

NAIRL Frontier LETTER

2026 | Vol.04

NAIRL Special Column

김종철 KT 전문위원, KAIST 겸직교수

중장기 AI 기술 로드맵과 피지컬 AI의 미래 전망

NAIRL Research Highlight

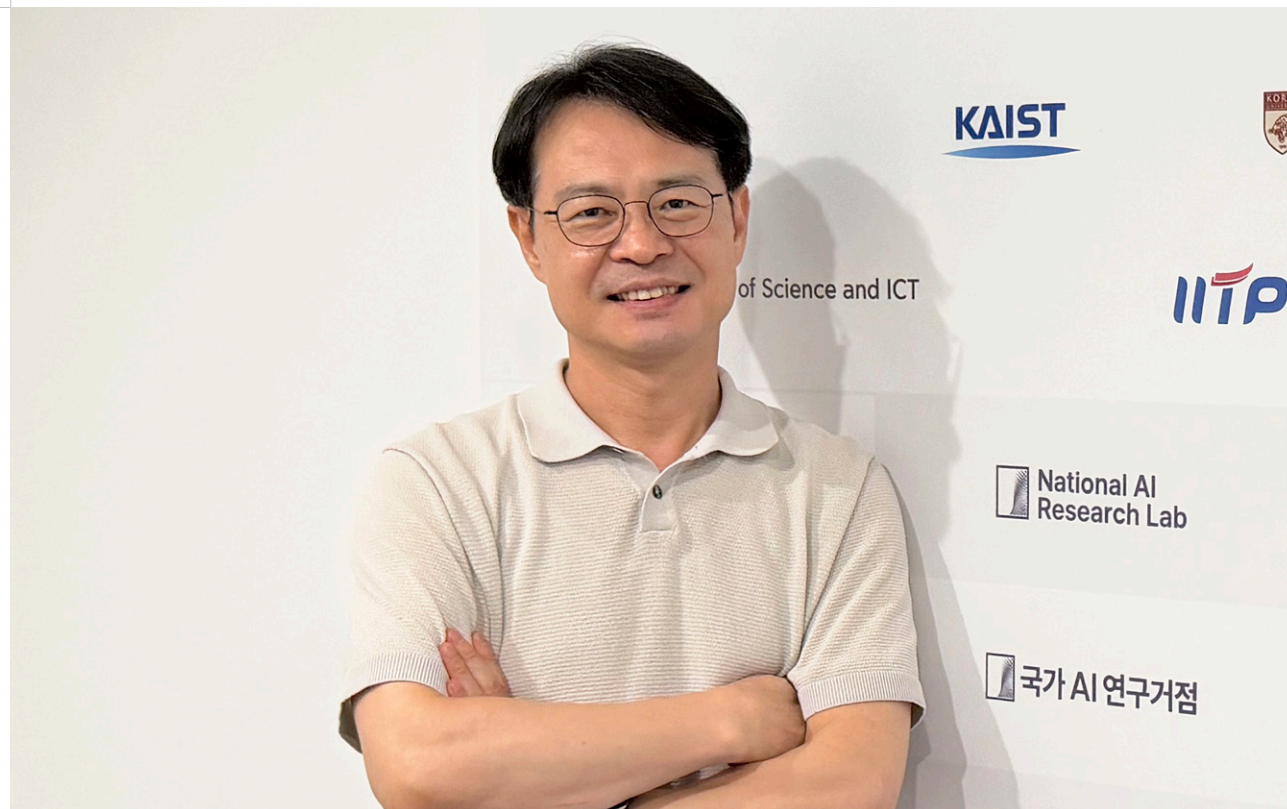
서울에서 열린 ICML 2026,
그리고 국가AI연구거점의 활약

NAIRL Tech Horizon

「글로벌 AI 프론티어 심포지엄 2026」 현장



National AI
Research Lab



NAIRL SPECIAL COLUMN

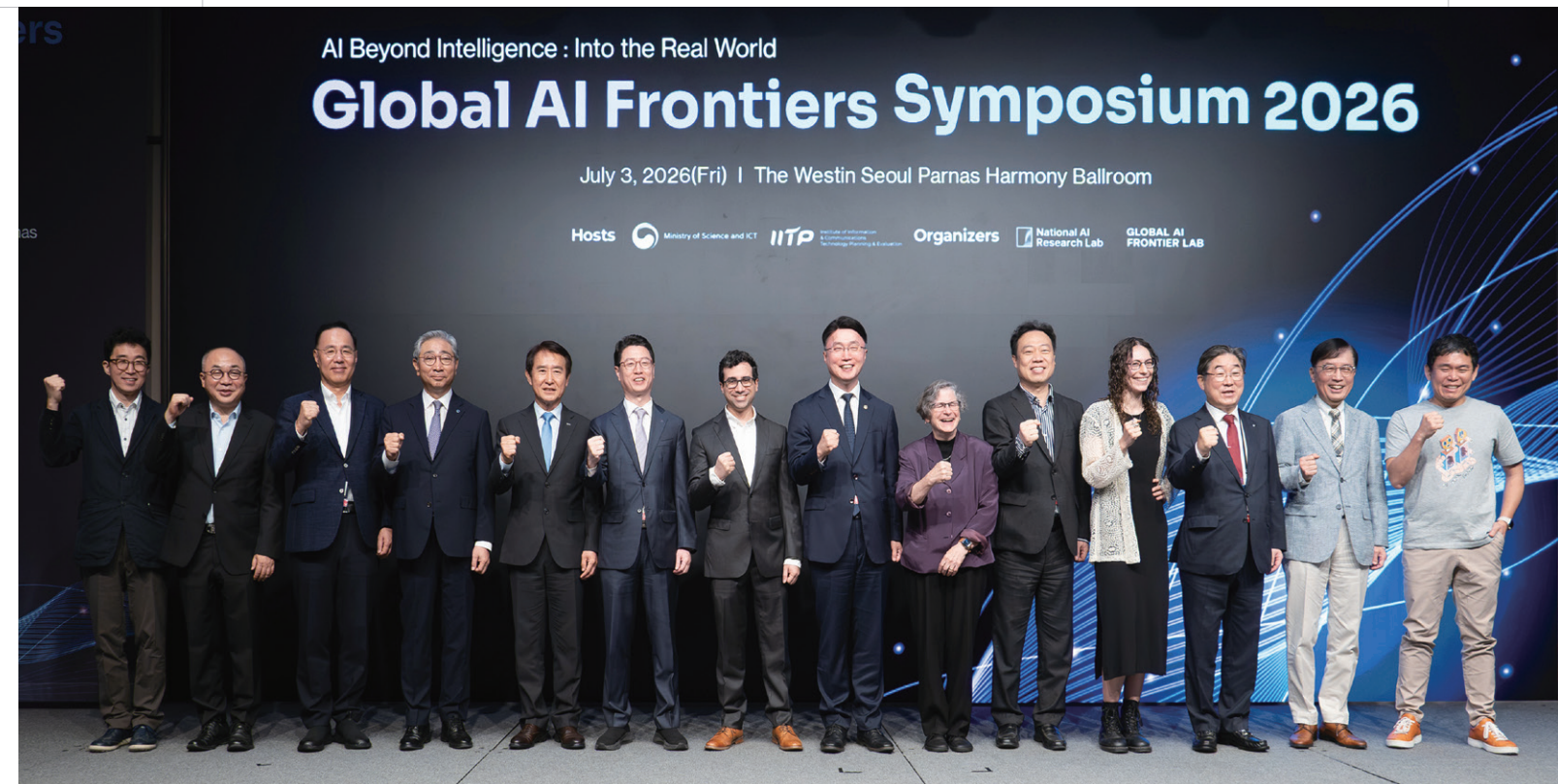
- 04 · 김중철 KT 전문위원, KAIST 겸직교수
중장기 AI 기술 로드맵과 퍼지컬 AI의 미래 전망

NAIRL RESEARCH HIGHLIGHT

- 12 · 서울에서 열린 ICML 2026,
그리고 국가AI연구거점의 활약

NAIRL TECH HORIZON

- 20 · 「글로벌 AI 프론티어 심포지엄 2026」 현장
- 26 · 세 나라의 AI 연구자들, AI의 다음을 논하다
- 30 · 제주에 모인 거점, 다음 걸음을 그리다
- 34 · 「GARL Challenge 2026」,
성황리에 막을 내리다



김종철
KT 전문위원, KAIST 겸직교수

중장기 AI 기술 로드맵과 피지컬 AI의 미래 전망



도입: AI 경쟁의 중심이 이동하고 있다

지난 수년간 AI 기술 경쟁은 모델·데이터·컴퓨팅 규모를 확대하는 스케일링 전략을 중심으로 전개되어 왔다. 대규모 Foundation Model은 방대한 지식을 학습하여 자연어 질의응답과 문서·이미지·영상·코드 생성 능력을 고도화했으며, 최근에는 복잡한 문제를 단계적으로 추론하고 외부 도구를 활용해 실제 과업을 수행하는 수준으로 발전했다.

그러나 주요 모델의 성능이 빠르게 향상되고 일부 영역에서 상향 평준화되면서 ‘잘 대답하는 능력’과 ‘장기 과업을 신뢰성 있게 완수하는 능력’의 간극이 새로운 한계로 부상하고 있다. 이에 AI는 생성 중심에서 추론·계획·실행 중심으로 진화하고 있으며, 멀티모달 AI가 다양한 정보를 통합하고, AI Agent가 Long-term Memory와 World Model을 활용해 과업을 계획·실행하며 행동 결과를 예측하는 자율 시스템으로 확장되고 있다. 이러한 자율성은 디지털 공간을 넘어 로봇·자율주행차·산업설비와 결합하면서, 현실 환경을 인식하고 물리적 상호작용의 결과를 예측해 행동하는 피지컬 AI로 이어지고 있다.

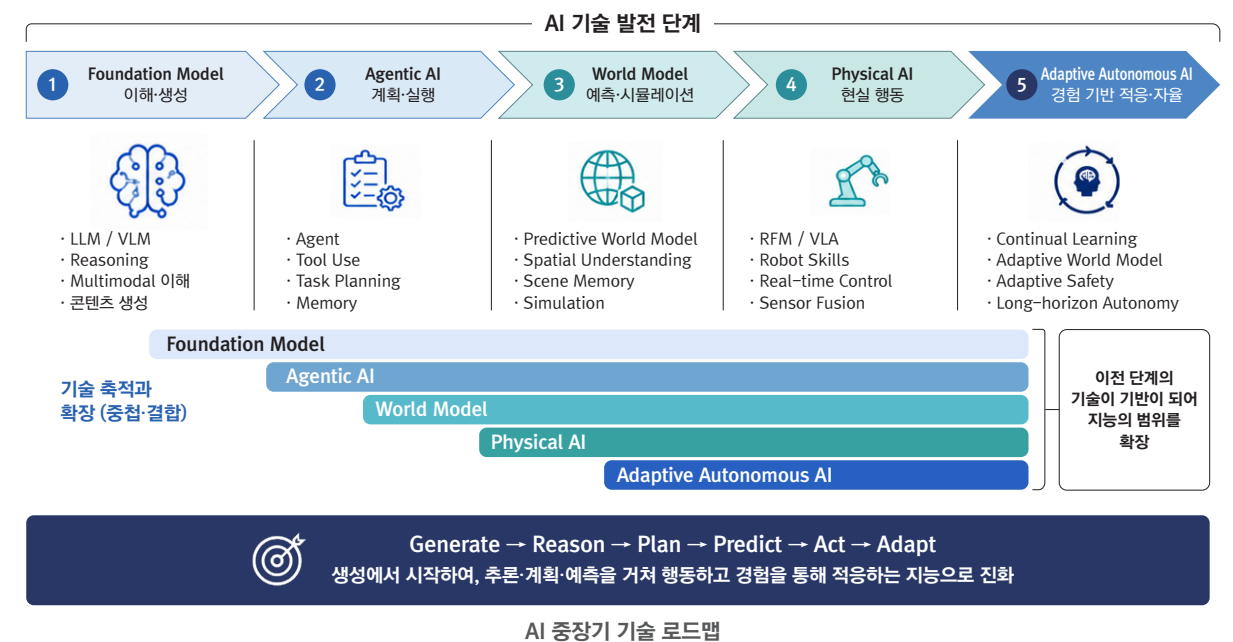
피지컬 AI는 대규모 Foundation Model을 로봇에 탑재하는 것만으로 구현되지 않으며, 인지·기억·예측·계획·행

동의 페루프에 안전을 통합해야 한다. 또한 실제 수행 과정에서 성공·실패 결과와 위험 상황·인간 개입 이력을 수집·평가하고, 검증된 경험만 재학습에 반영함으로써 미학습 환경에 대한 적응력과 안전성을 지속적으로 높여야 한다.

따라서 미래 AI 경쟁의 기준은 모델 규모와 벤치마크 성능을 넘어 장기 과업의 완수율과 현실 환경에서의 적응성·신뢰성·안전성으로 이동할 것이다. AI의 가치는 무엇을 얼마나 잘 생성하는가보다 현실을 얼마나 정확히 이해하고 예측하며 안전하게 행동하는가에 의해 결정될 것이다.

중장기 AI 기술 로드맵: 생성에서 경험으로

앞서 살펴본 경쟁 기준의 변화는 AI 기술 구조의 진화로 이어지고 있다. 중장기적으로 AI는 단일 거대 모델의 성능 확장에 머무르지 않고, Foundation Model을 기반으로 AI Agent를 중심으로 Memory·World Model·Action Model·Safety System을 결합한 자율 시스템으로 발전할 것이다. 기술의 중심도 콘텐츠 생성에서 목표 이해, 과업 계획·실행, 행동 결과 예측, 경험을 통한 지속적 개선으로 이동하고 있다.



발전 단계	핵심 기술	주요 능력
Multimodal Foundation Model	LLM·VLM·Reasoning Model	언어·영상·음성·센서 정보의 통합 이해와 생성
Agentic AI	AI Agent·Tool Use·Long-term Memory·Multi-Agent	목표 분해, 도구 활용, 장기 과업 계획·실행 및 결과 검증
World Model	Predictive·Spatial·Generative World Model·Scene Memory	환경 상태와 공간 구조 표현, 미래 상태 예측 및 행동 대안 시뮬레이션
Physical AI	VLA·Robot Foundation Model·Robot Skills·Real-time Control	인지·추론·예측·계획을 실제 로봇 행동으로 연결
Adaptive Autonomous AI	Adaptive World Model·Continual Learning·Adaptive Safety	실행 경험을 통한 모델 보정, 미학습 환경과 환경 변화에 대한 적응, 실패 복구 및 안전한 지속적 성능 개선

AI 중장기 기술 로드맵

이러한 AI의 발전은 기존 기술이 차례로 대체되는 선형적 전환이 아니라, 각 기술이 중첩·결합되며 지능의 범위를 확장하는 과정이다. 단기에는 추론·계획 능력과 Tool Use·Long-term Memory를 갖춘 Agentic AI가 고도화되고, 중기에는 Spatial Intelligence가 World Model에 결합되면서 현실을 이해하고 행동하는 Physical AI로 확장될 것이다. 장기적으로는 실제·가상 환경에서 축적한 경험을 재학습과 안전성 개선에 연결하는 Adaptive Autonomous AI로 진화할 것이다.

단기: Foundation Model에서 Agentic AI로

단기적으로는 Multimodal Foundation Model의 추론 능력과 AI Agent의 자율적 과업 수행 능력이 빠르게 고도화될 것이다. AI는 텍스트·이미지·음성·영상·센서 정보를 하나의 맥락으로 통합하고, 복잡한 목표를 세부 과업으로 분해한 뒤 외부 데이터와 도구를 활용해 계획·실행하고 결과를 확인하는 방향으로 발전할 것이다. 여기에 Long-term Memory와 Multi-Agent 협업이 결합되면 과거 경험과 업무 맥락을 장기간 유지하면서, 전문 Agent들이 역할을 분담하고 협업하는 자율 업무 시스템으로 확장될 수 있다.

이에 따라 평가 기준도 정적 벤치마크 점수에서 실제 과업 완수율·비용·지연시간·강건성·보안성·안전성과 같

은 시스템 성능으로 이동할 것이다. 특히 장기 과업에서는 단계별 오류가 누적되거나 잘못된 도구와 권한이 사용될 위험이 커지므로, 실행 전 계획 검증과 권한 통제, 실행 중 상태 추적과 이상 감지, 실행 후 결과 검토와 이력 관리, 전 과정의 인간 감독이 필요하다.

결국 단기 Agentic AI의 경쟁력은 모델의 추론 성능뿐 아니라, 복잡한 과업을 얼마나 신뢰성 있고 효율적으로 완수하느냐에 의해 결정될 것이다.

중기: World Model에서 Physical AI로

중기에는 Spatial Intelligence를 포함한 World Model이 고도화되면서 AI의 적용 범위가 현실 세계의 이해·예측·행동으로 확장될 것이다. World Model은 환경의 현재 상태와 공간 구조, 객체 간 관계를 표현하고 특정 행동에 따른 미래 상태의 변화를 예측한다. AI는 이를 바탕으로 여러 행동 대안을 내부적으로 시뮬레이션하고, 목표 달성 가능성과 위험을 비교해 더 적절한 행동을 선택할 수 있다. World Model은 미래 상태를 예측하는 Predictive World Model, 3D·4D 공간 구조와 객체 관계를 표현하는 Spatial World Model, 가상환경과 합성 데이터를 생성하는 Generative World Model로 분화·결합될 것이다. 여기에 Scene Memory가 결합되면 객체의 상태 변화와 과거 행동, 작업 진행 상태를 기억해 장기 계획에 활용할 수 있다.

이러한 예측·공간 지능이 VLA·Robot Foundation Model과 결합되면, AI의 판단은 Robot Skills를 통해 이동·조작·조립·운반과 같은 물리적 행동으로 전환된다. 다만 피지컬 AI는 언어·영상 이해를 넘어 물리적 접촉을 처리하고 균형을 유지하며 관절 제약을 준수하는 동시에, 환경 변화에 실시간으로 대응해야 한다. 따라서 고수준의 인지·추론·계획과 빠른 행동 생성, 실시간 제어를 계층적으로 결합하고 안전 기능을 전 과정에 통합해 불확실한 환경에서도 적응성과 안전성을 함께 확보해야 한다.

이 단계의 핵심 과제는 로봇 행동 데이터의 부족, 로봇별 센서 구성·관절 구조·Action Space의 차이, 시뮬레이션과 현실의 격차를 극복하는 것이다. 이를 위해 인간 시연 데이터 수집, 합성 데이터 생성, 시뮬레이션 학습과 실제 Rollout을 연계하고, Cross-Embodiment 학습을 강화하며 Sim-to-Real 격차를 줄여 다양한 로봇과 환경으로 일반화할 수 있는 기반을 확보해야 한다. 구체적인 기술 구조와 경험학습 체계는 다음 장의 ‘피지컬 AI의 미래’에서 살펴본다.

장기: 경험을 통해 적응하는 Adaptive Autonomous AI로

장기적으로 AI는 인간이 생성한 데이터를 학습하고 모방하는 단계를 넘어, 실제·가상 환경과 지속적으로 상호작용하며 경험을 축적하고, 과업 수행의 자율성을 경험 기반 적응과 지속적 개선으로 확장하는 Adaptive Autonomous AI로 발전할 것이다. 성공·실패 결과와 인간 개입 이력을 활용해 미학습 환경에 적응하고, 장기 과업의 오류를 복구하며, 다음 행동 정책을 개선한다. 최근 AdaJEPa는 실행 중 관측한 상태 전이로 World Model을 보정하고 행동을 재계획하는 ‘Plan-Act-Adapt-Replan’ 페루프를 제시함으로써, Physical AI가 실제 경험을 통해 적응하는 Adaptive Autonomous AI로 발전할 가능성을 보여준다.

환경 인식 → 미래 예측 → 행동 계획 → 안전 검증 → 실행 → 결과 평가 → 경험 선별·검증·학습

다만 경험 기반 학습은 허용된 운용 범위와 안전 정책 안에서 통제되어야 한다. 위험이 임계치를 넘으면 행동을 제한·중단하거나 안전 상태로 복귀하고, 필요하면 인간에게 제어권을 넘겨야 한다. 검증된 경험만 학습에 반영하고 기존 능력과 안전 정책을 보호하는 Safe Continual Learning이 뒷받침되어야 한다.

로드맵의 최종 방향

AI의 중장기 기술 로드맵은 다음의 확장적 발전 경로로 요약된다.

이해하고 생성하는 Foundation Model → 계획하고 실행하는 Agentic AI → 세계를 예측하는 World Model → 현실에서 행동하는 Physical AI → 경험을 통해 안전하게 적응하는 Adaptive Autonomous AI

이 경로의 종착점은 하나의 초거대 모델이 아니라, 예측·행동·학습의 페루프에 안전이 통합된 자율 시스템이다. 다음 장에서는 이러한 전환의 핵심 무대인 피지컬 AI의 기술 구조와 산업 확산 방향을 살펴본다.

피지컬 AI의 미래: 기술적 과제에서 산업 확산까지

앞서 제시한 중기 로드맵을 현실에서 구현하려면 Foundation Model을 로봇에 탑재하는 것만으로는 충분하지 않다. World Model이 환경의 상태와 공간 구조를 표현하고 미래를 예측하며, Scene Memory가 객체의 상태 변화와 작업 이력을 기억해야 한다. Robot Foundation Model이 Robot Brain과 Action Model을 연결하며, Real-time Control이 생성된 행동을 안정적인 로봇 동작으로 변환해야 한다. 여기에 안전 기능이 계획부터 실행·중단·복구까지 전 과정에 통합되어야 한다.

결국 피지컬 AI의 기술적 성패는 개별 모델의 성능보다 인지·기억·예측·계획·행동·제어가 하나의 페루프로 연결되고, 안전이 그 전반에 내재화된 시스템 완성도에 달려 있다.

피지컬 AI를 완성하기 위한 핵심 기술·역량

핵심 기술·역량	주요 내용
World Model	환경의 상태와 3D·4D 공간 구조, 객체 관계 및 물리적 인과관계를 표현하고, 행동에 따른 미래 상태와 위험을 예측
Scene Memory	객체의 상태 변화, 과거 행동과 작업 진행 상태를 기억해 장기 계획과 재계획에 활용
Robot Foundation Model	Robot Brain과 Action Model을 통합하거나 긴밀하게 연결하고, 로봇·환경·과업 간 지식과 행동 능력을 전이
Long-horizon Autonomy	장기 과업의 진행 상태와 누적 오류를 관리하고, 실패 시 재계획·재시도·중단·복구
Real-time Control	행동 명령을 안정적이고 정밀한 로봇 동작으로 변환하고 관절과 구동기를 실시간으로 제어
Adaptive Safety	불확실성과 위험 수준에 따라 행동 범위·자율성·인간 개입 수준을 동적으로 조절

피지컬 AI의 핵심 과제는 고수준의 판단과 빠른 물리적 실행 사이의 간극을 해소하는 것이다. Robot Brain은 목표와 상황을 바탕으로 ‘무엇을 해야 하는가’를 판단하고, Action Model은 이를 실행 가능한 동작으로 변환한다. Real-time Controller는 접촉력을 제어하고 균형을 유지하며 관절 제약을 준수하도록 관절과 구동기를 빠르게 조정한다. 이처럼 서로 다른 시간 주기로 작동하는 Robot Brain, Action Model, Real-time Controller가 계층적으로 연결되어야 안정적인 로봇 행동이 가능하다.

Long-horizon Autonomy는 단일 동작의 성공을 넘어 수분·수시간 동안 과업의 진행 상태와 선행 조건, 자원, 누적 오류를 관리하는 능력이다. 로봇은 진행 상황을 Scene Memory와 World Model에 갱신하고, 실패나 환경 변화가 발생하면 재계획·재시도·안전 중단 중 적절한 대응을 선택해야 한다.

Adaptive Safety는 특정 계층에 한정된 기능이 아니라 계획·행동·제어 전반을 관통하는 설계 원칙이다. 불확실성과 위험이 증가하면 행동 범위와 자율성을 축소하고 인간 개입 수준을 높여야 한다. 실행 중에는 Safety Controller

가 행동을 생성하는 주 경로와 별도의 감시·개입 경로에서 로봇과 환경의 상태를 확인하고, 위험 수준에 따라 동작 제한·안전 정지·안전 상태 복귀 등의 조치를 수행한다.

사전 평가와 시뮬레이션만으로는 실제 환경의 성능과 위험을 완전히 검증하기 어렵다. 따라서 실환경 검증과 운영 중 모니터링을 지속하고, 발견된 실패와 위험징후를 이후의 안전성 개선에 반영해야 한다.

범용 로봇보다 먼저 확산될 산업 지능

피지컬 AI는 완전한 범용 휴머노이드가 등장한 이후에 시작되는 기술이 아니다. 작업 범위와 안전 조건을 비교적 명확히 정의할 수 있는 제조·물류 등의 산업 현장에서 먼저 도입되고, 이후 건설·에너지·국방 분야로 확산되는 과정에서 현장 경험과 Robot Skills가 축적되면서 서비스·돌봄·가정처럼 복잡하고 개방된 환경으로 확장될 가능성이 높다.

아래 표와 같은 단계는 단순한 시간 순서라기보다 환경의 개방성, 과업 다양성, 요구되는 자율성 수준에 따른 확산 경로를 의미한다. 초기 경쟁력은 인간과 유사한 외

확산단계	적용 환경	기술적 특징	인간의 역할
1단계	구조화된 산업	정의된 작업의 반복 수행·제한적 환경 적응	상시 감독·예외 처리
2단계	반구조화된 환경	환경 변화 인식·재계획·복수 작업 수행	원격 감독·필요시 개입
3단계	개방형 생활 환경	미학습 상황 적응·검증된 경험 기반 성능 개선·다목적 행동	목표 설정·최종 통제

형이나 범용 대화 능력보다 정의된 운영설계영역^{ODD}에서 작업을 안정적으로 완수하고 예상하지 못한 변화와 실패에 안전하게 대응하는 능력에서 결정될 것이다.

로봇 제품에서 Physical AI Platform으로

경쟁의 단위도 개별 로봇 제품에서 다양한 로봇과 산업설비로 지능·기술·경험을 전이할 수 있는 Physical AI Platform으로 이동할 것이다. 플랫폼은 목표를 해석하고 과업을 분해·조정하는 Robot Agent, 상황을 인지·추론하고 행동을 계획하는 Robot Brain, 환경의 상태와 미래를 표현·예측하는 World Model, 작업 맥락과 진행 이력을 유지하는 Scene Memory, 행동 계획을 실행 가능한 동작으로 변환하는 Robot Action Model, 생성된 동작을 안정적으로 수행하는 Real-time Control로 구성된다.

Robot Agent → Robot Brain → Robot Action Model → Real-time Control → Robot

*World Model·Scene Memory: Robot Brain의 상태 표현·예측·기억 기반

여기서 World Model과 Scene Memory는 Robot Brain의 인지·추론·계획을 지원하는 상태 표현·예측·기억 기반으로 기능한다. Robot Foundation Model은 이러한 구조에서 Robot Brain과 Robot Action Model을 통합하거나 긴밀하게 연결하고, 로봇·환경·과업 간 지식과 행동 능력을 전이하는 공통 모델로 기능한다.

플랫폼은 환경과 과업의 변화에 따라 계획·행동·제어를 조정할 수 있어야 하며, Adaptive Safety는 목표 해석부터 계획·행동 생성·실행, 실패 시 중단·복구까지 전 과정과 전 계층에 적용되어야 한다. 불확실성과 위험이 증가하면 행동 범위와 자율성을 축소하고 인간 개입 수준을 높여야 한다. 실행 단계에서는 Safety Controller가 기능적으로 분리된 감시·개입 경로로 작동한다. 이러한 안전 기능은 하나의 모델이나 장치에 의존하지 않고, 계획 검증·행동 제한·실시간 안전 제어·비상정지·인간 개입이 중첩되는 다층 구조로 설계되어야 한다.

데이터에서 경험의 플라이휠로

생성형 AI의 핵심 자산이 대규모 문서·이미지·영상이었다면, 피지컬 AI에서는 상태·목표·행동·물리적 결과·위험도·인간 개입 이력이 연결된 경험 데이터가 핵심 자산이 된다.

인간 시연·텔레오퍼레이션 데이터 → 합성 데이터 생성·시뮬레이션 확장 → 학습·검증 → 실제 로봇 Rollout → 결과·위험도 평가 → 경험 선별·검증·재학습

텔레오퍼레이션은 초기 행동 정책을 제공하고, 합성 데이터와 시뮬레이션은 환경·과업·실패 조건을 확장하며, 실제 Rollout은 Sim-to-Real 격차를 보정한다. 이후 실행 결과와 실패·위험 사례를 수집해 검증된 데이터만 학습에 반영하고, 업데이트된 모델도 안전성 검증을 거쳐 현장에 적용해야 데이터가 지속 가능한 경쟁력으로 전환된다.

따라서 미래 경쟁력은 로봇을 얼마나 많이 생산하는가보다 현장 경험을 얼마나 빠르고 안전하게 재사용 가능한 학습 자산으로 전환하는가에 의해 결정된다.

국가와 기업의 대응 전략: 풀스택 경쟁보다 전략적 통합 역량

피지컬 AI의 통합 기술과 경험 플라이휠을 산업 규모로 구현하려면 개별 기업의 기술 개발을 넘어 국가와 산업 생태계 차원의 대응이 필요하다. 피지컬 AI는 Foundation Model, World Model, Robot Foundation Model, 로봇 하드웨어, 반도체·센서, 시뮬레이션, 실시간 제어, 안전·검증 기술이 결합되는 복합 산업이다. 하나의 국가나 기업이 전체 기술 스택에서 경쟁우위를 확보하기는 어렵다.

따라서 산업 경쟁력과 데이터 주권에 직결되는 핵심 기술은 선택적으로 내재화하고, 범용 모델·하드웨어·시뮬레이션 도구 등은 글로벌 생태계와 전략적으로 연계해야 한다. 이를 바탕으로 전체 시스템을 설계·통합·검증·운영하는 역량을 확보하는 것이 중요하다.

우선 확보해야 할 핵심 역량

전략 영역	주요 내용
산업 현장 데이터	제조·물류 등 산업 현장의 상태·목표·행동-결과-위험도-인간 개입 이력을 표준화해 수집·검증·축적
핵심 AI·Robot Skills	World Model·Robot Foundation Model과 산업 특화 Robot Skills를 전략적 중요도에 따라 선택적으로 내재화
학습·검증 인프라	인간 시연·합성 데이터·시뮬레이션·실환경 실증을 연결하는 Sim-to-Real 학습·검증 체계 구축
공통 플랫폼	다양한 로봇·센서·관절·Action Space를 수용하고 모델·기술·경험을 전이하는 Physical AI Platform 구축
안전·인증 체계	운영설계영역(ODD) 기반 위험 평가, Adaptive Safety, 인간 개입, 안전 중단·복구와 실환경 검증·인증 체계 확립
개방형 생태계	산학연 공동연구, 데이터·실증 인프라 공유, 상호운용성 표준과 공동 활용·확산 체계 구축

한국이 선택해야 할 전략

한국은 세계적 수준의 제조업과 반도체·ICT 인프라뿐 아니라 로봇 제조사와 대규모 수요기업을 함께 보유하고 있다. 따라서 범용 Foundation Model의 기술 경쟁력을 지속적으로 높이는 한편, 한국의 강점인 산업 현장을 피지컬 AI의 검증 가능한 경험 데이터와 실증 기반으로 전환하는 전략이 필요하다.

고부가가치 산업 과업 선정 → 인간 시연·행동 데이터 수집 → 합성 데이터 생성·시뮬레이션 학습 → 사전 안전성 검증 → 실제 로봇 실증 → 결과·위험 평가 → 검증된 경험·Robot Skills 축적 → 다른 로봇·현장으로 확산

국가는 공통 데이터 형식, 표준 시뮬레이션 환경, 운영 설계영역^{ODD} 기반 안전성 평가 기준, 공동 실증 인프라와 상호운용성 표준을 구축해야 한다. 기업은 산업별 World Model·Robot Action Model·Robot Skills, 현장 경험 데이터와 Adaptive Safety 기술을 전략적 중요도에 따라 선택적으로 내재화해야 한다. 범용 Foundation Model과 로봇 하드웨어·시뮬레이션 기술은 기술 주권과 사업 경쟁력을 고려해 글로벌 생태계와 전략적으로 연계할 필요가 있다. 자동차·전자·조선·물류처럼 경쟁력 있는 산업을 우선 실증 영역으로 선정하고, 현장에서 검증된 Robot Skills와 경험·안전 데이터셋을 표준화해 다른 로봇과 산업으로 전이해야 한다.

한국의 기회는 세계적 제조 현장을 검증 가능한 경험

과 재사용 가능한 Physical AI 자산으로 전환해 산업 전반으로 확산하는 데 있다.

결론: AI의 다음 10년은 현실에서 결정된다

AI의 중장기 경쟁은 생성 성능을 넘어, 현실을 이해하고 예측하며 안전하게 행동하고 경험을 통해 지속적으로 개선되는 시스템 역량으로 이동할 것이다. 피지컬 AI는 디지털 지능이 산업과 사회의 실질적인 생산성으로 전환되는 핵심 무대다. 모델과 플랫폼, 경험 데이터, 학습·검증·안전 체계는 개별 요소로 머무르지 않고 현실 과업의 안정적인 완수를 중심으로 통합되어야 한다.

이를 관통하는 원칙은 적응성과 안전성의 통합이다. AI는 환경과 과업의 변화에 맞춰 계획과 행동을 조정해야 하며, Adaptive Safety를 통해 불확실성과 위험 수준에 따라 행동 범위와 자율성, 인간 개입 수준을 동적으로 조절해야 한다. 이를 위해 사전 안전성 검증뿐만 아니라 실행 중 위험 감시와 인간 개입, 사고·위험징후·실패 이력을 반영한 지속적인 안전 개선이 함께 이루어져야 한다. 실패를 감지하고 재계획·중단·복구할 수 있을 때 자율성은 신뢰할 수 있는 산업 역량이 된다.

결국 AI의 다음 10년을 결정하는 것은 모델의 크기 자체보다 현실에서 축적하고 검증한 경험을 더 정확한 예측, 더 신뢰할 수 있는 행동, 더 안전한 자율성으로 전환하는 능력 일 것이다. [NAIRL](#)

【 참고문헌 】

1. LeCun, Y. (2022). “A Path Towards Autonomous Machine Intelligence.” OpenReview, Version 0.9.2.
2. Yao, S., et al. (2023). “ReAct: Synergizing Reasoning and Acting in Language Models.” International Conference on Learning Representations (ICLR).
3. Assran, M., et al. (2025). “V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning.” arXiv:2506.09985.
4. Wang, Y., et al. (2026). “AdaJEPA: An Adaptive Latent World Model.” arXiv:2606.32026.
5. Gu, Q., et al. (2023). “ConceptGraphs: Open-Vocabulary 3D Scene Graphs for Perception and Planning.” arXiv:2309.16650.
6. Agarwal, N., et al. (2025). “Cosmos World Foundation Model Platform for Physical AI.” arXiv:2501.03575.
7. Brohan, A., et al. (2023). “RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control.” Conference on Robot Learning (CoRL).
8. O’Neill, A., et al. (2024). “Open X-Embodiment: Robotic Learning Datasets and RT-X Models.” IEEE International Conference on Robotics and Automation (ICRA).
9. Reed, S., et al. (2023). “RoboCat: A Self-Improving Foundation Agent for Robotic Manipulation.” Transactions on Machine Learning Research.
10. Feng, T., et al. (2026). “Self-evolving Embodied AI.” arXiv:2602.04411.
11. Tobin, J., et al. (2017). “Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World.” IROS.
12. Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. National Institute of Standards and Technology.
13. ISO. (2025). ISO 10218-1:2025, Robotics—Safety Requirements—Part 1: Industrial Robots.
14. ISO. (2025). ISO 10218-2:2025, Robotics—Safety Requirements—Part 2: Industrial Robot Applications and Robot Cells.
15. The White House. (2025). America’s AI Action Plan.
16. Bengio, Y., et al. (2026). International AI Safety Report 2026. DSIT 2026/001.
17. 산업통상자원부. (2024). 「제4차 지능형로봇 기본계획(2024~2028)」.
18. 한국지능정보사회진흥원. (2025). 「피지컬 AI 글로벌 동향 및 대응 전략」.
19. 한국산업기술진흥협회. (2026). 「주요국의 피지컬 AI 도입 및 활용 현황」. KITI Purple Paper.
20. 삼일PwC경영연구원. (2026). 「지능형 로봇이 바꾸는 산업 현장: Physical AI가 공장을 움직인다」.

서울에서 열린 ICML 2026, 그리고 국가AI연구거점의 활약



아시아 세 번째, 한국 첫 개최 - ICML 2026

머신러닝 학계 최전선이 서울에 모이다

제43회 국제머신러닝학회 International Conference on Machine Learning, ICML 2026가 지난 7월 6일부터 11일까지 서울 코엑스COEX에서 열렸다. ICML은 NeurIPS, ICLR과 함께 세계 3대 머신러닝 학회로 꼽히며, 이 중에서도 가장 오랜 역사를 지니고 이론적 기초와 알고리즘, 수학적 증명을 중시하는 학회로 평가받는다. 1980년대 이후 40여 년의 역사에서 ICML이 아시아에서 개최된 것은 이스라엘, 중국(베이징)에 이은 세 번째이며, 이번 서울 개최는 한국의 AI 연구 위상이 그만큼 높아졌다는 것을 보여주는 상징적인 사건으로 받아들여진다.

이번 학회는 미국 일리노이대 어배너-섐페인 캠퍼스 UIUC의 통 장Tong Zhang 교수가 조직위원장을 맡아 진행했다. 행사는 7월 6일 엑스포·튜토리얼 데일로 시작해 7일부터 9일까지 메인 컨퍼런스, 10일과 11일에는 44개의 정식 워크숍과 4개의 어피니티 워크숍이 병행 진행됐다. 전 세계 AI 기업 및 학계 관계자 2만 명 이상이 서울을 찾았으며, 이번 학회에는 역대 최다인 2만 3,918편의 논문이 투고돼 전년(1만 2,107편)의 두 배에 가까운 규모를 기록했다. 채택률은 약 26.6%(6,352편)였고, 이 중 전체 투고의 2.2%에 해당하는 536편이 스포트라이트 Spotlight 논문으로, 0.7%인 168편이 구두 발표Oral 논문으로 선정됐다. 한편 LLM 사용 정책을 위반한 것으로 적발된 497편(약 2%)이 심사 중 데스크 리젝desk-reject된 점도 올해 학회의 특징적인 이슈로 언급됐다.

학계에서는 MIT, 스탠퍼드대, UC 버클리, 카네기멜런대 등 미국 주요 대학과 캐나다의 토론토대, 밀라Mila연구소, 영국의 옥스퍼드대, 케임브리지대, 프랑스 인리아Inria 등이 대거 참여했다. 기업 부문에서는 구글과 구글 딥마인드가 130편 이상의 논문을 발표했고, 애플, 아마존, 오픈AI, 메타, 마이크로소프트, 미스트랄 등 글로벌 빅테크가 총출동해 부스를 열었다.

국내에서도 KAIST(120편)와 서울대(77편)를 비롯해

총 357편의 논문이 채택됐으며, 연세대와 고려대, 포항공대POSTECH, 울산과학기술원UNIST, 중앙대 등 주요 대학 연구진이 발표에 나섰다. 산업계에서는 삼성전자, 네이버, LG AI연구원, 카카오, 크래프톤, SK텔레콤, 업스테이지, NC AI 등 기업과 디노티시아, 노타, 하이퍼엑셀, 베슬AI, 프렌들리AI 등 스타트업이 부스를 마련해 기술을 소개하거나 인재 채용에 나섰다.

올해의 화두, '에이전틱 AI와 그 안전성'

올해 ICML을 관통한 핵심 주제는 스스로 목표를 수행하는 자율 시스템, 즉 '에이전틱 AI(Agentic AI)'와 그 신뢰성·안전성이었다. 에이전틱 AI는 목표를 받아 스스로 다음 행동을 결정하고 도구나 API를 호출하며, 중간 단계마다 사람의 승인 없이 작업을 이어가는 자율적 행동 루프를 가진 시스템을 뜻한다.

에이전틱 AI가 올해의 의제로 떠오른 데에는 ICML이 전통적으로 최적화 이론, 통계적 학습 이론, 강화학습 비중이 높은 학회라는 점도 작용했다. 이 영역들이 자율 에이전트의 신뢰성·안전성 문제와 직접 맞닿아 있어, 올해 의제로 자연스럽게 부상했다는 평가다. 생성형 AI, 거대언어모델LLM 관련 주제가 전체 투고에서 가장 큰 비중을 차지했고, 그중에서도 에이전트 관련 주제가 두드러졌다.

열기는 워크숍에서 특히 뚜렷했다. 워크숍 제안서 중 60건 이상의 제목에 '에이전틱 AI'라는 표현이 포함돼 워크숍 의장단조차 '놀라운 집중도'라고 평가할 정도였다. 특히 235편 이상의 논문이 투고되고 800명 이상이 참석한 'Agents in the Wild' 워크숍은 에이전트 간 통신, 최소한의 인간 감독하에 작동하는 에이전트 커뮤니티, 멀티에이전트 협업 등에서 발생하는 위험을 다뤘다.

6인의 초청 강연자, 에이전트 시대의 질문을 던지다

이번 학회의 초청 강연Invited Talk에는 6인의 연사가 나서 머신러닝 이론, AI 안전·윤리, 경제·정책, 계산생물학, 자연어처리, 인간-컴퓨터 상호작용 등 폭넓은 분야를 다루었다. 세션은 총 7개로 편성되었는데, 이 중 하나는 청중이 물리면서 추가로 마련된 오버플로 세션이었다. 이

들의 강연은 각기 다른 분야에서 출발했지만 공통적으로 ‘에이전틱 AI를 어떻게 신뢰할 수 있게 만들 것인가’라는 질문으로 모였다.

메타 FAIR 출신이자 홍콩과기대 소속의 파스칼 펑 Pascale Fung은 물리 세계를 이해하는 ‘월드 모델링’ 기반 에이전트로의 전환을 주장했고, 스탠퍼드 경영대학원 소속의 기술경제학자 수전 애시 Susan Athey는 트랜스포머 기반 인과추론 방법론을 소개했다. 강화학습 이론가 삼 카케이 드 Sham Kakade는 복잡한 학습률·배치 크기 스케줄링이 단순한 이차함수 모델로 설명될 수 있음을 보이며 ‘SeeSaw’ 스케줄러와 ‘Horizon-Free Pretraining’ 기법을 제시했고, 제넨텍 리서치 총괄 겸 EVP 아비브 레게브 Aviv Regev는 실험 데이터와 AI 모델이 서로를 개선해 나가는 ‘랩 인 더 루프 Lab-in-the-Loop’ 구조로 신약 개발 사례를 소개했다. 구글 딥마인드의 리서치 리드 베레나 리저 Verena Rieser는 경직된 가드레일을 넘어 에이전트가 스스로 원칙을 해석·적용하는 ‘규범적 정렬’이라는 패러다임을 제안했으며, 프린스턴대의 아빈드 나라야난 Arvind Narayanan은 AI를 자동화가 아닌 증강 기술로 보는 관점에서 인간 역할의 이동을 전망해 오버플로 세션까지 마련될 정도로 큰 반향을 얻었다.

주목할 만한 구두 발표 연구

168편의 구두 발표는 올해 프로그램위원회가 가장 의미 있다고 판단한 연구를 가장 뚜렷하게 보여주는 목록이다. 게놈 모델링부터 게임이론, 확률밀도 추정, 에이전트 안전성까지 이어지는 연구들은 ICML이 여전히 이론적 엄밀성과 실용적 효율성을 함께 요구하는 학회임을 보여준다.

먼저 UIUC의 사그니크 무케르지 Sagnik Mukherjee는 ‘Do We Need Adam?’이란 제목의 발표에서 LLM 강화학습 미세조정 단계에서는 AdamW의 장점이 크지 않다는 점을 규명하고, 메모리 효율이 높은 SGD만으로도 전체 파라미터의 0.02% 미만만 갱신해 동등하거나 더 나은 성능을 내며 기존 LLM 훈련 관행에 의문을 제기했다. 캐나다 토론토대의 아르나브 샤 Arnav Shah가 공개한 dnaHNet은 서버워드 토큰나이저와 뉴클레오타이드 단위 모델

사이의 오랜 딜레마를 해결한 게놈 파운데이션 모델로, 동적 청킹 dynamic chunking을 통해 토큰나이저 없이도 3배 이상 빠른 추론 속도와 우수한 제로샷 성능을 보였다. 일본의 사이버에이전트 AI 랩의 아베 겐시 Kenshi Abe는 ‘Asymmetric Perturbation in Solving Bilinear Saddle-Point Optimization’이라는 제목의 발표를 통해 미니맥스·게임이론 문제에서 한쪽 플레이어의 페이오프 함수만 교란시키는 기법을 제시하고, 기존 대칭적 교란 방식보다 빠른 선형 수렴을 이론적으로 증명했다.

미국 워싱턴대 바실리 일린 Vasily Il'in은 ‘DiScoFormer’란 트랜스포머 모델을 제시하며 셀프어텐션이 커널 밀도 추정을 정확히 복원할 수 있음을 보이고, 재학습 없이 어떤 분포에도 적용 가능한 등변 트랜스포머로 확률밀도·점수 추정 문제를 해결했다. 오픈AI의 멜로디 관 Melody Guan은 ‘Monitoring Monitorability’라는 제목의 발표에서 에이전틱 AI 안전성과 직결된 연구를 소개하며 사고사슬을 관찰해 AI 오작동을 감지하는 ‘CoT 모니터링’의 취약성을 체계적으로 평가하고 안전성 평가 체계 설계에 시사점을 던졌다. 오스트리아 잘츠부르크 응용과학대의 슈테판 후버 Stefan Huber는 ‘Chebyshev Policies and the Mountain Car Problem’에 대한 발표를 통해 강화학습의 고전 벤치마크인 ‘마운틴카’ 문제를 36년 만에 해석적으로 완전히 풀어내며, 체비셰프 정책으로 파라미터 수를 277배 줄이면서도 성능은 오히려 개선했다.

학술의 장을 넘어선 인재 확보 전쟁

ICML은 학술 행사이자 동시에 빅테크의 기술 경쟁력 과시와 인재 채용의 장이기도 했다. 팀네이버는 12편의 논문(그중 1편은 상위 2.2% 스포트라이트)을 발표하며 국내 기업 중 가장 큰 부스를 열었고, LG AI연구원은 14편의 논문과 이를 산업에 적용한 실증 사례(신소재 발굴 플랫폼 ‘엑사원 디스커버리’ 등)를 함께 전시했다.

AI 양자화 기업 노타와 AI 인프라 스타트업 프렌들리시, 베슬AI는 기술 소개보다 채용에 무게를 둔 부스를 마련했으며, 추론 전용 가속기 LPU를 앞세운 하이퍼엑셀은 신



ICML 2026에 등록한 현장 참석자 수는 2만 456명으로 집계됐다. 이는 지난 10년간 열린 ICML 학회 중 최대 규모다. 사진은 포스터 발표 현장에 모인 국내외 참석자들.

제품 ‘베르다 500’을 처음으로 공개했다. 채용 공고를 붙여 두는 행사장 입구의 잡 보드 job board는 이제 AI 학회에서도 하나의 상징적인 문화로 자리 잡았다는 평가다.

학술 프로그램이 끝난 뒤에도 교류는 도심 한복판으로 이어졌다. 코엑스에서 차량으로 약 5분 거리인 서울 강남구 삼성동의 복합문화공간 ‘메자닌 삼성 MEZZANINE SAMSUNG’에서 국내외 참가자 약 300명이 모여 애프터파티를 열었다. 지정된 좌석 없이 참가자들이 자유롭게 이동하며 대화를 나누는 스탠딩 네트워킹 형태로 운영됐는데, 참가자들은 공식 세션에서 다 나누지 못한 이야기를 이어가며 공동 연구나 기업 협력 가능성을 논의했다.

국내 대학·연구기관은 세계적인 연구자들과 국제 공동연구를 모색할 기회를 얻었고, 국내 AI 기업과 스타트업 역시 해외 파트너를 발굴하는 계기로 삼았다는 평가다. 공식 프로그램에서 형성된 교류가 민간 공간으로 확장돼, 하나의 네트워크 플랫폼으로 기능한 셈이다. ICML 2026이 단순한 학술 행사를 넘어 서울의 MICE 산업 경쟁력을 보여준 사례로도 언급되는 이유다.

국가AI연구거점, 한국 AI 연구 역량을 보여주다

NAIRL, ICML 2026에서 논문 30편 발표

국가AI연구거점 NAIRL 참여 연구진은 이번 ICML 2026에서 총 30편의 논문을 발표했다. 학회가 서울에서 열린 만

큼, NAIRL 연구진이 전 세계 AI 연구 커뮤니티를 안방에서 맞이하며 한국 AI 연구의 역량을 직접 보여주는 자리였다.

이번 학회에는 NAIRL 소속 교수 19명과 1저자 학생 연구원 40여 명이 참여했으며, 기초 머신러닝 이론부터 생성 모델, 강화학습, AI 안전에 이르는 폭넓은 연구의 제를 다뤘다. 기관별로는 KAIST에서 8명의 교수가 이끈 12편으로 가장 많은 논문을 발표했고, 연세대는 5명의 교수를 통해 11편으로 그 뒤를 이었다. POSTECH은 5명의 교수가 7편을, 고려대는 2명의 교수가 2편을 발표하며 컨소시엄의 성과를 함께 완성했다(단, 공동연구 논문은 각 기관에 중복 계산).

특히 이번 성과에는 ICML 포지션 Position 트랙에 채택된 논문 2편이 포함돼 눈길을 끌었다. 하나는 ‘머신 언러닝 machine unlearning’ 개념이 LLM 연구에서 남용되고 있다는 문제를 제기했고, 다른 하나는 AI에 대한 과도한 의존이 낳는 ‘락인 Lock-In’ 현상에 대비해야 한다는 주장을 담았다. 이는 NAIRL 연구진이 기술 개발을 넘어 AI 연구 공동체의 방향성 논의에도 목소리를 내고 있음을 보여준다. 이번 결과는 KAIST, 연세대, 고려대, POSTECH 4개 대학이 기관 간 경계를 넘어 협력해 온 ‘원팀 One Team’ 체제의 결실로 평가된다.

NAIRL 연구는 주제에 따라 3개 세부과제로 나누어 운영된다. 1세부는 뉴럴 스케일링 법칙 초월 연구, 2세부는 로봇 파운데이션 모델 연구, 3세부는 고차원 생성형 파운데이션 모델 연구를 중점으로 하고 있다. 세부별 발표 논문은 다음과 같다.



국가AI연구거점에서는 김기웅 센터장을 비롯한 많은 인원이 ICML 2026에 참가했다.

1세부 논문, 언어-비전 파운데이션 모델의 효율과 양자화

1세부 연구진은 LLM과 비전-언어모델의 개인화, 추론 효율화, 양자화, 안전성 등 폭넓은 주제에서 16편의 논문을 발표했다.

논문명	저자(소속)	개요
Persona2Web: Benchmarking Personalized Web Agents for Contextual Reasoning with User History	김세린, 이상암, 이동하(연세대)	사용자가 의도를 명시적으로 밝히지 않아도 과거 이력을 바탕으로 모호한 요청을 해석하는 개인화 웹 에이전트를 평가하기 위한 최초의 벤치마크. 실제 열린 웹 환경에서 에이전트의 추론 궤적까지 정밀하게 평가한다.
BESPOKE: Benchmark for Search-Augmented Large Language Model Personalization via Diagnostic Feedback	김현서, 이상암, 서광욱, 이동하(연세대)	검색 결합형 LLM의 개인화 성능을 평가하기 위해 실제 사용자의 대화-검색 이력을 수집하고, 세밀한 선호도 점수와 진단적 피드백을 함께 제공하는 벤치마크.
Dependency-Aware Parallel Decoding via Attention for Diffusion LLMs (DAPD)	김범준, 전동재, 전문규, 노알버트(연세대)	디퓨전 언어모델의 병렬 디코딩 시 셸프어텐션으로 토큰 간 의존성 그래프를 구성하고, 독립집합을 선택해 동시에 여러 토큰을 안전하게 복원하는 학습 없는(training-free) 디코딩 기법.
Preserve-Then-Quantize: Balancing Rank Budgets for Quantization Error Reconstruction in LLMs	조윤준, 전동재, 김소은, 전문규, 노알버트(연세대)	양자화 시 가중치의 저랭크 구조가 훼손되는 문제를 해결하기 위해, 지배적인 특이값 부분공간을 먼저 보존한 뒤 나머지 랭크로 오차를 복원하는 랭크 배분 프레임워크(SRR). 2비트 양자화 파인튜닝(QPEFT)에서 평균 5.9%p 성능 개선을 보였다.
dgMARK: Decoding-Guided Watermarking for Diffusion Language Models	홍표민, 노알버트(연세대)	디퓨전 언어모델은 토큰을 임의 순서로 생성할 수 있다는 특성에 착안해, 언마스킹 순서 자체를 워터마크 채널로 활용하는 디코딩 기반 워터마킹 기법. 삽입-삭제-치환 등 후편집 공격에도 강한 탐지력을 보였다.
Position: The Term Machine Unlearning Is Overused in LLMs	윤상연, 전예찬, 노알버트(연세대)	현재 '머신 언러닝'이라는 용어가 거부(refusal), 지식 삭제, 표적 억제 등 서로 다른 목표를 가진 작업에 뒤섞여 쓰이고 있다고 지적하며, 데이터셋 단위의 삭제(retraining과 구분되지 않는 수준의 완전한 영향 제거)에만 이 용어를 한정해 사용해야 한다고 주장하는 포지션 논문.

논문명	저자(소속)	개요
Real-Time Visual Attribution Streaming in Thinking Model	강세일, 한우정, 김준혁, 김진영, 김영은, 황성재(연세대)	멀티모달 사고형(Thinking) 모델이 이미지의 어느 부분에 근거해 추론했는지 실시간으로 보여주는 프레임워크. 반복적인 역전파 없이도 기존 인과적 분석 기법에 준하는 신뢰도를 실시간 스트리밍으로 제공한다. 메인 트랙(일반 논문) 기준 상위 2.2% 스포트라이트 논문 선정 및 SCALE 워크숍 최우수논문상(구두 발표, 상위 5.8%) 수상.
Physics in 2-Steps: Locking Motion Priors Before Visual Refinement Erases Them	한우정, 강세일, 전영준, Min-Hung Chen, Fu-En Yang, 황성재(연세대)	이미지 기반 비디오 생성 모델에서 단 2스텝만 거친 결과가 오히려 50스텝을 모두 거친 결과보다 물리적으로 더 그럴듯한 움직임을 보인다는 역설적 현상을 규명하고, 노이즈 제거 과정에서 사라지는 '위상(phase)' 정보를 초기 단계에서 추출해 고정하는 학습 없는 프레임워크(PhaseLock)를 제안했다.
Mitigating Mask Prior Drift and Positional Attention Collapse in Large Diffusion Vision-Language Models	홍수정, 윤찬웅, 황성재(연세대)	대형 디퓨전 비전-언어모델이 장문 생성 시 반복 생성 및 시각적 근거 상실을 겪는 원인을, 마스크 토큰이 공유된 방향으로 표류하는 현상과 위치 편향의 불일치로 규명하고, 추가 학습 없이 이를 완화하는 기법을 제시했다.
Speculative Coupled Decoding for Training-Free Lossless Acceleration of Autoregressive Visual Generation	소준혁, 국현호, 장채연, 박은혁(POSTECH)	자기회귀 방식 시각 생성 모델의 추론 속도를 높이기 위해, 초안 토큰 샘플링의 불안정성을 정보이론적으로 안정화하는 '커플링' 기법을 제안. 단 한 줄의 코드 수정만으로 이미지 생성은 최대 4.2배, 비디오 생성은 최대 13.6배 가속하면서도 품질 손실이 없다.
Adversarial Robustness of Implicit Neural Representation-Based Classifiers	김자영, 이국진, 박노성, 홍상현(KAIST)	데이터를 신경망 함수로 표현하는 암묵적 신경 표현(INR) 기반 분류기의 적대적 공격 취약성을 최초로 체계적으로 분석. INR 생성 과정 자체가 최적화 루프를 포함해 기존 공격 기법을 무력화한다는 점에 착안, 이를 대신할 대리(surrogate) 공격 모델을 설계했다.
DistMatch: Adaptive Binning via Distribution Matching for Robust Sequential Conformal Prediction	Enver Menadjiev, 성지현, 여지수, 최재식(KAIST)	시계열 데이터의 순차적 컨포멀 예측(conformal prediction)에서 잔차의 교환 가능성 가정이 자주 깨지는 문제를, 콜모고로프-스미르노프 통계량으로 이진트리를 재귀적으로 분할해 지역적으로 교환 가능한 구간을 만들어 해결하는 적응형 구간화 기법.
Breaking the Lock-in: Diversifying Text-to-Image Generation via Representation Modulation (DAVE)	권다희, 이하은, 최재식(KAIST)	대규모 트랜스포머 기반 텍스트-이미지 생성 모델이 동일 프롬프트에서 서로 비슷한 이미지만 반복 생성하는 원인을, 생성 초기 여러 시드의 중간 특징이 빠르게 하나로 수렴하는 현상(DC 성분의 조기 고정)으로 규명하고, 이를 선택적으로 완화해 다양성을 끌어올리는 학습 없는 기법을 제안했다.
Making Models Unmergeable via Scaling-Sensitive Loss Landscape	장민우, 김호영, 구자빈, 옥정슬(POSTECH)	파인튜닝된 모델이 허가 없이 다른 모델과 병합(merge)돼 무단 활용되는 것을 막기 위해, 손실 지형을 스케일링에 민감하게 왜곡시켜 모델 병합 시 성능이 크게 저하되도록 하는 방어 기법(TRAP)을 제안했다.
Over-Alignment vs Over-Fitting: The Role of Feature Learning Strength in Generalization	염태선, 하태혁, 이재호(POSTECH)	모델 출력 스케일의 역수로 정의되는 '특징 학습 강도(FLS)'가 너무 크면 과잉 정렬(over-alignment)로, 너무 작으면 과적합(over-fitting)으로 일반화 성능이 떨어진다는 점을 규명하고, 두 극단 사이에 존재하는 '최적의 FLS'를 이론적으로 설명했다.
TABX: A High-Throughput Sandbox Battle Simulator for Multi-Agent Reinforcement Learning	이하영, 오준혁, 이병준(고려대)	협력형 멀티에이전트 강화학습(MARL) 연구를 위한 JAX 기반의 고속재구성 가능한 시뮬레이션 샌드박스. GPU 가속을 통한 대규모 병렬화로 다양한 난이도의 팀 전투 시나리오를 유연하게 설계할 수 있게 했다.

2세부 논문, 에이전트 안전과 생성모델의 신뢰성 제시

2세부 연구진은 8편의 논문을 통해 임바디드 에이전트 Embodied Agent의 안전, 역문제 inverse problem 해결, 코드

생성의 신뢰성, 안전한 이미지 생성, 강화학습 등을 다뤘다.

논문명	저자(소속)	개요
EMBGuard: Constructing Hazard-Aware Guardrails for Safe Planning in Embodied Agents	최동욱, 권태윤 외, 여진영(연세대)	멀티모달 LLM 기반 임바디드 에이전트가 현실 환경에서 마주치는 물리적 위험을 명시적으로 판단하지 못하는 문제를 해결하기 위한 최초의 MLLM 기반 안전 가드레일. 시각 정보와 행동을 함께 평가해 위험을 자연어로 설명하며, 소형 모델(2B~4B)로도 GPT-5.1급 상용 모델과 경쟁력 있는 성능을 보였다.
Measurement-Consistent Langevin Corrector for Stabilizing Latent Diffusion Inverse Problem Solvers (MCLC)	이효석, 임소희, 차은주, 오태현(KAIST)	의료·위성 영상 등에서 손상된 영상을 복원하는 잠재 디퓨전 기반 역문제 해결기가 겪는 불안정성을, 해법의 동역학과 실제 확산 동역학 간의 격차로 규명하고, 측정값과 일치하는 랑주병 업데이트로 이를 안정화하는 플러그인 모듈을 제안했다.
Zero-Shot Rankability: Revealing Latent Ordinal Structure in Multimodal Large Language Models via Language	남현우, 문예빈, 임소희, 권병기, 오태현(KAIST/POSTECH)	멀티모달 LLM이 별도 학습 없이도 ‘작다-크다’, ‘어둡다-밝다’와 같은 순서 관계를 텍스트 프롬프트만으로 포착할 수 있음을 규명. 기존 CLIP 같은 비전-언어모델이 61% 수준이던 성능을, 텍스트 기반 순위축만으로 지도학습 대비 90% 수준까지 끌어올렸다.
A Language-Guided Bayesian Optimization for Efficient LoRA Hyperparameter Search	백성은, 이정목, 김성빈, 오태현 (POSTECH/KAIST)	LoRA 파인튜닝의 하이퍼파라미터를 찾으려면 보통 수만 개 조합을 탐색해야 하는데, 사전학습된 LLM이 각 설정값의 역할을 자연어로 설명하게 해 탐색을 유망한 후보로 유도하는 베이지안 최적화 기법. 약 30회 탐색만으로 표준 탐색(약 4만 5,000개 조합) 대비 20% 이상 성능을 끌어올렸다.
Towards Functional Correctness of Large Code Models with Selective Generation	정재우, 김태수, 박상돈 (POSTECH)	코드 생성 모델의 환각(hallucination) 문제를 해결하기 위해, 동적 코드 분석으로 단위 테스트를 자동 생성하고, 이를 근거로 불확실한 생성 결과에 대해서는 답변을 보류하는 선택적 생성기(FuzzEval)를 제안했다.
Concept Removal Guidance: Evidence-Calibrated Negative Guidance for Safe Diffusion Sampling (CRG)	최윤석, 오채영, 최현준, 서석인, 김기웅(KAIST)	적대적 프롬프트로 유해 콘텐츠를 유도하는 공격에 대응하기 위해, 매 디노이징 단계에서 원치 않는 개념의 존재 정도를 모델의 노이즈 예측값에서 직접 추정하고, 이를 근거로 네거티브 가이드نس 강도를 자동으로 조절하는 학습 없는 기법. 스포트라이트 논문으로 선정됐다.
PS-PPO: Prefix-Sampling PPO for Critic-Free RLHF	황두환, 김기웅(KAIST)	비평가(critic) 없이 학습하는 RLHF 기법이 긴 추론 경로에서도 매 롤아웃마다 전체 궤적을 업데이트해야 하는 비효율을 겪는다는 점에 착안, 중간 지점(prefix)까지만 역전파하면서도 편향 없는 근사가 되도록 중요도 가중치를 적용한 계산 효율적 PPO 변형.
Constrained Multi-Objective Reinforcement Learning with Max-Min Criterion	박기승, 남현영, 변우현, Amir Leshem, 성영철(KAIST)	여러 목표 간 공정성을 추구하는 max-min 방식의 다목적 강화학습에 제약조건 만족을 함께 결합한 프레임워크. 건물 열 관리, 다목적 보행 제어, 온실가스 저감형 교통 관리 등 실제 응용에서 공정성과 제약 만족을 동시에 달성함을 보였다.

3세부 논문, 생성확산 모델의 구조와 표현 다뤄

3세부 연구진은 6편의 논문을 발표하며 확산 플로

우 기반 생성모델의 이론적 구조와 샘플링 품질을 다뤘다.

논문명	저자(소속)	개요
Contrastive Flow Map Matching	Junyu Zhang, Daochang Liu 외, 박은병(연세대)	플로우 맵 매칭(FMM) 학습이 실제 샘플링 과정과 어긋나는 문제를, 역방향 KL 발산에 대한 이론적 상한을 근거로 학습과 샘플링을 명시적으로 일치시키는 프레임워크(CFMM).
Routing by Reaching: Composition of Pre-trained GFlowNets for Multi-Objective Generation	윤석원, 최영빈, 조승혁, 이승범, 박문정, 김동우(POSTECH)	여러 개의 상충하는 목표를 만족시키는 후보를 탐색하는 GFlowNet을, 새로운 목표 조합이 생길 때마다 재학습하지 않고 추론 시점에 사전학습된 모델들을 조합해 즉시 적응시키는 학습 없는 프레임워크. 선형 가중합은 이론적으로 정확히 복원되며, 비선형 결합에 대해서도 근사 품질을 정량화했다.
TAG: Tangential Amplifying Guidance for Hallucination-Resistant Diffusion Sampling	조현민, 안동훈, 홍수성, 김지은, 김승룡, 진경환(KAIST)	디퓨전 모델의 이미지 생성 시 발생하는 의미적 불일치(환각)를, 별도의 외부 신호나 구조 변경 없이 샘플링 궤적 자체의 접선(tangential) 성분을 증폭시켜 데이터 분포의 고밀도 영역으로 유도해 줄이는 학습 없는 가이드نس 기법.
Lookahead Unmasking Elicits Accurate Decoding in Diffusion Language Models (LookUM)	이상현, 김승룡, 박종호, 박동민(KAIST)	마스크 디퓨전 언어모델의 디코딩 순서가 성능을 좌우한다는 점에 착안, 신뢰도 기반의 근시안적 순서 선택 대신 여러 언마스킹 경로를 미리 생성하고 불확실성을 계산해 오류가 전파되지 않는 경로를 선택하는 디코딩 기법.
Deep Forcing: Training-Free Long Video Generation with Deep Sink and Participative Compression	이정, 장우석, 조현빈, 남지수, 윤희지, 김승룡(KAIST)	실시간 스트리밍이 가능한 자기회귀 비디오 디퓨전 모델이 장시간 생성 시 반복·표류·움직임 둔화를 겪는 문제를 해결. 어텐션 싱크 절반을 유지하며 시간 위상을 재정렬하는 ‘Deep Sink’와 최근 프레임에 실제로 기여하는 토큰만 남기는 ‘Participative Compression’을 결합해, 학습 없이도 5초짜리 영상을 60초 이상으로(12배 이상) 확장했다.
Position: AI Lock-In Is in Progress, and We Must Be Prepared	김재호, 이석현, 이지은, 이창희(고려대)	AI 안전 연구가 기술적 정렬과 사회적 영향 규제에 주로 집중해 온 반면, AI에 대한 과도한 의존이 인간의 숙련도를 저하시키고 독립적 기능 수행 능력을 약화시키며, AI 시스템이 작동하지 않거나 오염됐을 때 발생할 시스템적 취약성(‘AI 락인’)은 상대적으로 덜 주목받았다고 지적하는 포지션 논문. 포지션 논문 트랙 기준 상위 5.3% 스포트라이트로 선정됐다.

논문에 담긴 기술 흐름으로 본 종합 평가

30편의 논문을 관통하는 흐름은 크게 세 갈래로 요약된다. 첫째, 효율성의 추구다. 양자화(Preserve-Then-Quantize), 파인튜닝 탐색 자동화(Language-Guided BO), 추론 가속(Speculative Coupled Decoding, DAPD), 장시간 생성(Deep Forcing) 등을 다룬 논문에서 보듯이 계산 자원을 절약하면서도 성능을 지키려는 연구가 다수를 이뤘다. 둘째, 신뢰성과 안전성이다. 임바디드 에이전트의 물리적 위험 판단(EMBGuard), 안전한 이미지 생성(Concept Removal Guidance), 워터마킹(dgMARK), 코드 환각 통제(FuzzEval), 그리고 AI 락인 과 머신 언러닝 개념 자체를 되짚는 두 편의 포지션 논문까지, 기술에 대해 ‘얼마나 잘 작동하는가’를 넘어 ‘얼마나

안전하고 통제 가능한가’를 묻는 연구가 뚜렷하게 늘었다. 셋째, 생성모델의 구조적 이해다. 디퓨전 플로우 모델의 디코딩 순서, 위상 정보, 표현 공간의 기하학적 구조를 파고드는 이론적 연구들(TAG, LookUM, Physics in 2-Steps, Contrastive Flow Map Matching)은 생성모델이 여전히 활발한 기초 연구의 대상임을 보여준다. 한국에서 처음 열린 ICML이 국내 AI 생태계에 남긴 것은 무대 그 자체만이 아니다. KAIST, 연세대, 고려대, POSTECH 4개 대학의 연구진이 ‘원팀’으로 쌓아 올린 30편의 성과는 국가AI연구거점이 지향하는 산학연 협력 체계가 실질적인 연구 역량으로 이어지고 있음을 보여준다. 국가AI연구거점은 이번 성과를 발판으로 글로벌 개방형 혁신 생태계 구축에 더욱 속도를 낼 계획이다. [NAIRL](#)

지능을 넘어 현실 세계로, 서울에서의 하루

「글로벌 AI 프론티어 심포지엄 2026」 현장



「글로벌 AI 프론티어 심포지엄 2026」에서 개회사를 한 류제명 과학기술정보통신부 제2차관, 축사를 맡은 배충식 KAIST 총장, 김성근 POSTECH 총장, 기조연설에 나선 레슬리 팩 캄블링 MIT 교수, 노엄 브라운 오픈AI 부사장 등 주요 인사가 단체 사진 촬영에 임했다.

지난 7월 3일 서울 강남구 웨스틴 서울 파르나스에 산·학·연 전문가와 정부 관계자 1,000여 명이 모였다. 국가 AI연구거점^{NAIRL}과 글로벌AI프론티어랩이 공동 주관한 「글로벌 AI 프론티어 심포지엄 2026」이다. 지난해 학계 위주로 치러진 1회 행사와 달리 올해는 오픈AI·엔트로픽·퍼플렉시티의 핵심 인사와 프랑스 프레리^{PRÉRIE} 연구소, 캐나다 벡터연구소^{Vector Institute}의 연구자들이 합류했다. 주제는 ‘AI, 지능을 넘어 현실 세계로’. 모델 성능 경쟁을 지나, AI가 공장과 병원과 일상에서 실제로 무엇을 해내고 무엇에 걸려 넘어지는지를 묻는 하루였다.

오프닝과 기조연설: 무엇을 미리 알려주고 무엇을 측정할 것인가

류제명 과학기술정보통신부 제2차관은 개회사에서 나흘 전 발표된 ‘대한민국 대도약 3대 메가프로젝트’를 언급하며 2029년까지 8.4GW, 2035년까지 18.4GW 규모의 AI 데이터센터 확보 계획과 피지컬 AI 세계 1강 목표를 짚었다. 다만 발언의 무게는 마지막에 실렸다. AI가 가져올 혜택을 나누고 국가 간 격차를 줄이려면 조율된 국제적 대응이 필요하며, 두 연구 거점이 그 다리가 되어달라는 당부였다.

이어진 축사에서 배충식 KAIST 총장은 자신을 취임한 지 며칠 되지 않은 신임 총장이라 소개하며, AI의 미래가 기술보다 ‘우리가 던지는 질문과 우리가 길러내는 사람’에 달려 있다고 말했다. 김성근 POSTECH 총장은 한발 더 나아갔다. 김 총장은 AI를 지구에 착륙해 눌러앉기로 결정한 외계 지성에 비유하며, 인류가 처음으로 자신이 아닌 지성과 함께 사는 법을 배워야 하는 시대에 들어섰다고 말했다. 그러면서 이날 여섯 개 트랙 가운데 다섯이 AI를 어디에 쓸 것인가를 다루고 하나만이 AI와 인간의 관계를 다룬다는 점을 짚으며, 할 수 있는 일에 매혹된 나머지 이 방문객의 본질을 놓치지 말자고 당부했다. 지난해 무대에 섰던 안 르쿤 뉴욕대 교수와 최예진 스탠퍼드대 교수, 푸시미트 콜리 구글 딥마인드 부사장은 영상으로 축하를 보냈다.



학계를 대표해 레슬리 팩 캄블링 MIT 파나소닉 석좌교수가 ‘합리적 로봇’을 주제로 기조연설을 했다.

학계 기조연설에 나선 레슬리 팩 캄블링 MIT 파나소닉 석좌교수는 무대에 오르자마자 좌장을 가리키며 김기응 센터장의 박사학위 심사위원이었다는 사실이 자랑스럽다고 말해 웃음을 끌어냈다. ‘합리적 로봇’이라는 제목 아래 그가 붙든 문제는 재료 배분(무엇을 사람이 미리 넣고 무엇을 데이터로 배우게 할 것인가)이었다. 로봇을 만드는 데 필요한 것은 엔지니어와 데이터뿐인데, 엔지니어가 완벽한 프로그램을 써낼 수도 없고 모델이 유연할수록 필요한 데이터는 기하급수적으로 늘어난다. 그래서 어떤 규칙을 미리 정의해둘 것인가가 문제가 된다. 그가 고른 규칙은 3차원 세계, 행동 효과의 희소성과 국소성, 요구되는 과제의 다양성이었다. 그는 리모컨을 위로 집어 던져 보이며, 이렇게 해도 뒤쪽 음향팀은 아무 영향을 받지 않는다고 말했다.

차이는 영상에서 드러났다. 커피 캡슐 세 개를 트레이에 넣으라는 지시에 기존 비전-언어-행동 모델^{VLA}은 곧바로 움직였고, 그의 로봇은 잠시 멈춰 있다가 트레이 위의 캔부터 치웠다. 커피 캡슐 세 개를 넣으려면 캔을 먼저 옮겨야 한다는 판단에는 앞을 내다보는 인과 추론이 필요했다. 자기 부엌에 들어올 로봇이라면 최소한 이 정도는 해야 한다는 것이 그의 기준이었다. 그는 데이터를 충분히 쏟아부으면 무엇이든 학습할 수 있다는 말이 사실이라고 인정하면서도, 앞으로 10년, 20년의 지평에서는 데이터와 구조 사이의 교환 관계를 진지하게 따져야 한다고 맺었다.



산업계를 대표해 노엄 브라운 오픈AI 리서치 부문 부사장이 '대규모 추론 연산의 시사점'을 주제로 발표했다.

산업계 기초연설에 나선 노엄 브라운 오픈AI 리서치 부문 부사장의 주장은 한 문장으로 요약됐다. '우리는 지금 AI를 잘못된 방식으로 평가하고 있다.' 벤치마크 막대그래프로는 몇 퍼센트포인트 차이에 불과했던 최신 모델을 x축에 출력 토큰 수를 놓고 다시 그리자 곡선이 확연히 갈라졌다. 성능이 포화될 때까지 돌려보라는 상식적 해법도 이제 통하지 않는다. 최신 모델의 포화 지점이 너무 멀어졌기 때문이다. 영국 AI보안연구소AI Security Institute의 사이버보안 평가에서 모델은 1억 토큰을 생성할 때까지 성능이 올랐고, 거기서 멈춘 이유는 연구소가 더는 돌릴 수 없었기 때문이었다.

문제는 안전성으로 이어진다. 지능이 투입 연산량의 함수라면 평가는 어느 예산에서 수행해야 하는가. 현실의 평가는 낮은 예산에서 이뤄지지만, 어떤 조직은 1,000만 달러를 쓸 수 있다. 10달러 예산에서 위험하지 않다고 판정해 배포한 모델에 누군가 100만 달러를 투입하면, 평가가 보여준 것보다 훨씬 강력한 능력이 나온다. 그는 벤치마크를 단일 수치가 아니라 토큰·비용·시간의 함수로 공개하고, 안전 임계값 판단에 추론 연산량을 명시적으로 반영해야 한다는 제언으로 발표를 마쳤다.

오전 토론: AI 시대의 프런티어, 한 번도 가지 않은 길

두 기조연사는 곧 심층 패널로 자리를 옮겼다. 조경현 뉴욕대 교수와 에밀리 블랙 뉴욕대 교수가 합류했고, 김

기응 국가AI연구거점 센터장이 좌장을 맡았다. 그는 이 패널을 준비하느라 지난 한 주가 힘들었다고 털어놓으며 객석을 향해 물었다. “오늘 유행하는 문제가 아니라 경력을 걸 만큼 깊은 문제를 어떻게 알아보는가?”

먼저 패널들은 각자의 난제부터 꺼냈다. 캘블링 교수는 기억과 부분관측성을 꼽았다. 집이나 병원에서 몇 달을 작동할 로봇이라면 자신이 무엇을 했는지 기억해야 하는데, 토큰 스트림이나 파일 쓰기로는 ‘소금통을 어디 뒀는지’ 같은 간단한 기억도 붙들어두기 어렵다. 브라운 부사장은 에이전트의 작동 지평을 주 단위로 늘리는 일과 에이전트끼리 역할을 나눠 협력하게 만드는 일을 들었다. 조경현 교수는 연구 정체성의 위기를 겪고 있다고 고백했다. 코딩과 실험 설계가 자동화되면서, 인과추론처럼 기술적 세부에 막혀 손대지 못했던 문제로 들어갈 자유가 생겼다는 것이다. 블랙 교수는 평가의 불안정성을 지적했다. 벤치마크를 내놓고 조건 몇 개만 바꾸면 불확실성이 커진다는 것이다.

연구의 취향으로 넘어가자 캘블링 교수가 이날 가장 자주 인용된 문장을 남겼다. 잘 될 것 같은 프로젝트를 하고 있다면 이미 실패한 것이고, 충분히 멀리 손을 뻗지 않은 것이라고 했다. 자신은 늘 안 되는 일을 하고 있으며 거의 언제나 부수적으로 흥미로운 무언가가 생긴다고 덧붙였다. 브라운 부사장은 게임이론 전문가였다가 확신이 선 방향으로 전공을 옮긴 자신의 선택을 들려줬고, 조경현 교수는 검증 없이 쏟아지는 가설 생성 연구를 공상 생성기라 부르며 자신도 예외가 아니라고 했다.

마지막 라운드의 주제는 한국이었다. 조경현 교수의 답이 가장 날카로웠다. 그는 새로운 방향이 나오면 왜 하지 말아야 하는지 이유부터 찾는 분위기를 지적하며 한국에서 만나는 학생 대부분이 실리콘밸리나 뉴욕으로 옮기기만을 고대한다고 말했다. 아무도 입 밖에 내지 않지만, 학생들이 그 공기를 먼저 감지한다는 것이다. 그는 빅테크와의 경쟁을 피하고 변명거리부터 찾는 이 패배주의를 넘어서야 인재를 붙잡을 수 있다고 못 박았다. 블랙 교수는 한국의 AI 기본법을 언급하며 영향평가에 강건성을 넣으라고 조언했



김기응 국가AI연구거점 센터장이 좌장을 맡아 '글로벌 AI 리더십: 산·학·관 협력'을 주제로 진행한 심층 패널 토론. 캘블링 MIT 교수, 브라운 오픈AI 부사장과 함께 조경현 글로벌AI프론티어랩 공동소장 겸 뉴욕대 교수, 에밀리 블랙 뉴욕대 교수가 참여했다.

다. 하나의 결과만 제출하게 하지 말고 여러 영역에서 같은 것을 측정해 패턴을 드러내게 하라는 주문이었다.

객석에서 질문은 두 개를 받았다. 첫 질문자가 연구를 이끄는 윤리적 원칙이 있느냐고 묻자 캘블링 교수는 부끄럽지만 없다고 답했다. 평생 로봇을 더 잘 작동시키려 애써왔고 로봇은 여전히 거의 작동하지 않으며, 정말 잘 작동하기 시작하면 막대한 사회적 영향이 있으리라는 것은 사실이지만 그 부분에 대한 생각이 부끄러울 만큼 빈곤하다고 했다. 두 번째 질문자인 뉴욕대 박사과정생은 산업계와 자원 격차가 이토록 큰데 학계에서 같은 연구를 하는 가치가 무엇이나고 물었다. 브라운 부사장은 연산이 적게 드는 영역, 특히 평가 연구에서는 산업계의 우위가 없다고 답했다. 조경현 교수는 더 냉정했다. 투입과 성능의 관계가 예측 가능하고 결과에 금전적 이익이 따르는 문제라면 대학이 아니라 회사에서 할 일이며, 지속 가능한 사업을 만들기엔 대학은 끔찍한 장소라고 했다. 캘블링 교수가 마지막을 맡았다. 여러분은 그들만큼 똑똑하고, 팔 물건은 기업이 만들지만, 방법을 발명해 작은 규모에서 입증하는 일은 누구나 할 수 있다는 것이었다.

오후 분과별 세션: 여섯 개 트랙, 하나의 병목

점심시간 후에는 여섯 개 트랙이 병렬로 열렸다. 거대 언어모델·에이전틱 AI 트랙과 AI for Life 트랙은 같은 전

환을 기업과 소비자 양쪽에서 짚었다. 임우형 LG AI연구원 공동 연구원장은 에이전틱 AI의 본질이 인지 비용을 붕괴시키는 데 있고 그 붕괴가 산업 질서를 재편한다고 진단했다. 엑사원EXAONE 4.5와 K-엑사원을 축으로, LG그룹 8만여 명이 매일 쓰는 챗엑사원, 원자재 공급 스케줄링을 자율 최적화하는 석유화학 사례, 병리 전문 모델 엑사원 패스가 근거로 제시됐다. 그가 정리한 인간의 역할은 셋이다. 최종 위험을 지는 승인자, 여러 에이전트의 흐름을 통제하는 감독자, 어떤 문제를 풀지 기획하는 설계자. 두 번째를 그는 오케스트라 지휘자에 빗댔다. 모리타 준 퍼플렉시티 APAC 대표는 같은 전환을 웹에서 봤다. 링크를 일일이 열어 비교하던 검색이 항공편을 찾아 예약하고 일정까지 관리하는 행동으로 넘어갔다는 것이다. AI 브라우저 코멧Comet이 사례로 나왔다.

이어진 패널토의에서 걸림돌로 지목된 것은 검증이었다. PoC는 잘되지만 왜 그런 결과가 나왔는지 설명하지 못해 프로덕션에 오르지 못한다는 현장의 증언이 나왔다. 해법으로 거론된 것은 모델을 둘러싼 도구와 검증 절차를 통칭하는 하니스harness다. 같은 모델을 쓰는 두 회사가 다른 결과를 얻는다면 그 차이가 모델이 아니라 하니스에서 온다는 지적이었다.

피지컬 AI·체화형 지능 트랙과 멀티모달 AI 트랙은 언어 바깥의 신호를 다뤘다. 조민수 POSTECH 교수는 체



오후에는 분과별 세션으로 여섯 개 트랙이 병렬로 진행됐다. 사진은 ‘멀티모달 AI’를 주제로 한 트랙 B의 패널토의.

화형 AI가 긴 조작 과제를 수행하려면 3차원 공간에서 개방형 어휘 기반의 행동 유도성 그라운드링이 필요하다고 봤고, 2D 파운데이션 모델을 이용한 자동 라벨링과 제한된 시야에서 구조를 재구성하는 능동 지각을 해법으로 내놓았다. 패널토의에서는 현장 데이터 수집의 위험과 비용을 줄이기 위해 인터넷 영상과 시뮬레이션을 결합하는 계층적 혼합 훈련, 그리고 촉각 센서 데이터의 표준화가 논의됐다. 멀티모달 트랙의 기조는 하나의 비유로 요약됐다. 인터넷의 정적인 데이터를 가득 실은 화물선 대신, 환경과 계속 상호작용하며 적응하는 항해사 같은 모델을 만들자는 것이다. 아기의 시점에서 촬영된 연속 영상으로 시각과 행동을 폐쇄 루프에서 다시 학습시키는 접근이 소개됐고, 패널은 구조가 편이한 데이터를 언어 모델에 역지로 끼워 넣는 관성에서 벗어나야 한다는 데 뜻을 모았다.

AI for Science 트랙과 신뢰·안전·거버넌스 AI 트랙에서는 검증이 다시 중심에 놓였다. 황성주 KAIST 교수는 문헌 탐색부터 가설 설정과 코드 구현, 실험 검증까지 스스로 수행하는 연구 에이전트를 선보이며 신약 발견 사례를 들었다. 패널토의의 화두는 AI가 만든 가설을 어떻게 엄격하게 검증할 것인가, 벤치마크가 실제 성과를 반영하지 못하는 한계를 어떻게 넘을 것인가였다. 자연과학에는 언어와 달리 자연법칙이 있는 만큼, 인간의 사전 지식을 귀납적 편향으로 주입하는 일이 과제로 남았다.

김명주 AI안전연구소장은 정적인 위험 분류의 한계를 지적하며 원인과 영향을 입체적으로 있는 위험 매핑을 제

안했다. 주체와 수명 주기, 결함 영역을 분석하고 그에 따른 사회적 피해와 책임 부처를 명시해, 분류표에 고치지 않는 거버넌스 도구를 만들겠다는 구상이다. 패널토의에서는 에이전트 간 상호작용이 낳는 시스템 위험, 지식이 서로 얹혀 있어 머신 언러닝이 근본 해법이 되기 어렵다는 지적, 컴퓨팅 자원을 원천적으로 검증할 글로벌 인프라의 필요성이 이어졌다.

연구에서 일상까지, 네 개의 전시 존

무대 밖 아틀라스홀에는 ‘AI 가치사슬의 완성: 연구에서 일상까지’를 주제로 한 특별 전시가 차려졌다. 연구에서 나온 것이 산업으로 옮겨가고 끝내 일상에서 경험되는 경로를 하나의 동선으로 만든 구성이다. 주최 측은 이를 연구라는 심부^{The Core}, 인프라라는 골격^{The Backbone}, 산업 전환이라는 동력^{The Engine}, 일상이라는 경험^{The Experience} 네 단계로 나눴다.

첫 번째 존은 연구였다. 글로벌시프론티어랩은 견고함과 책임성, 생명을 구하는 기술을 앞세워 서울과 뉴욕 두 축의 협력을 정리했고, 국가AI연구거점은 한국을 넘어 AI 혁신의 중심이 되겠다는 문장을 부스 정면에 세웠다. 아직 답이 없는 문제를 함께 붙든다는 것이 이 존의 취지였다.

두 번째 존의 제목은 ‘칩에서 에이전트까지’였다. 칩이 없으면 연산이 없고, 언어 이해가 없으면 답이 없고, 에이전트가 없으면 일이 되지 않는다는 순서로 부스가 놓였다. AIM퓨처는 국산 NPU IP로 설계한 엡지 AI를, 퍼블릭AI는



‘AI 가치사슬의 완성: 연구에서 일상까지’를 주제로 마련된 특별 전시. 국가AI연구거점과 협력 기업이 참여했다.

구축부터 운영까지 이어지는 기업용 AI를, AX는 명령 없이 알아서 일하는 에이전트를 내걸었다. 포티투마루는 ‘묻고, 답하고, 행동까지’를 테마로 환각 제어 기술과 프라이빗 보안 모드를 적용한 ‘신뢰할 수 있는 답변 AI’ 플랫폼의 구조와 적용 시나리오를 공개했다. 검색증강생성과 기계독해, 도메인 특화 경량 모델을 묶은 결과물로, 민간 AI-MASTER와 공공 TTA CAT 인증을 함께 받은 이력을 내놴다.

세 번째 존은 현장이었다. 플레이오나는 현장 데이터로 학습해 설비를 직접 움직이는 공정 제어를, 대동AI는 스스로 감지하고 판단하고 실행하는 농업 에이전트를, AIS타는 배터리의 성능이 실험실이 아니라 데이터에서 결정된다는 접근을, 4D비전은 공장 바닥을 살피는 엡지 AI 3D 비전을, 하이퍼클라우드는 현실과 가상을 잇는 산업용 환경을 소개했다. 오전 기초연설과 오후 트랙에서 반복된 피지컬 AI 논의가 이 존에서는 제조와 농업의 언어로 번역돼 있었다.

네 번째 존은 이용자 쪽이었다. 에버엑스는 근골격계 재활을, 디지털메딕은 정신건강 관리 서비스 디딤스^{DIDIX}를, 헬프티처는 감정을 읽고 돌봄으로 잇는 기술을, 코르카는 함께 생각하는 업무 도구를 내놓았다. 코르카는 거대언어 모델·에이전틱 AI 트랙의 패널토의에 참석한 정영현 대표가 이끄는 회사이다. 연구실에서 출발해 공장을 지나는 기술이 마지막에 개인의 몸과 마음에 닿는 자리였다.

전시장 한편에는 퍼플렉시티와 협업한 AI ART 섹션이 놓였다. 우주적 풍경과 디지털 몽상, 알고리즘적 창작에

서 출발한 몰입형 시각 서사로 인간의 상상력과 AI의 관계를 다룬 구성이다. AI를 도구가 아니라 창작의 동반자로 두고, 호기심과 지각과 이야기가 어떤 시각 언어로 바뀔 수 있는지 보여주는 자리였다.

이날 무대에는 두 개의 언어가 나란히 놓였다. 하나는 규모의 언어다. 18.4GW의 데이터센터, 세계 1강, 독자 파운데이션 모델. 다른 하나는 규모로 풀리지 않는 문제들의 언어다. 몇 달을 일한 로봇이 소금통을 어디 뒀는지 기억하는 일, 에이전트가 자기가 내놓은 답을 스스로 확인하는 일, 만들어진 가설을 실험대에 올리는 일, 그리고 될 것 같지 않은 문제를 일부러 골라잡는 일. 앞의 것은 예산과 시간을 들이면 되지만, 뒤의 것은 그렇지 않다.

조경현 교수가 지적한 패배주의는 아마 이 두 언어 사이의 간격에서 자란다. 프런티어를 규모로만 정의하면 우리가 설 자리는 늘 남의 뒤가 되고, 야망은 자연히 낮아진다. 이날 오간 질문들은 그 정의를 다시 열어놓았다. 로봇에게 무엇을 미리 알려줄 것인가, 우리는 무엇을 측정하고 있는가, 검증되지 않은 결과를 어떻게 신뢰할 것인가. 어느 것도 GPU 숫자로 답할 수 있는 질문이 아니다.

행사가 내건 주제는 지능을 넘어 현실 세계로 가자는 것이었다. 심포지엄이 끝날 무렵 분명해진 것은 현실 세계가 연구실보다 훨씬 까다로운 심사위원이라는 사실이었다. 그 심사를 통과하는 일은 더 큰 모델이 아니라 더 나은 질문에서 시작된다. 서울에 모인 1,000명이 하루 중 일 붙들고 있었던 것도 그 질문들이었다. **NAIRL**

RIKEN-Vector-NAIRL 워크숍 현장 스케치

세 나라의 AI 연구자들, AI의 다음을 논하다

지난 7월 5일 서울 서초구 AICT빌딩 프론티어홀에서 국가AI연구거점(NAIRL)과 일본 이화학연구소 첨단지능프로젝트(RIKEN AIP), 캐나다 벡터연구소(Vector Institute) 소속 연구자들이 한자리에 모였다. 'RIKEN-Vector-NAIRL 워크숍'에 참석하기 위해서다. 이번 워크숍에서 참가자들이 어떤 연구를 공유했는지 살펴보자.

세 연구기관, 서울에서 마주하다

행사 시작에 앞서 김기웅 국가AI연구거점NAIRL 센터장은 “가능한 한 편안하고 격식 없이 즐겨 달라”는 인사로 개회를 알리고 워크숍 식순을 소개했다. 오전에는 기관 소개와 연구 발표가 이어졌고, 오후에는 MOU 체결식과 라이트닝 토크·포스터 세션이 진행됐다.

그렇다면 NAIRL과 함께 워크숍을 진행한 두 기관은 어떤 곳일까. 먼저 RIKEN AIP의 모 기관인 RIKEN은 1917년 설립됐으며, 노벨화학상 수상자 노요리 료지가 이사장을 지낸 연구소다. 그 산하 조직으로 2016년 출범한 AIP는 올해 3월 첫 10년 프로젝트를 마치고 2기 체제에 들어섰다. 또 다케우치 이치로(chiro Takeuchi PI)의 설명에 따르면 연구자 10명 중 7명이 외국인이며, 지난 10년간 240명의 해외 인턴을 받아들인 것이 특징적이다.

이어 벡터연구소는 2017년 제프리 힌턴 등이 주축이 되어 설립된 캐나다의 국가 AI 연구기관이다. 전국 단위로 대학과 네트워크를 이루고 산업 파트너 30여 곳, 병원과도 협력한다. 연구 분야는 헬스와 과학 같은 실용 분야부터 AI 안전성, 최근 주목받는 생성형 AI까지 넓혔다. 연구를 산업 현장으로 옮기는 엔지니어링팀도 따로 두고 있다.

한편 NAIRL은 아직 3년 차에 접어든 곳으로, 10년을 넘긴 RIKEN AIP, 9년 차 Vector Institute와 비교하면 역사가 짧다. 그럼에도 2024년 KAIST·고려대·연세대·POSTECH 컨소시엄이 국가AI연구거점으로 선정되며 2028년까지 440억 원 규모의 정부 지원을 확보했다. 출범 시기는 달라도 세 기관 모두 각국의 AI 연구 생태계를 잇는 허브라는 공통점을 지니며, 이번 워크숍에서 한자리에 모여 연구 의제를 공유했다.



국가AI연구거점(NAIRL), 일본 이화학연구소 첨단지능프로젝트(RIKEN AIP), 캐나다 벡터연구소(Vector Institute)의 조인트 워크숍에 참여한 연구자들의 단체 사진.



벡터연구소를 소개하고 있는 파스칼 푸파르 교수.

효율·현실·신뢰, AI의 다음 과제를 묻다

기관 소개 이후 본격적인 발표가 이어졌다. NAIRL·RIKEN·Vector 소속 연구원들이 한 명씩 돌아가며 발표를 진행했고, 첫 세션은 ‘커지는 AI 앞에서 무엇을 다시 봐야 할까’로 시작됐다. 먼저 NAIRL 소속의 이재호 교수(POSTECH)는 “우리 예산은 프린티어 AI 모델 하나를 학습시키는 비용의 20%도 안 된다”고 말했다. 한국이 미국·중국과 규모로 경쟁할 수 없는 만큼, 이론으로 효율을 찾아야 한다는 문제의식이다. 그는 신경망 초기화 스케일을 조절하면 학습 속도와 일반화 성능이 트레이드오프 관계를 이루며, 그 사이 최적의 ‘스위트스팟’이 존재함을 이론

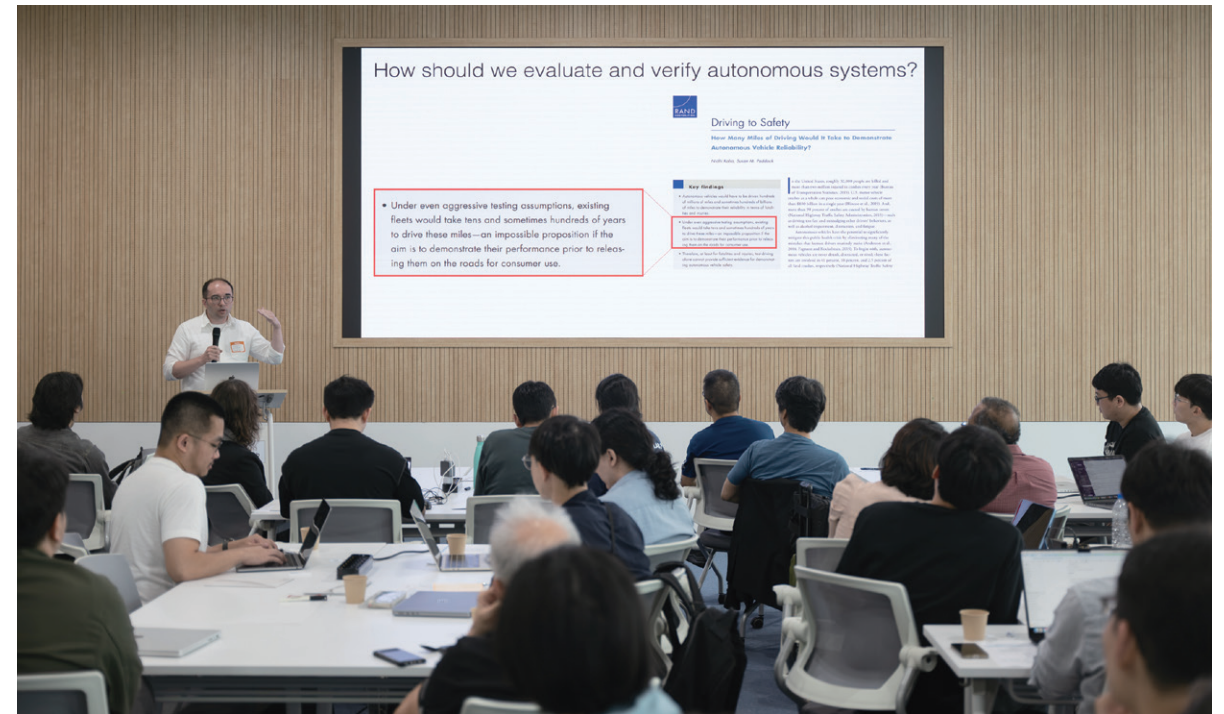
이날 워크숍에서는 3개 기관 소속의 연구원들이 활발한 토론을 벌였다.

적으로 규명한 연구를 소개했다.

다케우치 PI는 AI가 가설 생성부터 실험 설계, 데이터 수집까지 관여하는 ‘AI 주도 과학’을 짚었다. 이런 환경에서는 절차를 사전에 고정해야 한다는 전통 통계학의 원칙이 무너지는 것이다. 이에 그는 탐색이 끝난 뒤에도 신뢰성을 사후적으로 복원할 수 있는 선택적 추론(selective inference) 기법을 제안했다.

Vector Institute 소속의 샤이 벤데이비드(Shai Ben-David) 교수는 동일한 분포를 가정하는 기존 학습이론이 현대 AI에는 더 이상 들어맞지 않는다고 지적했다. 자릿수를 늘려가며 일반화해야 하는 지금, 학습 데이터를 제시하는 순서와 문제의 복잡도 구조가 학습의 성패를 가른다는 이론적 결과를 제시했다.

두 번째 세션은 ‘현실에서 통하는 AI를 만들려면 무엇이 필요할까’로 이어졌다. NAIRL 소속의 박은병 교수(연세대)는 여러 장의 이미지를 한 번에 3D로 복원하는 feed-forward 방식의 3D Gaussian Splatting 연구를 소개했다. 현장에서는 화면을 직접 터치해 조작하는 실시간 데모로 청중의 관심을 끌었다.



벡터연구소 소속의 플로리안 슈쿠르티 교수가 로봇 정책의 안전한 평가 방법에 대해 발표하고 있다.

RIKEN 소속의 요코야 나오토(Naoto Yokoya) 교수는 소수 빅테크가 독점하다시피 한 위성 데이터 생태계에 맞서, 전 세계 어디에나 적용할 수 있는 개방형 지구관측 AI 벤치마크를 구축해 온 성과를 전했다. 최근에는 원격탐사 특화 질의응답 에이전트로 연구를 확장하고 있다고 밝혔다. Vector Institute 소속의 플로리안 슈쿠르티(Florian Shkurti) 교수는 로봇을 실제로 수없이 돌려보지 않고도, 오프라인 데이터와 확산 모델 기반 기법으로 로봇 정책을 안전하게 평가하는 방법을 소개했다.

세 번째 세션은 ‘AI를 신뢰할 수 있는가’로 마무리됐다. NAIRL 소속의 심현정 교수(KAIST)는 오늘날 파운데이션 모델의 한계를 짚었다. 예시로 ‘우산을 쓰지 않은 사람을 찾아 달라’는 요청에도 부정 표현을 무시한 채 우산을 쓴 사람까지 찾아내는 문제점이 있는데, 현재 모델들이 부정 표현 이해와 3D 공간 이해에 취약하다는 것이다.

RIKEN 소속의 아라이 히로미(Hiromi Arai) 교수는 일본 사회 맥락에 맞춘 편향 평가 벤치마크(JBBQ)를 소개했다. 그는 설명할 수 있는 AI가 오히려 불공정성을 감추는 도구로 악용될 수 있다는 ‘페어워싱’ 위험도 경고했다.

Vector Institute 소속의 샤오샤오 리(Xiaoxiao Li) 교수는 장기 과제를 수행하는 자기개선형 AI 에이전트에서 발생하는 오류를 진단·복구하는 연구를 소개했다. 신경망의 잔차연결(ResNet) 구조에서 착안한 기법으로, 신소재 설계 분야와의 실제 협업 사례까지 공유했다.

연구 교류에서 협력으로, 연결을 넓히다

오전 발표를 마친 참석자들은 낮 12시경부터 로비와 포럼티어홀에서 열린 네트워킹 런치에서 자연스럽게 어울렸다. 세 기관에서 온 연구자들은 서로의 발표에 관한 질문을 주고받으며 편안한 분위기 속에 오전 일정을 마무리했다.

점심식사 후에는 김기웅 센터장과 파스칼 푸파르(Pascal Poupart) Vector Institute 리서치 디렉터 겸 워털루대 교수가 양 기관을 대표해 업무협약에 서명했다. 양측은 이번 파트너십이 더 깊은 협력의 토대가 되어 글로벌 AI 연구 커뮤니티의 발전에 기여하기를 기대했다. NAIRL은 앞으로도 국내 연구 커뮤니티와 세계 유수의 프런티어 AI 기관을 연결해, 과학적 발견과 산업적 혁신이 함께 이어지는 연구 생태계를 넓혀갈 계획이다. [NAIRL](#)

「2026 국가AI연구거점 성과·발전 방향 워크숍」 개최

제주에 모인 거점,
다음 걸음을 그린다



「2026 국가AI연구거점 성과·발전방향 워크숍」에 참가한 사람들의 단체 사진.



워크숍의 개최사를 하고 있는 김기응 센터장.

1단계 성과를 돌아보고, 2단계 도약을 논의하다

지난 6월 25일, 국가AI연구거점^{NAIRL}에 참여하는 연구 책임자와 교수진, 대학원생 연구원들이 제주국제컨벤션 센터ICC 제주에 한자리에 모였다. 한국정보과학회 하계학 술대회^{KCC} 기간에 맞춰 열린 「2026 국가AI연구거점 성과·발전 방향 워크숍」은 1단계 사업의 연구 성과를 점검 하고, 사업 종료 전 보완이 필요한 과제를 짚기 위해 마련 됐다. 여기에 빠르게 변화하는 AI 연구 환경에 대응할 2단 계 추진 방향을 함께 모색한다는 목표가 더해졌다.

이번 워크숍의 핵심 질문은 명확했다. “1단계에서 우리 는 무엇을 이루었는가”, 그리고 “2단계에서는 무엇을 선택하고 집중할 것인가.” 참석자들은 각 세부 과제의 연구 성과를 객관적으로 돌아보는 동시에, 국가AI연구거점이 하나의 연구 공동체로서 어떤 방향으로 협력하고 발전해 나갈 것인지 깊이 있게 논의했다.

세부과제 성과를 공유하고, 다음 연구 방향을 함께 고민하다

워크숍은 김기응 센터장의 개최사로 막이 올랐다. 이어 ‘NVIDIA B200 GPU 인프라 운영 현황과 향후 통합 클러 스터 구축 계획’ 발표가 이어졌다. 현재 240장 규모로 확 대된 GPU 인프라의 운영 현황과 함께, 연구 경쟁력 강화 를 위한 통합 클러스터 구축 방향이 공유됐다.

이후 1·2·3세부 책임교수(양은호, 조민수, 예종철 교수) 는 각 세부 연구 진행 현황과 주요 성과를 발표하고, 1 단계 사업의 성과를 돌아보며 향후 보완 과제를 제시했 다. 다른 연구 일정으로 현장 참석이 어려웠던 예종철 교 수와 노알버트 교수는 화상으로 참여해 발표를 이어갔으 며, 실시간으로 발표 자료를 공유하고 질의응답에 참여하 며 현장과 활발한 논의를 이어갔다.

이어진 세부 과제별 토론은 이번 워크숍의 핵심 프로그



1세부 소속 구성원들의 토론 모습. 1단계 연구성과를 점검하고, 에이전트 중심 연구를 2단계 핵심 방향으로 제시했다.

램이었다. 참가자들은 ① 1단계 목표 대비 성과, ② 보완이 필요한 연구와 추진 과제, ③ 학계가 수행해야 할 차별화된 연구, ④ 2단계에서 집중해야 할 연구 방향을 중심으로 의견을 나눴다.

이날 토론 장소 배정에는 작은 에피소드도 있었다. 3세부는 참가 인원이 가장 적었던 덕분에 회의실에서 자리를 옮겨야 했고, 결국 컨벤션센터 로비에서 토론을 이어갔다. 참가 인원은 가장 단출했지만, 로비에서 이어진 토론의 열기만큼은 다른 세부에 결코 뒤지지 않았다.

뉴럴 스케일링 법칙 초월 연구를 진행하고 있는 1세부에서는 모델 경량화와 학습 알고리즘, 데이터셋 구축 등 1단계 연구 성과를 점검하는 한편, Tool-use Agent, Vision-Language-Action^{VLA}, 에이전트 간 지식 종류 등 에이전트 중심 연구를 2단계 핵심 방향으로 제시했다. 또

한 연구실 간 협력과 기업 연계 확대를 위한 다양한 방안도 함께 논의됐다.

로봇 파운데이션 모델을 연구하는 2세부에서는 18명의 교수가 참여해 퍼셉션·컨트롤·리징 분야의 연구 성과를 공유했다. 상위 국제학회 논문 실적이 목표 대비 약 두 배에 달하는 성과를 거둔 가운데, 연구자 간 협업 체계를 더욱 강화하기 위해 학생 중심 정례 연구 미팅, 연구자 매칭, ICRA 2027 참여 등 구체적인 협력 방안이 논의됐다.

고차원 생성 파운데이션 모델에 집중하고 있는 3세부에서는 논문 실적이 목표 대비 약 세 배에 달한 성과와 국내외 특허 실적을 공유하며, 비디오·의료·분자 생성 모델을 하나의 물리 기반 월드 모델 연구로 연결하는 2단계 연구 비전을 제시했다. 아울러 의료, 신약 개발, 강화학습 분야 연구자 간 협업 가능성도 활발히 논의됐다.



세부과제별 토론 후 발표하는 시간은 토론 결과를 공유하는 시간이었다.

함께 성장하는 거점을 위한 다음 단계

이번 워크숍은 교수진뿐 아니라 대학원생들에게도 특별한 자리였다. 같은 세부과제 소속이라도 학교와 연구실이 다르다면 서로 마주칠 일이 좀처럼 없었던 학생들이 이번 워크숍을 계기로 얼굴을 맞대고 이야기를 나눴다. 발표와 토론 사이사이 쉬는 시간마다 삼삼오오 모여 서로의 연구 근황을 묻고 실험 노하우를 나누는 모습이 여기저기서 눈에 띄었다. 같은 주제를 다른 각도에서 파고들던 학생들이 한자리에서 마주하는 경험은 논문이나 이메일로는 좀처럼 얻기 힘든 소득이었다.

토론 결과를 공유하는 시간에는 세부과제별 논의를 바탕으로 공동 GPU 활용 연구 확대, 학생 중심 협업 체계 구축, 수요기업과의 협력 강화 등 거점 차원의 협업 방안이 제안됐다. 참석자들은 개별 연구의 경쟁력을 높이는

것과 함께 세부과제 간 협력 구조를 더욱 강화해야 한다고 한목소리를 냈다.

워크숍을 마무리하며 참석자들은 “왜 우리는 하나의 거점으로 함께하고 있는가”라는 질문을 다시 한번 공유했다. 개별 연구실의 독립성과 자율성은 유지하되, 공동 연구와 기업 연계, 연구 인프라를 중심으로 협력 체계를 더욱 강화해 나가는 것이 국가AI연구거점 2단계의 중요한 방향이라는 점에 뜻을 모았다.

이번 워크숍은 1단계 사업의 성과를 정리하는 자리를 넘어, 빠르게 변화하는 AI 연구 환경 속에서 국가AI연구거점이 함께 고민하고 협력하며 다음 도약을 준비하는 의미 있는 시간이었다. 제주에서 이어진 논의는 앞으로의 공동 연구와 국제 협력으로 확장되며 국가AI연구거점의 새로운 성장 기반이 될 것으로 기대된다. **NAIRL**

36개 팀의 도전, 22개 팀의 결실, 그리고
150여 명이 함께한 글로벌 네트워킹 데이

「GARL Challenge 2026」, 성황리에 막을 내리다



Top 3 수상팀이 국내외 석학들 앞에서 15분간 영어로 자신의 연구 내용을 피칭했다.
사진은 허치영 학생 연구팀이 터미널 에이전트 벤치마크 연구를 발표하는 장면.

국경을 넘은 연구 아이디어, 세계 석학과 만나다

국가AI연구거점^{NAIRL}이 올해 처음으로 선보인 「Global AI Research Leap Challenge 2026」(이하 GARL 챌린지)이 지난 7월 8일 시상식과 글로벌 네트워킹 데이를 끝으로 막을 내렸다. 이번 챌린지는 논문으로 완성되기 이전 단계의 연구 아이디어를 조기에 발굴하고, 국내외 연구자들과 함께 검증하고 발전시킬 수 있는 교류의 장을 마련하자는 취지에서 기획됐다.

보통의 학술대회가 이미 완성된 연구 성과를 발표하는 자리라면, GARL 챌린지는 거꾸로 접근했다. ‘아직 논문이 되지 않은, 그래서 더 날카로운 피드백이 필요한 아이디어’를 무대에 올린 것이다. 여기에 7월 3일 개최된 ‘Global AI Frontiers Symposium 2026’과 연계해 해외 석학 및 글로벌 AI 기업 연구진과의 네트워킹 기회까지 더해, 차세대 AI 연구자들이 국제 학술무대에 한 걸음 먼저 다가설 수 있도록 설계했다.

날짜 선정에도 계산이 있었다. 세계 3대 머신러닝 학회로 꼽히는 ‘ICML 2026’이 7월 6일부터 11일까지 서울 코엑스에서 열리며, 평소라면 각자의 대륙에 흩어져 있을 연구자들이 같은 도시에 모여 있는 시기였다. 그 주를 겨냥한 덕에 국내 대학생들의 포스터 앞에 세계 각지의 연구자들이 서는 장면이 만들어졌다.

서류에서 무대까지, 4개월의 여정

지난 3월 6일 공고를 시작으로 4월 24일 1차 서류 접수, 5월 29일 15분 분량의 발표 영상 접수를 거쳐 6월 12일 결과가 발표됐다. 국가AI연구거점 참여교수 10명(KAIST 4명, 고려대 2명, POSTECH 2명, 연세대 2명)이 비대면으로 서류와 발표 영상을 심사했다. 평가 기준은 연구 주제의 창의성, 학술적 기여도, 실현 가능성, 확장성과 파급력, 그리고 제한된 시간 안에 얼마나 논리적으로 전달하는지를 보는 피칭 역량까지 다섯 갈래로 나뉘었다.

국가AI연구거점 소속 교수 45명이 지도하는 대학원생을 대상으로 한 이번 공모에는 총 36개 연구팀이 지원했고, 치열한 심사 끝에 Top 3와 우수 19개 팀, 총 22개 팀

이 최종 선정됐다.

시상식은 ‘Global Scholars Networking Day’라는 이름으로 7월 8일 열렸다. Top 3 수상팀은 국내외 석학들 앞에서 15분씩 영어로 연구를 발표하고 깊이 있는 질의응답을 나눴으며, 우수상을 받은 19개 팀을 포함한 22개 전체 팀은 포스터 세션을 통해 연구 내용을 소개했다.

영예의 Top 3에는 세 팀이 이름을 올렸다. 허치영 외 1명(POSTECH 인공지능대학원 옥정술 교수 연구실)으로 구성된 팀 ‘MMTB’는 터미널 에이전트가 이미지·영상 등 멀티미디어 파일 작업을 얼마나 잘 수행하는지 평가하는 벤치마크 연구 ‘MMTB: Benchmarking Terminal Agents on Multimedia File Tasks’를 발표했다. 이수용(KAIST 김재철시대학원 신기정 교수 연구실) 연구자는 AI 행동에 대한 과학적·공학적 연구 과정 자체를 자동화하는 ‘Automating Research on the Science and Engineering of AI Behavior’를 선보였다. 차재영 외 1명(KAIST 김재철시대학원 윤철희 교수 연구실)으로 구성된 팀 ‘OptiML’은 더 나은 검증 전략 설계를 통해 트리 기반 추측 디코딩을 개선하는 ‘Improving Tree-Based Speculative Decoding by Designing a Better Verification Strategy’를 발표했다.

Top 3의 발표 뒤에는 조금 다른 성격의 세션이 이어졌다. AI 안전과 거버넌스를 연구하는 국제 비영리 단체 ‘퓨처 오브 라이프 인스티튜트(Future of Life Institute)’의 제임스 페트리(James Petrie) 컴퓨터 보안·거버넌스 총괄과 사비나 농(Sabina Nong) AI 안전 조사관이 컴퓨터 거버넌스와 하드웨어 검증을 주제로 발표한 것이다. 발표의 골자는 모델의 성능이 아니라 모델을 돌리는 연산 자원 자체를 어떻게 들여다볼 것인가, 그 검증을 정책 문서가 아니라 하드웨어 층위에서 구현할 수 있는가였다. 벤치마크와 디코딩 최적화를 이야기하던 자리에 거버넌스 논의가 얹히면서, 연구실의 문제와 국제 규범의 문제가 한 테이블에 놓였다.

이날 현장에는 벡터연구소(Vector Institute), 밀라(Mila), 코넬대 등 세계적 연구기관의 석학들과 엔비디아, 퍼플렉시티 등 글로벌 AI 기업의 관계자가 함께했다. 국제 공동연구 교수



Top 3 수상팀이 시상식을 마친 뒤 김기응 센터장과 함께 단체 기념사진을 촬영했다.

진으로는 밀라의 아론 쿠르빌(Aaron Courville), 이리나 리시(Irina Rish) 교수, 워털루대와 벡터연구소의 파스칼 푸파르(Pascal Poupart) 교수, 코넬대의 대니얼 리(Daniel D. Lee) 교수가 자리를 지켰다. 학생, 교수, 기업 인사를 합쳐 총 150여 명이 참석해, 시상식이라기보다는 하나의 국제 학술 교류회에 가까운 풍경이 연출됐다.

현장에서 건진 뒷이야기 셋

① 동점이 만든 두 자리

원래 심사위원단이 계획했던 우수상 규모는 17개 팀, Top 3를 더해 총 20개 팀이었다. 그런데 막상 점수를 집계해 보니 우수상 커트라인에서 동점이 나온 것. 논의 끝에 심사위원단은 같은 점수를 받은 팀 모두를 우수상으로 인정하기로 결정했고, 우수상은 17개 팀에서 19개 팀으로, 전체 수상팀은 20개 팀에서 22개 팀으로 늘어났다.

포스터를 붙일 자리를 부랴부랴 두 자리 더 마련해야 했던 담당자들의 유쾌한 소동은 이번 챌린지의 첫 번째 뒷이야기로 남았다.

② 마감 직전의 화이트보드

15분짜리 영어 발표 영상은 결코 만만한 도전이 아니었다. 마감일이던 5월 29일 자정 무렵까지 스튜디오 대신 연구실 화이트보드 앞에서 리허설을 반복하며 대본을 고쳐 쓰던 팀들도 있었다는 후문이다. 그렇게 다져진 실력은 시상식 무대에서 드러났다. 세계적 석학들의 날카로운 질문에 진땀을 빼면서도 또박또박 자신의 아이디어를 설명해내는 학생들의 모습에, 지켜보던 참석자들 사이에서

22개 전체 팀이 참여한 우수 연구성과 포스터 세션 현장. 국제 학술 교류회에 가까운 풍경이 연출됐다.

는 “저 패기가 진짜 국가AI연구거점의 미래”라는 뿌듯한 이야기가 오갔다.

③ 행사 뒤에 도착한 메일

해외 석학들의 관심은 예정된 순서에서 끝나지 않았다. 행사가 마무리된 뒤에도 여러 교수가 학생들의 발표 자료를 다시 볼 수 있도록 공유해달라고 따로 요청해온 것이다. 15분으로 끝날 수도 있었던 아이디어들이 석학들이 돌아가는 비행기 안에서 한 번 더 읽히게 된 셈이다. 담당자들 사이에서는 “가장 확실한 성공 지표는 행사 후에 도착하는 메일”이라는 말이 나왔다.

경쟁을 넘어, 함께 성장하는 연구문화로

GARL Challenge 2026은 단순한 논문 피칭 대회로

넘어, 국내 AI 연구자가 세계적 연구자들과 직접 연구를 공유하고 토론하는 국제 연구 교류 프로그램으로 자리매김했다는 평가를 받는다. 논문 게재 이전 단계의 아이디어를 조기에 발굴해 완성도를 높이는 계기가 됐다. 참가 학생들에게는 국제 학술무대에서 발표하고 소통하는 값진 경험이, 국가AI연구거점에는 해외 우수 연구기관 및 글로벌 AI 기업과의 협력 기반 확대라는 성과가 남았다.

행사를 마친 뒤에도 여러 팀에서 해외 연구자들과의 공동연구 논의가 이어지고 있다는 반가운 소식이 들려오고 있다. 경쟁이 아니라 공유와 피드백을 중심에 둔 이 새로운 연구 문화가 앞으로도 국가AI연구거점 연구실들 사이의 교류를 더 활발하게 이어가는 씨앗이 되기를 기대해본다. [NAIRL](#)



국가AI연구거점은 4개 대학(KAIST, 고려대, 연세대, POSTECH) 컨소시엄 외
미국, 캐나다, 프랑스, UAE의 총 13개 국제공동연구기관, 국내 12개 협력기업,
서울특별시 및 서초구청 지자체 협력으로 구성된 역동적인 산·학·연·관 협력 생태계입니다.



NAIRL Frontier Letter | 2026 | Vol.04

발간인 국가AI연구거점 김기응 센터장
발행일 2026년 9월
발행처 서울특별시 서초구 양재대로12길 56
제작 (주)동아에스앤씨

