AISI Korea AI Safety Institute

제4차
AI 안전 정책 연구 세미나

Agentic AI와 AI 안전 연구

DATE
2026. 8. 12. 수
18:00 — 20:00

VENUE
온라인 (Zoom)
한영 동시 통역 제공

안녕하세요, 인공지능안전연구소입니다. 국내외 전문가들과 함께 AI 안전의 주요 쟁점과 정책적 과제를 논의하는 「제 4차 AI 안전 정책 연구 세미나」를 진행합니다.
이번 세미나에서는 The Centre for Long-Term Resilience와 Singapore AI Safety Hub (SASH) 연구진을 초청하여 Research on Agentic AI and AI Safety를 주제로 AI의 Scheming 및 Loss of control, Agent IDs를 중심으로 최신 연구와 정책적 시사점을 논의할 예정입니다.

Program

18:00	**Introduction**	10 min
18:10	**Scheming in the wild –** ***key findings from the Loss of Control Observatory*** **Tommy Shaffer Shane**, Senior AI Policy Manager, The Centre for Long-Term Resilience	40 min + 10 min Q&A
19:00	**AgentID:** **Identifying, authorizing and shutting down AI agents** **Sam Boger**, Technical Lead, Singapore AI Safety Hub (SASH) **Amin Oueslati**, Governance Lead, Singapore AI Safety Hub (SASH) & Research Affiliate, Oxford Martin AI Governance Initiative	40 min + 10 min Q&A
19:50	**Closing**	10 min

참가 신청
다음 링크로 등록 신청해 주세요
https://forms.gle/Pj7XFbWkjgg9PyGK6
문의 · mypark56@etri.re.kr

ETRI

AISI AI Safety Institute

AISI Korea AI Safety Institute

The 4th
AI Safety Policy Research Seminar

Research on Agentic AI and AI Safety

DATE

2026. 8. 12. Wed

18:00 — 20:00

VENUE

Online (Zoom)

KR-EN simultaneous interpretation available

The Korea AI Safety Institute regularly hosts the AI Safety Policy Research Seminar, bringing together domestic and international experts to discuss key issues and policy challenges in AI safety. The fourth seminar will focus on **"Research on Agentic AI and AI Safety"**

Researchers from The Centre for Long-Term Resilience and the Singapore AI Safety Hub (SASH) will present recent research on real-world AI scheming, loss-of-control incidents, and Agent IDs for trustworthy AI agents, and discuss methodological approaches and policy implications for AI safety and governance.

Program

18:00	**Introduction**	10 min
18:10	**Scheming in the wild –** ***key findings from the Loss of Control Observatory*** **Tommy Shaffer Shane**, Senior AI Policy Manager, The Centre for Long-Term Resilience	40 min + 10 min Q&A
19:00	**AgentID:** **Identifying, authorizing and shutting down AI agents** **Sam Boger**, Technical Lead, Singapore AI Safety Hub (SASH) **Amin Oueslati**, Governance Lead, Singapore AI Safety Hub (SASH) & Research Affiliate, Oxford Martin AI Governance Initiative	40 min + 10 min Q&A
19:50	**Closing**	10 min

Registration

Please register by using the link below.

https://forms.gle/Pj7XFbWkjgg9PyGK6

Contact · mypark56@etri.re.kr

ETRI

AISI AI Safety Institute