 살펴봤다.

첫날인 19 일에는 AI 국가안보와 CBRN·사이버 위협을 중심으로 논의가 진행됐다. ETRI 인공지능안전연구소는 AI 국가안보의 개념과 범위, 연구소가 추진하고 있는 관련 연구와 대응 방향을 소개했으며, OpenAI 는 첨단 AI 발전에 따른 글로벌 위협 환경과 AI 안전·보안에 대한 접근방식을 공유했다.

CBRN 분야에서는 AI 발전에 따라 새롭게 나타날 수 있는 위협과 기존 CBRN 위협의 차이, 이에 대한 대응전략을 논의했다. OpenAI 는 특히 생물안보 분야에서 AI 가 악용될 가능성과 이를 줄이기 위한 안전조치를 소개했으며, 관련 기술에 대한 설명과 시연도 진행했다.

이어 에이전틱 AI(Agentic AI)와 AI 기반 코딩 도구의 발전에 따른 사이버 위협 환경 변화도 주요 의제로 다뤄졌다. OpenAI 는 최신 AI 모델을 사이버 방어 업무에 활용하고, 기존 방어 업무 흐름에 통합하는 방안을 소개했다.

20 일에는 국가연계 행위자, 테러와 폭력적 극단주의, 초국가범죄, 데이터 유출 및 AI 기반 사기 등 보다 구체적인 안보위협으로 논의를 확대했다. 국가연계 행위자들이 AI 를 정보·영향력 활동, 첩보, 이중용도 연구 등에 활용하는 사례와 AI 가 테러·범죄 활동에 악용될 가능성, 관련 위험을 탐지·분석하는 방안 등이 공유됐다.

ETRI 인공지능안전연구소는 외부 도구를 사용하는 LLM 기반 AI 에이전트의 데이터 유출 위험을 소개하고, 민감한 정보의 노출 가능성과 이를 완화하기 위한 대응방안을 발표했다.

이와 함께 국가안보 분야에서 AI 를 실무에 활용하는 방안도 논의됐다.

이번 워크숍은 ETRI 인공지능안전연구소의 AI 국가안보 분야 연구·평가 경험과 OpenAI 의 AI 안전·보안 관련 기술적 경험을 국내 관계기관과 공유했다는 점에서 의미가 있다. 특히 국방·외교 등 다양한 분야의 관계자들이 한자리에 모여 AI 가 국가안보 전반에 미치는 영향을 살펴보고, 기술적 안전조치와 정책적 대응을 연계하기 위한 협력 방향을 논의했다.

김명주 ETRI 인공지능안전연구소장은 "AI 기술이 빠르게 발전하면서 국가안보 영역에서 고려해야 할 위험 역시 더욱 다양하고 복합적인 형태로 변화하고 있다"며 "이번 워크숍은 국내 관계기관과 글로벌 프론티어 AI 개발 기업이 실제 위협 양상과 대응 경험을 공유하고 함께 대응방안을 모색했다는 점에서 의미가 크다"고 말했다. 이어 "앞으로도 CBRN 과 사이버를 비롯한 AI 국가안보 분야의 위험을 선제적으로 평가하고, 국내외 관계기관 및 글로벌 AI 기업과의 실질적인 협력을 지속적으로 확대해 나가겠다"고 밝혔다.

양일간의 발표 이후에는 질의응답과 참석자 간 의견교환을 통해 향후 협력 분야도 논의됐다. ETRI 인공지능안전연구소는 이번 워크숍을 계기로 국내외 관계기관 및 프론티어 AI 개발 기업과의 연구·정보 교류를 확대하고, 새롭게 등장하는 AI 안보위협에 대한 평가와 대응 역량을 지속적으로 강화해 나갈 계획이다.

Korea AI Safety Institute Holds Joint AI Security Workshop with OpenAI

About 60 participants from defense, foreign affairs, and security discuss national security responses to AI
Focused review of new security threats posed by advanced AI, including CBRN, cyber, and state-linked threats

The Korea AI Safety Institute (K-AISI) held a two-day "Joint Workshop on AI Security" with OpenAI on October 19–20 at L Tower in Yangjae, Seoul. The workshop was organized to examine how the rapid advancement of artificial intelligence (AI) is affecting the national security environment and to discuss technical and policy responses to emerging threats.

About 60 participants, including officials from the defense, foreign affairs, and security sectors, attended the event to discuss key security issues arising from AI advancement and directions for response. Participants covered a wide range of security issues, including AI and national security; chemical, biological, radiological, and nuclear (CBRN) threats; cyber threats; state-linked actors; terrorism and transnational crime; and data leakage.

On the first day, the 19th, discussions centered on AI and national security, along with CBRN and cyber threats. The Korea AI Safety Institute presented the concept and scope of AI national security, as well as the institute's related research and response strategies. OpenAI shared its view of the global threat landscape shaped by advanced AI development and its approach to AI safety and security.

In the CBRN session, participants discussed threats that may newly emerge as AI advances, how these differ from existing CBRN threats, and strategies for responding to them. OpenAI focused on the potential misuse of AI in the biosecurity domain and the safeguards in place to reduce it, and also explained and demonstrated related technologies.

Changes in the cyber threat landscape driven by advances in agentic AI and AI-powered coding tools were also a major agenda item. OpenAI introduced ways to use its latest AI models for cyber threat response and defense operations, and demonstrated technical approaches to applying AI in real-world cyber defense.

On the 20th, the discussion expanded to more specific security threats, including state-

linked actors, terrorism and violent extremism, transnational crime, and data leakage and AI-enabled fraud. Participants shared cases of state-linked actors using AI for information and influence operations, espionage, and dual-use research, as well as the potential for AI to be exploited in terrorist and criminal activities and methods for detecting and analyzing related risks.

The Korea AI Safety Institute presented the data leakage risks of LLM-based AI agents that use external tools, outlining the potential exposure of sensitive information and measures to mitigate it. Participants also discussed practical applications of AI in national security.

The workshop is significant in that it shared the Korea AI Safety Institute's research and evaluation experience in AI national security, along with OpenAI's technical experience in AI safety and security, with relevant domestic agencies. In particular, it brought together officials from diverse fields such as defense, foreign affairs, and security to examine AI's impact on national security as a whole and to discuss ways to link technical safeguards with policy responses.

"As AI technology advances rapidly, the risks that must be considered in the national security domain are also becoming more diverse and complex," said Kim Myuhng-joo, Director of the Korea AI Safety Institute. "This workshop is highly meaningful in that domestic agencies and a global frontier AI developer shared actual threat patterns and response experience and explored countermeasures together." He added, "We will continue to proactively assess risks in AI national security, including CBRN and cyber, and steadily expand practical cooperation with domestic and international agencies and global AI companies."

Following the presentations on both days, participants discussed future areas of cooperation through Q&A sessions and open exchanges of views. Building on this workshop, the Korea AI Safety Institute plans to expand research and information exchange with domestic and international agencies and frontier AI developers, and to continue strengthening its capabilities to evaluate and respond to emerging AI security threats.


AISI
Korea AISI×
OpenAI:
Joint Safety &
Security
Intelligence
Workshop
AISI
Korea AISI×
OpenAI:
Joint Safety &
Security
Intelligence
Workshop


AISI
Korea AISI×
OpenAI:
Joint Safety &
Security
Intelligence
Workshop
AISI
Korea AISI×
OpenAI:
Joint Safety &
Security
Intelligence
Workshop