

# NAIRL Frontier LETTER

2025 | Vol.02

NAIRL Tech Horizon

## ‘AI 프런티어 국제 심포지엄 2025’

“AI, 이제는 세계를 이해해야 한다”

안 르균·최예진 교수가 제시한  
차세대 AI의 미래

“AI는 공공 인프라가 돼야 한다”

글로벌 석학들과 국내 전문가들이 그린  
AI의 미래 비전

세션1 - 차세대 AI를 위한 기초와 혁신

세션2 - 실세계 자율성을 위한 피지컬 AI

파트너 기업 전시 부스 스케치

NAIRL Insight Talk

**장 풍스** 프랑스 파리 고등사범학교(ENS-PSL) 교수 | 파리 AI 융합연구소 PR[AI]RIE-PSAI 과학디렉터

“기초를 묻고, 이론의 밑그림을 그리는 것이야말로 AI 시대의 연구”

 National AI  
Research Lab



## NAIRL SPECIAL COLUMN

- 04 · 홍진배 정보통신기획평가원<sup>ITP</sup> 원장  
AI G3 달성,  
GPU × NPU 인프라와 생태계 확보해야

## NAIRL RESEARCH HIGHLIGHT

- 06 · 이병준 고려대 인공지능학과 교수  
배우면서 더 잘 배우는 법도 같이 배우기
- 10 · 박은병 연세대 첨단컴퓨팅학부 교수  
초거대 3D 재구성 모델
- 14 · 박상돈 POSTECH 인공지능대학원 교수  
인간과 AI의 가치 정렬

## NAIRL TECH HORIZON

- 18 · 안 르쿤·최예진 교수가 제시한 차세대 AI의 미래  
“AI, 이제는 세계를 이해해야 한다”
- 20 · 글로벌 석학들과 국내 전문가들이 그린  
AI의 미래 비전  
“AI는 공공 인프라가 돼야 한다”
- 22 · 차세대 AI를 위한 기초와 혁신  
“비용 넘는 창의적 해법 필요”
- 24 · 실세계 자율성을 위한 피지컬 AI  
“현실 세계에서 작동하는 AI,  
이제는 ‘물리적 지능’으로 간다”
- 26 · ‘글로벌 AI 프론티어 심포지엄 2025’  
파트너 기업 전시 부스 스케치  
AI 산업과 일상 속으로



## NAIRL INSIGHT TALK

- 28 · 컴퓨터 비전 분야의 세계적 석학  
장 풍스 교수 인터뷰  
“기초를 묻고 이론의 밑그림을  
그리는 것이야말로 AI 시대의 연구”

## NAIRL IMPACT

- 32 · 산업 AI 혁신을 선도하는 AI 스타트업  
플레이오니  
AI로 제조에서 로봇틱스까지 도전한다

홍진배  
정보통신기획평가원<sup>ITP</sup> 원장

AI G3 달성,  
GPU × NPU 인프라와 생태계 확보해야



전 세계 AI 패권 경쟁의 중심축이 데이터·모델을 넘어 ‘AI 인프라’로 이동하고 있다. 생성형 AI에서 AI 에이전트, 퍼지컬 AI로 일상과 산업 전반에 확산되면서 컴퓨팅 파워는 더 이상 연구용 자원이 아니라 국가 경쟁력을 좌우하는 전략적 기반이 되고 있다. AX(AI 대전환) 시대를 뒷받침하는 것은 결국 연산 능력이며, 이를 확보하는 국가만이 기술 주권을 지킬 수 있다.

이런 맥락에서 APEC 기간에 성사된 그래픽처리장치 GPU 26만 장 확보는 단순한 컴퓨팅 장비 도입이 아니다. 대한민국이 본격적인 글로벌 AX 경쟁의 전면에 등장한다는 신호이며, 그동안 병목이었던 학습 자원 부족 문제를 해소하는 결정적 전환점이다. 국내 연구자와 스타트업도 이제 세계적 수준의 모델 개발과 연구 환경에 접근할 수 있는 기반을 갖추게 되었다.

GPU는 초대규모 모델의 학습을 실질적으로 가능하게 하는 병렬 연산 엔진이다. 대규모 연산을 동시에 처리해 학습 시간을 단축하고, 모델의 성능을 끌어올리는 데 핵심 역할을 한다. 그동안 국내에서는 GPU 자원이 부족해 대형 언어모델<sup>LLM</sup> 개발 등에서 구조적 제약이 컸지만, 이번 대규모 컴퓨팅 자원 확보로 본격적인 연구·개발 경쟁력이 확보되었다. 한편, 네트워크처리장치<sup>NPU</sup>는 동일 연산 대비 전력 소모가 낮아 대규모 추론 서비스 운영 비용을 획기적으로 줄일 수 있다. AI 서비스 운영 단계에서는 전력 효율과 처리 속도를 극대화하는 NPU가 핵심이다. 정부와 정보통신기획평가원<sup>ITP</sup>이 2018년부터 꾸준히 투자한 결과, 리벨리온·퓨리오사AI·딥엑스·모빌린트 등 국내 기업은 데이터센터용부터 엣지·산업특화형까지 다양한 NPU를 상용화 수준으로 발전시켰다.

GPU 확보와 국산 NPU 육성은 학습-추론-응용으로 이어지는 한국형 풀스택 구조를 가능하게 한다. 학습 단계에서 대규모 GPU 클러스터가 모델 개발을 지원하고, 추론

단계는 국산 NPU가 실제 산업 서비스의 효율을 극대화하며, 응용 단계에서 산업 현장에 확산된 서비스가 시장 확장을 견인할 수 있다. 이 구조가 자리 잡으면 한국은 AI 학습·추론 인프라의 기술 자립 기반을 확보하고, 전력 효율·비용 우위까지 갖춘 독자적 생태계를 구축할 수 있다. 특히 NPU 중심의 추론 인프라가 확산되면 산업별 AI 전환 속도가 한층 빨라지고, 한국만의 고성능·고효율 모델 운영 체계가 강점으로 부상할 것이다.

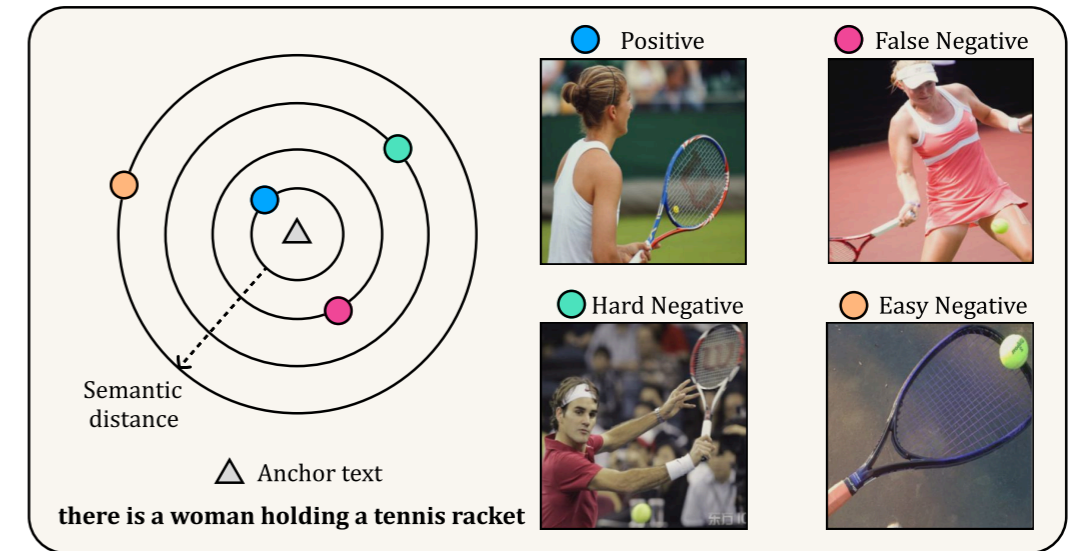
물론 AI 인프라만 갖췄다고 혁신이 자동으로 일어나는 것은 아니다. 중요한 것은 누가 인프라를 활용해 새로운 서비스를 만들고 산업을 확장하느냐이다. GPU는 연구·개발을 위한 개방형 클러스터로 운영하고, NPU는 제조·로봇·헬스케어·공공 분야 등 실제 산업 서비스에 적용해 국내 기술이 산업 전반으로 확산되어야 한다. 이를 위한 전력망 확충, 규제 정비, 고급 AI 인재 양성 같은 기반 체력을 함께 강화해야 한다. GPU 기반의 학습 생태계, NPU 중심의 추론 생태계를 바탕으로 응용 서비스 생태계가 유기적으로 결합할 때 비로소 한국은 글로벌 AI 경쟁에서 독자적 위치를 확보할 수 있다.

대한민국은 지금 AI 대전환의 문턱에서 현실적이고 거대한 기회를 마주하고 있다. GPU × NPU 투트랙 전략은 기술 자립과 산업 확장의 길을 동시에 여는 결정적 전략이며, 한국형 AI 생태계를 글로벌 시장으로 확장할 수 있다. 나아가 AI 인프라·모델·반도체를 결합하는 국가 전략은 기술패권 경쟁의 새로운 판도 속에서 한국의 주도권을 강화할 것이다. 기술 혁신과 안전·신뢰의 균형을 유지하며 AX 시대의 도전을 정면으로 받아들이 때, 대한민국의 AI 미래는 더욱 단단해질 것이다. 우리나라가 인공지능<sup>AI</sup> 분야에서 세계 3대 강국으로 도약하겠다는 국가 전략을 세웠는데, AI G3를 향한 힘찬 도약을 기대해 본다. [NAIRL](#)

## 이병준

고려대 인공지능학과 교수

## 배우면서 더 잘 배우는 법도 같이 배우기



시각 언어 대응 공간 예시. 기준점이 되는 텍스트에서 의미적 거리가 가까운 순으로 양성 쌍, 거짓 음성 쌍, 어려운 음성 쌍, 쉬운 음성 쌍 순서대로 위치한다. 효율적인 학습을 위해서는 거짓 음성 쌍과 어려운 음성 쌍을 정확하게 구분할 수 있어야 한다.

### 전 세계적인 컴퓨팅 부족, 이제 핵심은 '학습 효율성'이다

뉴럴 스케일링 법칙은 AI 시스템의 성능이 훈련 데이터, 모델 매개변수, 컴퓨팅 자원이 증가함에 따라 예측 가능하게 향상되는 현상을 말한다. 이 법칙에 따라 글로벌 AI 리더들은 AI 투자에 총력을 다하고 있지만, GPU, 메모리, 에너지 등 인프라 전반에서 병목이 발생하고 있어 갈수록 스케일 중심 경쟁보다는 효율성 중심으로의 전환이 강조되고 있다.

이러한 측면에서, 우리는 지금까지 모종의 관행이라고 생각되었던, 그저 해왔던 그대로 AI 학습 루틴들을 하나하나 점검하고 의구심을 가져볼 필요가 있다. 둘다리도 하나씩 두드려보고 개선하는 와중에서 큰 효율성 개선이 올지 모르니 말이다.

### 이 음성<sup>negative</sup> 샘플이 진짜 음성 샘플인가?

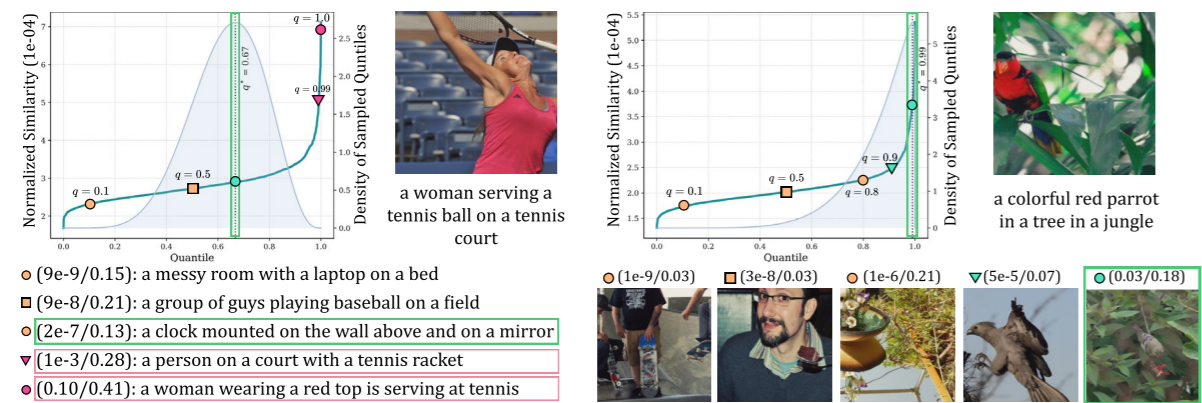
시각 언어 사전학습<sup>VLP</sup>는 이미지-텍스트 검색, 시각 질문 응답, 시각 추론 등 다양한 태스크에서 높은 성과를 보여주고 있다. VLP의 핵심 아이디어는 양성<sup>positive</sup> 텍스트-이

미지 쌍과 음성<sup>negative</sup> 텍스트-이미지 쌍을 준비해서 이를 대조하며 학습<sup>contrastive learning</sup>해 나가는 것이다. 하지만 이미지와 텍스트의 다대다 대응으로 인해, 실제 음성 쌍을 정확히 구별해내는 것은 그다지 자명하지 않다.

학습의 효율성을 위해서는 최대한 양성에 가까운 음성 쌍<sup>hard negative</sup>을 학습에 사용해야 하는데, 대부분의 기존 VLP 모델에서는 이렇게 하다 실제로는 양성 쌍인 샘플을 음성으로 잘못 취급하는 거짓 음성 쌍<sup>false negative</sup>이 생겨나는 일이 빈번하다. 이는 학습되는 텍스트-이미지 대응의 질을 저하시킨다. 게다가 현재 학습이 초기·중기·후기 중 어떤 단계에 있는지, 그리고 얼마나 복잡한 이미지와 텍스트 쌍을 고려하고 있는지에 따라 거짓 음성 쌍을 골라낼 수 있는 텍스트-이미지 유사도 임계치는 천차만별이기에 이 문제를 더욱 어렵게 만든다.

### FALCON: 음성 쌍 구분법을 배우면서 VLP 하기

지속적으로 변화하는 학습 과정 중, 거짓 음성 쌍<sup>false negative</sup>을 고르지 않으면서 최대한 어려운 음성 쌍<sup>hard</sup>



학습된 FALCON이 골라내는 음성 쌍의 예시. 왼쪽은 이미지 기준점에서 텍스트를 골라내는 예시이고, 오른쪽은 텍스트 기준점에서 이미지를 골라내는 예시이다. 각 기준점마다 최적의 어려운 음성 쌍에 해당하는 유사도는 매우 다르지만, 스케줄러가 정확하게 최적의 어려운 음성 쌍을 골라내는 것을 확인할 수 있다.

negative을 정확하게 골라내는 것이 결국 VLP 성능의 핵심이다. 이에 착안한 고려대 연구팀은 'False-Negative Aware Learning of Contrastive Negatives<sup>FALCON</sup>'라는 새로운 학습 구조를 제안했다.

FALCON의 핵심은 '학습 배치 구성 시 음성 쌍의 난이도를 학습 기반으로 조절하는 것'이다. 음성 쌍의 난도를 너무 높게 잡으면 거짓 음성 쌍이 너무 많이 나올 수 있고, 너무 낮게 잡으면 학습에 도움이 되지 않는다. 그러나 적절한 음성 쌍의 난이도는 학습 과정에 따라, 그리고 기준

점에 따라 계속해서 변화한다. 따라서 필요한 음성 쌍의 난이도를 매 경우마다 정확하게 정하는 건 불가능하지만, 대신 기준점에 따라 필요한 음성 쌍의 난이도를 결정하는 스케줄러를 두고, 해당 스케줄러의 각 결정에 따라 VLP의 손실함수가 어떻게 변화하는지를 관측하며 스케줄러를 VLP와 동시에 학습시키는 방식이다. 최종적으로 VLP 학습에 도움이 되는 방향으로 음성 쌍을 골라낼 수 있게 해주는 것이다.

스케줄러 학습의 효율성도 중요하다. 빠르게 스케줄러

연구소에서 개발중인 모습



고려대 의사결정연구실을 이끌고 있는 이병준 교수가 연구원들과 함께했다.

학습이 완료되어야 VLP 학습의 효율성을 변화하는 학습 과정에 따라 맞춤으로 끌어올릴 수 있다. 연구팀은 이에 대한 해법으로 다양한 음성 쌍 후보들과 기준점 사이의 유사도 분포만을 스케줄러 입력으로 받는 가벼운 스케줄러를 고안했다.

#### 단순히 개선된 음성 쌍 샘플링을 넘어서

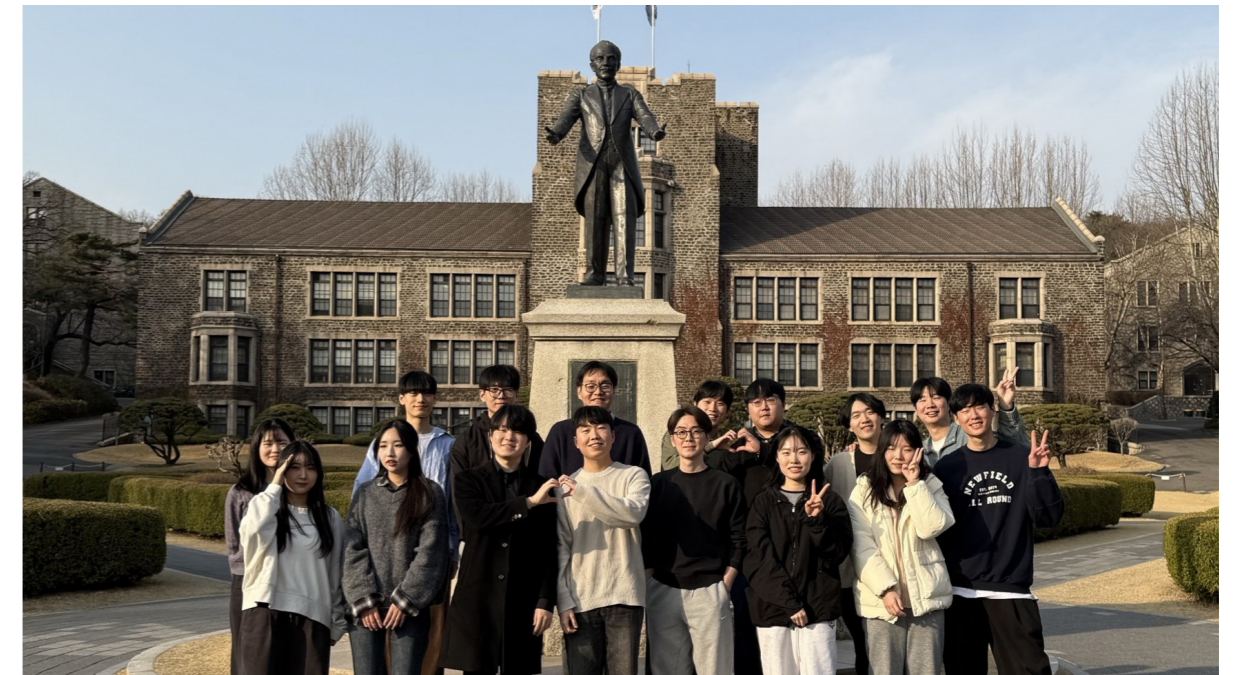
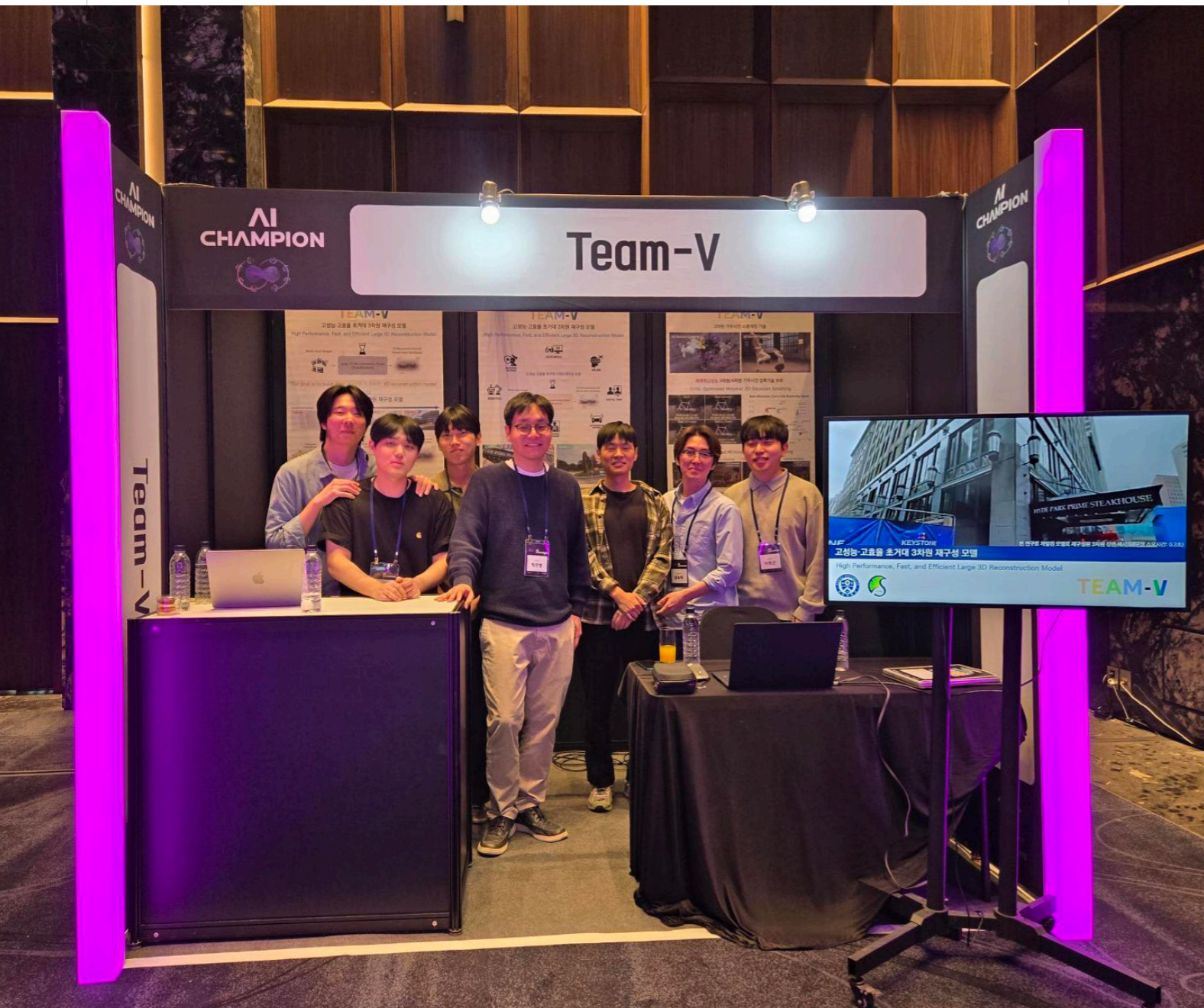
본 연구에서는 VLP에서 더욱 개선된 음성 쌍을 얻어내어 학습 효율성을 올리는 방향으로 시연되었지만, '학습에서의 최종 손실함수의 개선 폭을 보고 업데이트되는 스케

줄러의 도입'이라는 프레임워크는 현 도메인에 국한되는 구조가 아니다. 기계학습의 도메인과 각 도메인에서의 학습 알고리즘은 매우 다양하고 관행처럼 굳어 내려오는 학습 테크닉은 매우 많다. 딥러닝에서 조절해야 하는 알고리즘 하이퍼파라미터가 매우 많다 보니 기존 성공적이었던 예시들에서 많은 부분을 그대로 가져오는 경우가 많고, FALCON은 이를 효율화할 수 있음을 보여주는 하나의 예시에 불과하다. 뉴럴 스케일링 법칙을 극복하기 위해, 또 어떤 도메인에 어떤 개선점을 찾을 수 있을지, 본 고려대 연구진은 지금도 열심히 연구 중이다. [NAIRL](#)

박은병

연세대 첨단컴퓨팅학부 교수

## 초거대 3D 재구성 모델



연세대 V-Lab 연구실을 이끌고 있는 박은병 교수(뒷줄 가운데)가 연구원들과 함께 교내에서 찍은 단체 사진.

### 3D 재구성의 새로운 관점, 일반화된 3D 재구성

디지털 트윈, 로봇틱스, 자율주행, 확장현실XR 등 현대의 주요 기술들은 현실 공간을 빠르고 정확하게 디지털 3차원 형태로 재구성하는 능력이 필요하다. 전통적인 3D 재구성은 여러 시점의 이미지를 바탕으로 점군point cloud, 메시mesh, 또는 볼류메트릭voxel 형태의 3D 모델을 만들어내는 과정으로, 수십 년 동안 컴퓨터 비전과 그래픽스 연구의 중심에 있었다.

그러나 기존 방식은 대체로 시간을 많이 소모하고 계산량이 높은 반복적 최적화iterative optimization가 필요했다. 예를 들어 NeRFNeural Radiance Fields 혹은 3DGS3D Gaussian Splatting 기반 모델은 한 장면scene을 재구성하기 위해 수십 분에서 수 시간의 최적화 과정이 필요하다. 이러한 방식은 고정된 장면이나 오프라인 렌더링에는 적합하지만, 로봇 실시간 인지, 대규모 공간 스캔, AR·VR에서의 즉각적인 상호작용에는 한계가 있다.

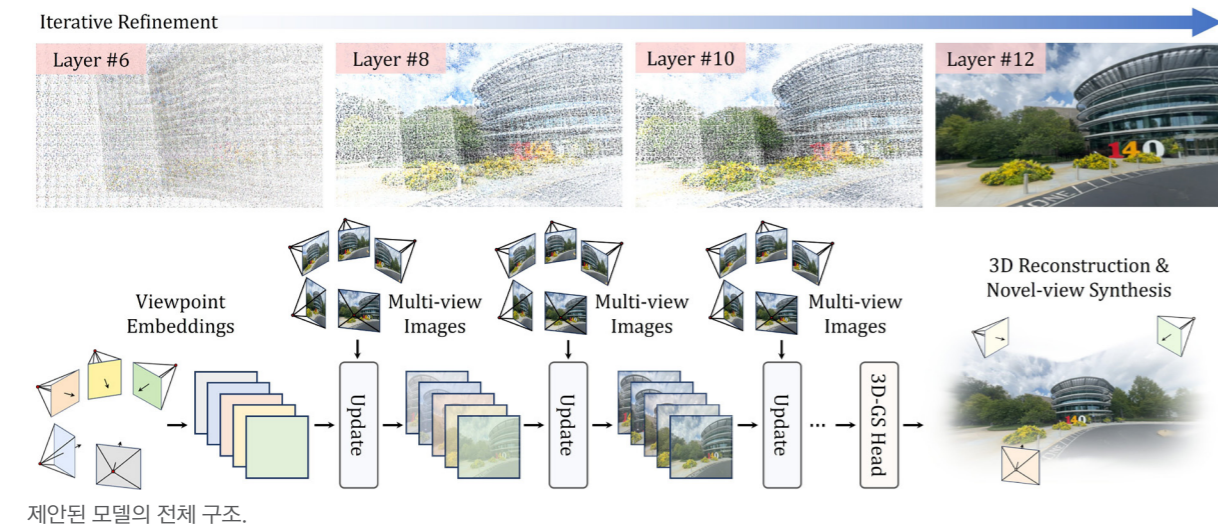
최근에는 대규모 데이터셋과 강력한 신경망 모델을 활

용해 하나의 학습된 모델이 여러 장면을 빠르게 재구성하는 일반화된generalizable 3D 재구성 모델이 각광받고 있다. 이러한 연구 방향에서 feed-forward 3D reconstruction, 즉 입력 이미지를 한 번의 전 방향 계산으로 즉시 3D 장면을 예측하는 패러다임이 새로운 돌파구로 등장했다.

### 왜 Feed-forward 방식인가?

Feed-forward 방식의 가장 큰 장점은 속도다. per-scene 최적화 기반 모델이 매 장면마다 반복 계산을 수행하는 반면, feed-forward 방식은 이미지를 입력하면 즉각 3D 재구성 결과를 획득할 수 있다. 이로 인해 다음과 다양한 응용 분야에 사용될 수 있다.

① 실시간 로봇틱스 및 자율주행: 로봇이나 자율주행차는 주변 환경을 즉시 이해해야 한다. 수 초에서 수십 초의 계산 지연은 치명적인 오류로 이어질 수 있다. Feed-forward 모델은 초 단위 이하의 속도로 환경을 인식할 수 있어 실전 배치가 가능하다.



② AR·VR에서의 즉각적인 공간 이해: 사용자가 스마트폰이나 HMD로 방을 스캔할 때, 기다리지 않고 실시간으로 장면이 3D로 변환된다면 완전히 새로운 사용자 경험이 가능하다.

③ 대규모 3D 지도 제작의 효율성 향상: 대규모 실내·실외 공간을 스캔하는 기업이나 기관은 수백~수천 개의 장면을 처리해야 한다. 다수의 이미지를 효율적으로 처리할 수 있는 기술을 개발한다면 개발 및 운영 비용을 획기적으로 줄일 수 있다.

#### 또 너야? Transformer

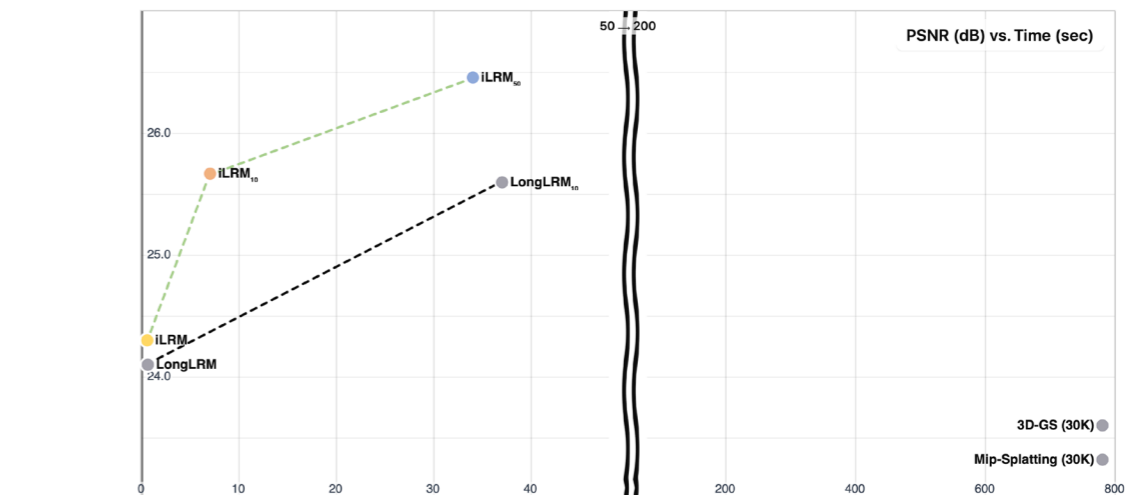
최근 Dust3R를 비롯한 여러 후속 모델들은 다시점 이미지를 통합해 3D 구조를 복원하는 데 Transformer 기반 아키텍처가 매우 효과적임을 보여주고 있다. ‘어디서나 Transformer!’라고 느껴질 정도지만, 실제로 3D 재구성 문제에서는 Transformer가 갖는 자연스러운 시점 간 상호작용 능력이 강력한 무기다. 특히 Transformer의 self-attention은 이미지 간의 관계를 세밀하게 파악하고, 그 과정에서 시점 간 대응(correspondence)을 포착한다. 이는 3D 재구성의 본질적 요소와 맞닿아 있어, 다시점 구조 이해라

는 문제에서 Transformer가 선택되는 것은 우연이 아니다. 그러나 이러한 장점에도 불구하고, 기존 multi-view Transformer는 다음과 같은 구조적 한계를 갖고 있다.

① 계산량 폭증: Attention의 속명으로, 이미지 수가  $N$ 이면 multi-view attention의 계산량은  $O(N^2)$  수준으로 폭발적으로 증가한다. 예를 들어 입력 이미지가 100장만 되어도, Attention 연산은 수십만 쌍의 관계를 한꺼번에 고려해야 하는 상황이 된다. 이는 메모리와 시간 측면 모두에 큰 부담을 준다.

② 대규모 장면에서는 곧바로 한계에 도달: 현실의 건물, 실내 공간, 건축물 단위의 스캔을 위해서는 수백 장의 이미지가 필요하다. 그러나 기존 모델들은 GPU 메모리가 부족해지거나 속도가 극단적으로 느려지며 확장성(scalability)에서 명확한 한계를 드러낸다.

결국 기존의 multi-view Transformer들은 강력한 표현력을 가지고 있음에도 불구하고, 실시간·대규모 3D 재구성이라는 목표에는 부적합한 부분이 있었다. 이러한 이유로, Transformer의 장점은 취하되 구조적 비효율을 해결할 새로운 방식이 필요해졌다.



기존의 최고 성능 모델 LongLRM과의 성능 비교. x-축은 재구성 소요 시간(초), y-축은 시점 합성 성능 PSNR을 각각 나타낸다.

#### 확장성이 뛰어난 multi-view Transformer

제안하는 방법은 기존의 feed-forward 3D Gaussian Splatting 기반 재구성 모델들이 갖고 있던 가장 큰 약점(입력 뷰 수가 늘어날수록 폭증하는 비효율적인 multi-view attention 구조)을 해결하기 위해 고안된 새로운 대규모 3D 재구성 프레임워크다. 기존 방법의 대부분은 고해상도 다중 뷰 이미지의 모든 토큰을 서로 Attention 시키는 방식에 의존해, 계산량이 뷰 수 및 해상도에 따라 이차적으로 증가(Quadratic complexity)하는 구조적 한계를 갖는다. 제안하는 방법은 이 문제를 해결하기 위해 ① 입력 이미지와 3D 표현(gaussians)을 완전히 분리(decoupling)하고, ② multi-view attention을 두 단계(per-view cross-attention → compact viewpoint self-attention)로 분해해 계산량을 획기적으로 줄였으며, ③ 매 레이어에서 ‘업데이트 블록’을 반복 적용하는 iterative refinement 구조로 기존의 정교한 3D 일관성을 확보한다. 이를 통해 iLRM은 많은 입력 뷰를 처리하면서도 실시간에 가까운 속도로 고품질 3D 재구성을 수행할 수 있는 모델이다.

또한 viewpoint embedding에서 직접 3D Gaussian을 생성하는 독창적인 방식과 다양한 설계를 통해 고해상도 시각 정보를 유지하면서도 컴팩트한 3D 표현을 생성한다. RE10K, DL3DV 등 대규모 벤치마크에서 기존 feed-forward 모델은 물론 일부 최적화 기반 3DGS 방법까지

능가하는 품질과 속도를 보이며, 특히 많은 수의 입력 뷰를 사용할수록 더 높은 효율성을 발휘해 확장성 측면에서 새로운 표준을 제시한다. 이러한 특징 덕분에 iLRM은 향후 pose-free, dynamic reconstruction 등 다양한 3D 분야로 확장 가능한 다음 세대 대규모 3D 재구성 아키텍처의 기반이 될 것으로 기대된다.

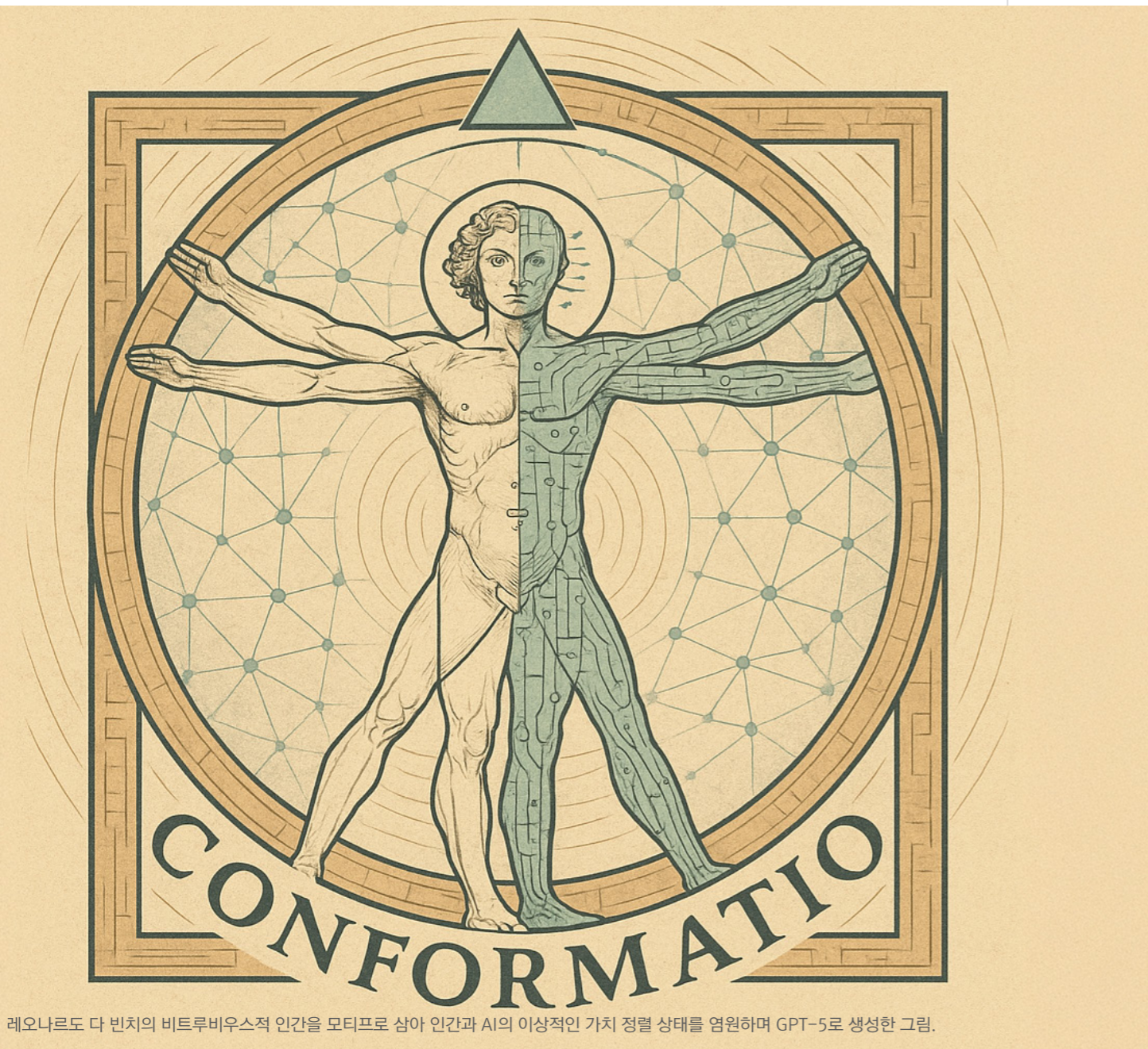
#### 앞으로의 확장 가능성: 기초 모델로서의 잠재력

현재는 정적 장면을 대상으로 한 feed-forward 3D 재구성에 초점을 맞추고 있지만, 제안된 구조는 훨씬 더 넓은 문제 설정으로 자연스럽게 확장될 수 있는 잠재력을 지니고 있다. 시간 정보를 통합하면 동적 장면 재구성, 동작 캡처, dynamic NeRF와 같은 시계열 기반 3D 이해로 확장할 수 있으며, 깊이·LiDAR와 같은 기하정보를 예측하도록 하여, 정밀한 메트릭 수준의 3D 복원, 대규모 매핑, SLAM과 같은 기하 중심 문제에도 적용 가능하다. 또한 제안된 방법의 모듈성과 확장성은 대규모 데이터 기반의 범용 multi-view foundation model로 발전할 수 있는 기반을 제공한다. 다양한 모달리티(영상·포인트·텍스트)를 통합하는 방식으로 확장하면, 복잡한 3D 환경을 포괄적으로 학습하고, 다양한 3D 이해와 생성 과제를 하나의 모델이 처리하는 방향으로 진화할 수 있다. [NAIRL](#)

박상돈

POSTECH 인공지능대학원 교수

## 인간과 AI의 가치 정렬



레오나르도 다 빈치의 비트루비우스적 인간을 모티브로 삼아 인간과 AI의 이상적인 가치 정렬 상태를 염원하며 GPT-5로 생성한 그림.

### AI를 믿고 사용하려면? AI 정렬 문제

2025년 현재 생성 AI가 사람의 수준에 견주어질 정도로 발전하였고 일반인의 일상 생활에 깊게 침투되어 있다. 이런 사용률의 증가는 AI의 신뢰성에 대한 갈증을 불러일으키는 계기가 되고 있다. 가령, AI가 거짓된 대답을 자신 있게 말하는 환각<sup>hallucination</sup> 효과를 비롯하여 AI가 알려진 가장 강한 신경독 중의 하나의 다이메틸수은<sup>Dimethylmercury</sup>을 만드는 법을 알려주는 등의 안전성 문제가 연구자와 일반인의 걱정을 불러일으키고 있다.

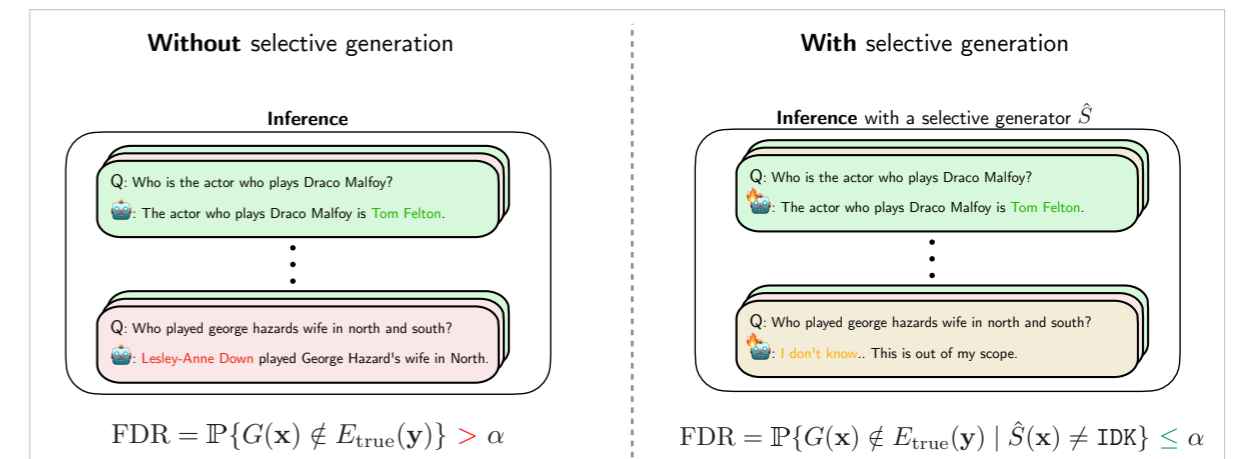
이런 신뢰성 문제를 해결하여 AI의 가치와 인간의 가치를 맞추는 문제를 AI 정렬<sup>AI Alignment</sup> 문제라고 부른다. 이 문제를 해결하기 위해서 인간 피드백을 통한 강화학습<sup>RLHF</sup>으로 대표되는 강화학습을 통한 미세 조정<sup>finetuning</sup>과 라마การ์ด<sup>Llama Guard</sup>로 대표되는 후처리 과정이 사용됨에 따라 AI 정렬 문제가 점차 해결되고 있다. 하지만 과연 이런 방법으로 정렬된 AI를 우리가 믿고 사용할 수 있을까?

이 질문에 대해서 대답하기 위해 우리가 현재 신뢰하며 사용하는 소프트웨어를 떠올려 보자. 아마도 가장 먼저 떠오르는 소프트웨어는 암호 알고리즘일 것이다. 그럼 왜 암호 알고리즘을 믿고 사용할 수 있을까? 암호 알고리즘은 다양한 실패를 통해 정립된 다음의 세 가지 원칙을 따라

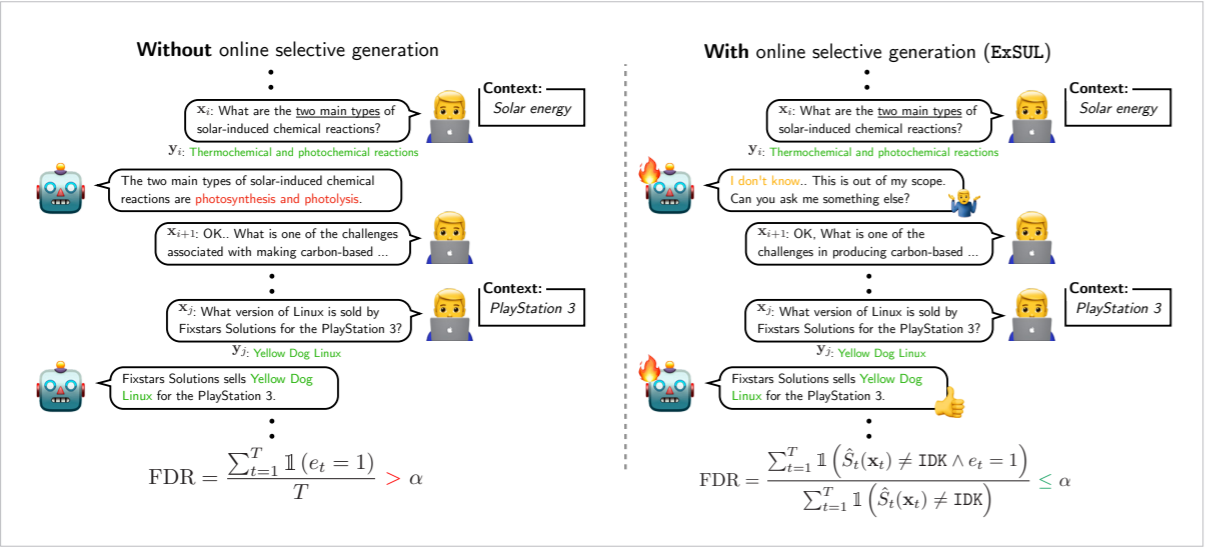
개발되어서 그 신뢰성이 보장된다고 알려져 있다. ‘원칙 1. 암호 알고리즘에서 고려하는 보안 목적을 정의’, ‘원칙 2. 암호 알고리즘이 사용되는 환경에 대한 가정을 명시’, ‘원칙 3. 가정된 환경에서 암호 알고리즘이 정의된 보안 목적을 달성할 수 있다는 증명을 제시’. 본 연구팀은 AI 학습 알고리즘의 개발도 이 원칙에 따라 이론적인 보장을 줄 수 있어야 믿고 사용할 수 있다고 생각한다. 이 여정을 위한 시작점으로 선택적 생성<sup>selective generation</sup>을 포함하여 이론적 보장을 할 수 있는 알고리즘 개발이 학계에서 활발히 진행되고 있다.

### 선택적 생성을 통한 LLM 환각 효과 제어

거대 언어 모델<sup>LLM</sup>의 환각 효과 제어란 언어 모델의 오류를 일정 수준으로 제어하는 것을 말한다. 이는 일반적으로 강화학습을 통한 미세 조정을 통해서 모델 성능을 높이는 방식으로 부분적으로 해결할 수 있지만, 모델의 크기로 인하여 학습 비용 및 시간이 많이 든다는 어려움이 있다. 이에 더해 희망 오류 수준을 갖도록 학습된 모델을 제어하기 힘들다는 단점이 있다. 이 문제를 해결하기 위해서 POSTECH 박상돈 교수 연구진은 고전적인 선택적 분류기<sup>selective classifier</sup>를 언어 모델을 위한 선택적 생성기<sup>selective</sup>



‘선택적 생성’을 통해 LLM의 환각 효과를 제어하는 예시.



‘온라인 선택적 생성’을 통해 LLM의 환각 효과를 제어하는 예시.

generator로 확장하여 학습하는 방법을 최초로 제시하였다.

여기서 선택적 생성기는 주어진 질의에 대하여 거대 언어 모델의 생성 대답이 올바르지 못하다고 판단될 경우 생성 대답을 회피abstain하는 생성기이다. 이런 선택적 생성을 통해서 모든 대답이 아니라 선택된 대답에 대한 조건부 오류를 제어할 수 있는 가능성을 제공해 준다.

선택적 생성기를 학습하기 위해서는 두 가지 도전 과제가 있다. 우선 선택적 언어 생성 모델의 학습을 위해서 자연어로 표현된 대답과 정답의 의미적 동치 관계를 판단하는 것은 극히 어려운 문제이다. 이를 위해서 연구진은 대답과 정답의 수반 관계entailment를 활용하여 동치 관계를 판단하는 기준을 제시하였다. 하지만 수반 관계를 통해 학습하려면 사람의 도움을 통해 라벨label을 얻어야 한다. 이는 많은 시간이 걸리고 비용이 비싼 작업이므로 라벨이 없는 데이터를 활용할 수 있는 준지도 학습semi-supervised learning 방법을 최종적으로 제시하였다. 이렇게 제시한 선택적 생성 모델 학습 기법 및 조건부 오류에 대한 이론적 보장에 대한 연구 결과는 최대 인공지능 학회 중 하나인 ‘NeurIPS 2024’에 spotlight 논문으로 게재됨으로써 학계의 큰 관심을 받았다.

온라인 선택적 생성을 통한 LLM 환각 효과 제어

선택적 언어 생성 모델은 대답이 확실하지 않은 경우에 대답을 회피하는 능력이 있지만, 이것은 학습의 결과이다. 모델이 실제 사용되고 학습되는 환경을 고려해 보면 온라인 학습이 필요하지만, 이는 이전 연구에서 전혀 다루지 않은 영역이다. 특히, 선택적 언어 생성 모델을 학습하기 위해서 대답의 올바름을 사용자에게 물어보는 온라인 학습 환경을 고려해 보자. 대답이 틀린 경우 사용자는 올바른 대답을 제공해 주는 것보다 틀렸다는 피드백feedback을 주는 것이 더 용이하다. 또한 물어 보는 질문 및 사용자의 피드백 분포는 시간에 따라 변하기 마련이다. 이런 극한의 환경에서 선택적 생성 모델을 온라인 학습시키는 것은 난해한 문제이고 개척되지 않은 문제이다.

POSTECH 박상돈 교수의 연구진은 해당 문제를 해결하기 위해서 고전적인 적대적 밴딧adversarial bandit 문제를 활용하는 방법을 채택하였다. 즉, 선택적 생성 모델 학습 문제를 적대적 밴딧 문제로 변환하여 임의의 적대적 밴딧 알고리즘을 활용하여 선택적 생성 모델을 학습시키는 방법을 제안하였다. 여기서 조건부 오류에 대한 보장을 하기 위해서 적대적 밴딧이 보장하는 후회값regret을 조건부 오

류로 변환하는 보조 정리도 세계 최초로 제안하였다. 하지만 고전적인 적대적 밴딧 알고리즘은 선택적 생성 모델 학습 문제의 특성을 사용하지 않는 문제가 있다. 연구진은 이 특성을 활용하여, 선택적 생성에 특화되어 피드백이 효율적인 알고리즘을 최종적으로 제안하였다. 이 방법은 선택적 생성 모델에 대해 적대적 밴딧 피드백을 이용한 최초의 학습 방법으로 보장 가능한 AI 정렬 문제를 효과적으로 해결할 수 있는 단초가 되는 연구결과이다.

LLM을 넘어 Physical AI의 신뢰성을 위하여

앞으로 연구팀은 가장 기본적인 생성 AI인 LLM을 넘어

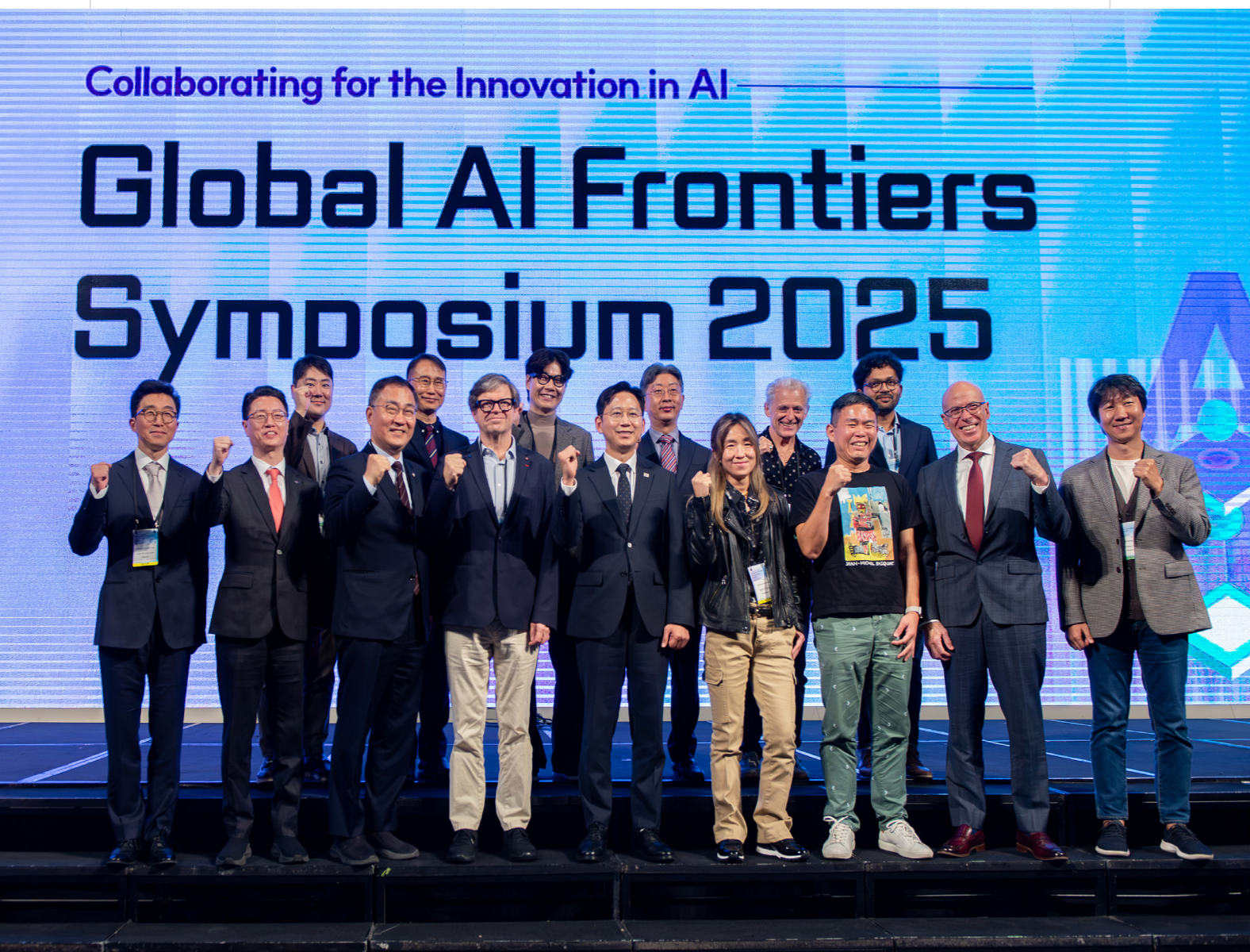
서 Physical AI의 정렬 문제에 힘을 계획이다. 현재 사람의 두뇌에 해당하는 LLM의 발전은 급진적이지만 손과 발에 해당하는 물리 세계로의 지능 확장은 천천히 진행 중이다. 더 나아가 물리 세계와 상호 작용하는 Physical AI는 인간 및 다른 AI와의 관계를 생각해야 하므로 사람의 가치에 따르는 엄밀한 AI를 개발하는 문제가 더욱더 중요해질 것이다. 이를 위해 선택적 생성 기법을 포함하여 신뢰성을 증명 가능한 AI 정렬 방법을 개발함으로써 믿고 사용 가능한 Physical AI를 개발하려는 목표를 가지고 있다. [NAIRL](#)



POSTECH 머신러닝 랩의 박상돈 교수(아랫줄 가운데)와 연구원들.

안 르쿤·최예진 교수가 제시한  
차세대 AI의 미래

“AI,  
이제는 세계를 이해해야 한다”



2025년 10월 27일 서울 용산에서 열린 ‘AI 프론티어 국제 심포지엄 2025’에는 세계적 석학들이 모여 생성형 AI 이후의 방향성을 논의했다. 기조강연에 나선 메타 수석 AI 과학자 안 르쿤(Yann LeCun) 뉴욕대 교수와 스탠퍼드대 최예진 교수는 각각 ‘월드 모델 훈련(Training World Models)’과 ‘생성형 AI의 민주화: 스케일링 법칙을 초월하여’를 주제로 발표했다.

이번 심포지엄은 과학기술정보통신부와 정보통신기획평가원(ITP)이 주관하고, 국가AI연구거점(NAIRL)과 글로벌AI프론티어랩이 공동 주최했다. AI의 미래 연구 방향, 성과 공유, 국제 협력 방안을 논의하는 자리였다.

#### 르쿤 교수, “언어 예측만으론 지능에 도달하지 못해”

르쿤 교수는 현재의 거대언어모델(LLM)이 텍스트 예측에는 능하지만, 현실 세계의 물리적 구조나 인과관계를 이해하지 못한다고 지적했다. 그는 “AI가 수학 문제는 잘 풀지만 물건 하나 제대로 못 잡는다”며 모라벡 역설을 인용해, 단순 작업이 오히려 어려운 AI의 한계를 꼬집었다. 이 한계를 넘는 방법으로 그는 ‘월드 모델(World Model)’을 제안했다. 이는 AI가 세상의 구조와 인과를 스스로 학습하며 예측 능력을 기르는 방식이다. 특히 그는 메타에서 개발 중인 JEPA(Joint Embedding Predictive Architecture) 구조를 소개하며, 기존 픽셀 예측 방식보다 의미 기반 예측이 훨씬 효율적임을 강조했다.



그는 궁극적으로 인간처럼 사고하는 AI(AGI)보다 실용적인 ‘기계지능(AMI)’의 발전이 중요하다고 덧붙였다. 즉, 특정 분야에서 인간보다 뛰어난 자율 AI가 현실적인 목표라는 설명이다.

#### 최예진 교수, “AI는 인간과 함께 작동해야 한다”

최예진 교수는 AI 기술의 민주화와 공정한 접근성을 강조했다. 그는 “현재 AI 연구가 초거대 모델과 GPU 자원 경쟁에 치우쳐 있다”고 지적하며, “인터넷 데이터는 이미 고갈 상태”라고 진단했다. 대안으로 그는 AI가 스스로 사고하고 학습하는 능력과, 합성 데이터(synthetic data)를 통한 새로운 학습 방식이 필요하다고 강조했다. 그는 또한 소형 모델, 오픈소스 생태계, 국가 간 협력을 통해 AI 접근성을 높이는 방향이 중요하다고 밝혔다.

그는 “AI는 인간에 의해 설계되고, 인간을 위해 존재하며, 인간과 함께 작동해야 한다”고 말하며, 윤리적이고 신뢰 가능한 AI 생태계 구축의 필요성을 거듭 강조했다.

두 사람은 기조강연을 통해 한목소리로 다음 단계의 AI는 단순히 크기가 아니라 ‘세계에 대한 깊은 이해’가 핵심이라고 강조했다. 언어 예측만으로는 부족하며, AI가 감각 정보, 인과 추론, 자기 주도 학습, 윤리와 신뢰성 등을 바탕으로 현실 세계를 이해하고 작동할 수 있는 존재로 진화해야 한다는 메시지를 전했다. [NAIRL](#)

글로벌 석학들과 국내 전문가들이 그린  
AI의 미래 비전

## “AI는 공공 인프라가 돼야 한다”



지난 10월 27일 서울 용산에서 열린 ‘AI 프론티어 국제 심포지엄 2025’에 세계적 AI 석학들과 국내 연구자들이 모여 AI의 민주화, 공공성, 책임 있는 발전 방향을 논의했다. 이번 좌담에는 안 르콘 메타 수석 AI 과학자 겸 뉴욕대 교수, 최예진 스탠퍼드대 교수, 조경현 뉴욕대 교수(글로벌 AI프론티어랩 공동소장), 김기응 KAIST 교수(국가AI연구거점 센터장)이 참여했고, 배경훈 부총리 겸 과학기술정보통신부 장관이 사회를 맡았다.

“이런 원칙하에 오픈 코드와 데이터 공유 문화가 더욱 확산돼야 한다”고 주장했다. 그는 또한 소형 모델과 다국적 협업이 자원 격차를 줄이는 열쇠가 될 것이라고 덧붙였다.

### “혁신과 신뢰, 둘 다 포기할 수 없다”

조경현 교수는 시민의 목소리를 반영하는 AI 거버넌스의 필요성을 강조하며, “AI 설계와 규범에 일반 시민의 참여가 이뤄져야 한다”고 말했다. 이에 대해 최예진 교수는



왼쪽부터 사회자 배경훈 과기정통부 장관, 안 르콘 뉴욕대 교수(메타 수석 AI 과학자), 최예진 스탠퍼드대 교수, 김기응 KAIST 교수, 조경현 뉴욕대 교수.

### “AI는 모두를 위한 공공 자산이어야”

안 르콘 교수는 “AI가 특정 기업의 전유물이 아니라 ‘모든 인류를 위한 공공 인프라’가 돼야 한다”고 강조했다. 그는 인터넷처럼 AI 기술도 오픈 소스 생태계로 가야 하며, 독점은 장기적으로 기술 발전을 저해한다고 지적했다. 그는 특히 “미래의 AI 비서가 소수 기업에 의해 통제된다면 정보 불균형이 심화될 것”이라며 ‘AI 주권’의 필요성을 역설했다.

김기응 교수는 “AI 기술이 지속 가능한 방식으로 발전하려면, 투명성, 책임성, 재현 가능성 같은 원칙이 필수”라며

“자원이 부족한 국가일수록 ‘비전통적 데이터 + 협업’ 전략을 통해 AI 개발에 참여할 수 있다”고 제안했다.

최 교수는 또한 “AI 생태계의 민주화를 위해 합성 데이터, 오픈 협업, 경량 모델이 핵심”이라고 강조하며, “AI는 인간을 위한 기술이며, 모두에게 열려 있어야 한다”고 말했다.

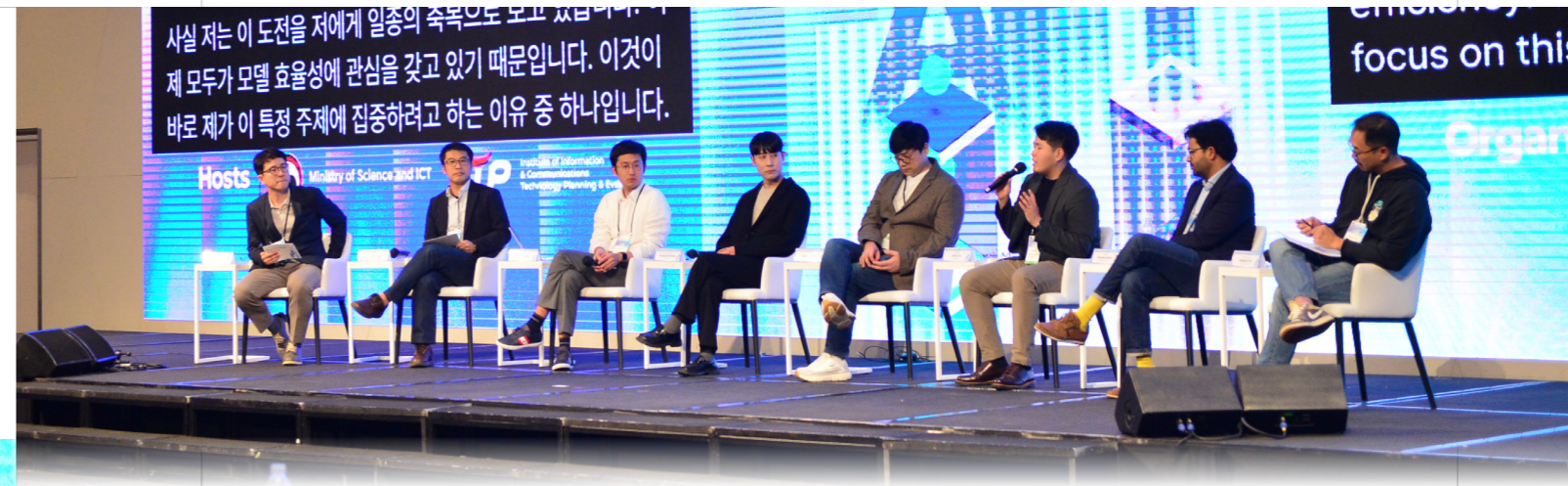
이번 좌담은 AI가 더 이상 단순한 기술이 아니라 사회 인프라와 민주적 자산으로서 자리매김해야 한다는 공감대를 형성했다. 배경훈 장관은 “오늘 논의가 한국의 AI 정책과 생태계를 더욱 성숙하게 발전시키는 이정표가 되길 바란다”고 말했다. [NAIRL](#)

차세대 AI를 위한 기초와 혁신

## “비용 넘는 창의적 해법 필요”



‘다음 세대 AI 기술의 기초’라는 세션1의 발표자들. 왼쪽 위부터 일본 이화학연구소RIKEN의 이치로 타케우치 수석연구원, 고려대 감태의 교수, 연세대 황성재 교수, KAIST 박노성 교수, POSTECH 이재호 교수, 프랑스 국립디지털과학기술연구소INRIA의 카르틱 알라하리 교수.



‘AI 프론티어 국제 심포지엄 2025’에서 오후에 진행된 세션1에서는 ‘다음 세대 AI 기술의 기초Foundations of Next-Generation AI’를 주제로 국내외 주요 연구자들이 AI의 기반 기술과 학문적 도전 과제들을 발표했다. 본 세션은 양은호 KAIST 교수가 좌장을 맡아 진행됐다.

### 과학적 기반·약자친화·스케일링의 한계 극복까지

이번 발표는 단순히 성능 향상이나 파라미터 확대를 넘어, 자원 제약 속에서의 효율성, 과학적 타당성, 그리고 범용성을 중심으로 AI의 다음 단계를 구체화하는 내용이 주를 이뤘다. 총 6명의 연구자가 나서서 최신 연구 내용을 공유했다.

일본 이화학연구소RIKEN의 이치로 타케우치 수석연구원은 ‘데이터 기반 과학에서의 연구 설계’를 주제로, 신뢰성과 투명성을 갖춘 AI 연구 프레임워크의 필요성을 강조했고, 고려대 감태의 교수는 ‘AI 기반 신약 개발’을 소개하며, 분자 특성 예측부터 분자 생성까지의 자동화 흐름을 설명했다. 이어 연세대 황성재 교수는 ‘비학습 기반Training-Free의 비전-언어 모델 해법’을 발표해, 거대 학습 없이도 활용 가능한 VLM 구조 분석을 소개했고, KAIST 박노성 교수는 ‘과학용 파운데이션 모델Scientific Foundation Models’의 개념과 활용 가능성을 설명했다. 또 POSTECH 이재호 교수는 ‘압축 스케일링 법칙 극복Overcoming the Compression Scaling Laws’을

주제로, 고성능을 유지하면서도 연산 비용을 줄이는 모델 압축 기법을 제안했고, 프랑스 국립디지털과학기술연구소INRIA의 카르틱 알라하리 교수는 ‘저자원 기반 비전-언어 모델’의 가능성을 타진하며, 자원 한계 속에서의 VLM 연구 방향을 논의했다.

### AI 연구 생태계, 어떻게 지속가능할 것인가

발표 이후 이어진 패널토론 ‘고비용 인프라 시대의 지속 가능한 AI 연구 생태계Sustainable AI Research Ecosystems in the Era of High-Cost Infrastructure’에서는 고성능 AI 모델 개발에 따르는 막대한 연산 자원 문제와 이를 극복하기 위한 학계·공공·산업계의 협력 방안이 집중 논의됐다. 이 토론은 양은호 교수가 사회를 맡았으며, 발표자 전원이 패널로 참여했다. 여기에 네이버클라우드의 윤상두 리더도 패널로 함께했다.

이번 세션은 단순히 기술적 성능의 경쟁을 넘어, 누구나 접근 가능한 AI, 사회적 책임을 갖는 AI, 그리고 과학적으로 타당한 AI라는 공통된 방향성을 확인하는 자리였다. 스케일링의 시대는 끝나가고 있으며, 이제는 ‘더 나은 설계’와 ‘더 똑똑한 학습’이 AI의 진짜 진화를 이끌 것이라는 뜻이다. 이날의 발표와 토론은 ‘다음 세대 AI’의 기초는 기술이 아니라 철학과 구조에 있다는 점을 다시금 일깨워주었다. **NAIRL**

## 실세계 자율성을 위한 피지컬 AI

“현실 세계에서 작동하는 AI,  
이제는 ‘물리적 지능’으로 간다”

‘실세계 자율성을 위한 피지컬 AI’라는 세션2의 발표자들. 왼쪽 위부터 프랑스 고등사범학교<sup>ENS-PSL</sup>의 장 풍스 교수, 연세대 이종민 교수, KAIST 심현정 교수, 고려대 최성준 교수, POSTECH 광수하 교수, UAE 무함마드 빈 자이드 인공지능대학<sup>MBZUAI</sup>의 이반 라프테프 교수.

인공지능<sup>AI</sup>이 실험실을 넘어 현실 세계에서 실제로 작동하려면 무엇이 필요할까? ‘AI 프론티어 국제 심포지엄 2025’에서는 오후에 이 물음에 답하기 위한 세션이 열렸다. ‘실세계 자율성을 위한 피지컬 AI<sup>Physical AI for Real-World Autonomy</sup>’라는 주제로 열린 세션2였다. 사회는 POSTECH 조민수 교수가 맡았다.

## 물리 세계에 강한 AI, 어떻게 만들까?

이 세션에서는 6명의 발표자가 나서서 물리적 환경에서 적응력 있는 AI 시스템 구현을 위한 최신 연구를 공유했다. 즉 시각지능, 자율학습, 강화학습, 로봇틱스 등 다양한 분야의 연구자들이 ‘현실 세계에서 작동하는 AI’를 위한 기반 기술과 구조적 과제를 논의했다.

프랑스 고등사범학교<sup>ENS-PSL</sup>의 장 풍스 교수는 ‘비대칭적 자기지도학습<sup>Non-Contrastive SSL</sup>’의 이중 관점을 소개하며, 사전 라벨 없이 AI가 스스로 특징을 학습하는 새로운 구조를 설명했고, 연세대 이종민 교수는 ‘부분 대칭성 환경에서의 강화학습’ 연구를 발표해, 현실의 비대칭적 물리 환경에서도 일반화 가능한 정책 학습 방식을 소개했다. 이어 KAIST 심현정 교수는 ‘범용 로봇틱스를 위한 강건한 3D 인지 기술’을 주제로, 자율주행 및 조작 로봇에 필요한 정밀 공간지각 기술을 발표했고, 고려대 최성준 교수는 ‘비전-언어-행동 모델의 추론 능력’에 주목해, 시각 정보와 언어 명령, 실제 동작을 연결하는 인공지능 구조를 제안했다. 또 POSTECH 광수하 교수는 ‘게이트-랭크<sup>Gated-Rank</sup> 적

응 방식에 기반한 비전 파운데이션 모델 강화’를 통해 다양한 환경 변화 속에서도 일관된 성능을 유지하는 시각 모델 개발 방안을 제시했고, UAE 무함마드 빈 자이드 인공지능대학<sup>MBZUAI</sup>의 이반 라프테프 교수는 ‘인간형 에이전트의 지각-행동 연계와 월드 모델 활용’을 발표해, 실제 환경에서 센서 기반으로 주변을 이해하고 행동으로 연결하는 로봇형 AI 시스템 구현을 강조했다.

## AI, 실험실 밖에서 어떻게 신뢰받을 수 있을까?

발표 이후 이어진 패널토론에서는 ‘현실 자율성의 조건: 피지컬 AI의 가능성과 한계’를 주제로 AI의 실제 환경 적용 가능성과 기술적 장벽이 논의됐다. 토론은 조민수 교수가 좌장을 맡았으며, 발표자 전원과 함께 포스코홀딩스 AI로봇융합연구소의 김용수 소장이 패널로 참여했다.

이번 세션은 AI가 단지 ‘두뇌(알고리즘)’만으로는 충분하지 않으며, 센서, 물리적 움직임, 맥락 이해 등을 통해 ‘몸’을 갖춘 지능으로 진화해야 한다는 점을 강조했다. 이런 접근은 자율주행, 로봇, 산업 현장 등에서 신뢰할 수 있는 AI의 미래상을 구체화하는 데 기여했다. AI는 이제 실험실 너머로 나가야 하고, 물리 세계와 상호작용하며 배우는 AI야말로 진정한 의미의 자율성과 지능을 보여줄 것이란 말이다. 이날 발표와 토론은 AI가 세상을 이해하고 스스로 움직일 수 있는 물리적 자율지능<sup>Physical AI</sup> 시대로 접어들고 있음을 강하게 시사했다. [NAIRL](#)



‘글로벌 AI 프론티어 심포지엄 2025’  
파트너 기업 전시 부스 스케치

## AI, 산업과 일상 속으로



‘글로벌 AI 프론티어 심포지엄 2025’에서는 국가AI연구 거점의 파트너 기업들이 다양한 산업 및 일상 영역에서 AI를 어떻게 실용적으로 적용하고 있는지를 보여주는 전시 부스를 운영해 주목을 받았다. 철강, 공공, 연구, 가정 등 전방위에 걸친 AI 활용 사례들이 생생하게 소개됐다.

### 산업을 움직이는 AI, 철강에서 공공까지

포스코홀딩스는 철강 제조와 의사결정 전반에 AI를 본격 도입한 사례를 전시했다. AI로봇융합연구소 김용수 소장이 과학기술정보통신부 배경훈 장관에게 부스를 직접 안내하며, “AI가 용광로 온도 제어, 코팅 두께 예측, 표면 결함 인식 등 주요 공정을 자동화하면서 생산성과 안전성을 동시에 높이고 있다”고 설명했다. 실제 현장에서는 4족 보행 로봇, 밀봉 로봇, 감시 드론 등이 위험한 작업을 대체하며, 안전한 작업 환경을 구현하고 있다. 또 AI는 글로벌 원자재 시장 예측, 고객 맞춤 소재 설계, 경영 전략 수립 등에도 적극 활용되고 있다.

포티투마루는 초거대 언어모델의 환각 문제를 줄이기 위한 검색증강 생성RAG42과 독해 AIMRC42 기술을 소개했다. 자체 경량 모델 ‘LLM42’를 기반으로 내부 데이터를 안전하게 활용할 수 있는 프라이빗 AI와 기업 맞춤형 전환<sup>AX</sup> 솔루션도 공개됐다. 김동환 대표는 “신뢰성과 확장성을 갖춘 글로벌 실용 AI 시장을 선도하겠다”고 밝혔다.

### 일상을 바꾸는 AI, 연구자와 가정의 변화

코르카는 연구자를 위한 AI 동료 ‘문라이트Moonlight’를 전시했다. 논문 탐색과 분석을 돕는 이 시스템은 실제 논문을 열고 사용자와 질의응답을 나누는 데모로 주목을 받았으며, 한 관람객은 “이거 없으면 논문 못 읽겠다”는 반응을 보이기도 했다. 정영현 대표는 “문라이트는 단순한 도우미를 넘어 AI 과학자<sup>AI Scientist</sup>, 혹은 과학자를 위한 AI<sup>AI for Scientist</sup>로 진화하고 있다”며 “앞으로 실험 설계와 사유의 흐름까지 지원할 수 있는 지능형 플랫폼이 되겠다”고 밝혔다.

LG전자는 ‘다정하게 살아가는 LG AI 홈Affectionately live in LG AI Home’이라는 주제로 온디바이스 AI 기술을 중심으로 한 스마트홈 미래를 제시했다. 클라우드가 아닌 기기 자체에서 AI 연산이 이뤄지는 구조로 빠르고 안전한 사용자 경험을 제공하며, 냉장고, 세탁기, 로봇청소기 등 가전이 사용자의 맥락을 이해하고 배려하는 존재로 진화하고 있음을 보여줬다.

이 외에도 네이버클라우드, LG에너지솔루션, 에버엑스, 베슬에이아이, 인이지, 플레이오니, 대동에이아이랩, HD 한국조선해양 등 다양한 기업들이 제조 효율화, 공공 인프라의 스마트화, 일상 속 편의 향상 등 각 분야에서 AI가 가져올 변화를 실감나게 제시했다. **NAIRL**



컴퓨터 비전 분야의 세계적 석학  
장 풍스 교수 인터뷰

“기초를 묻고, 이론의 밑그림을 그리는 것이야말로  
AI 시대의 연구”

장 풍스 Jean Ponce

- 프랑스 파리 고등사범학교 ENS-PSL 컴퓨터과학부 교수
- 뉴욕대학교 Courant 수학과학연  
구소 및 데이터사이언스센터 글  
로벌 석좌교수
- 파리 AI 융합연구소 PR[AI]RIE-  
PSAI 과학디렉터
- Enhance Lab 공동 창립자 및  
CEO



“AI를 ‘더 잘 작동하게’ 만들기보다는 AI가 왜 작동하는  
지를 이해하고 싶었습니다.” 파운데이션 모델이 연구와 산  
업 전반을 휩쓸고 있는 지금, 근본적인 원리와 역사적 시  
야를 되짚는 연구는 어떤 의미가 있을까? 컴퓨터 비전 분  
야의 세계적 권위자인 장 풍스Jean Ponce 교수는 AI가 어떻  
게 작동하는지를 깊이 탐색하는 연구가 여전히 중요하다  
고 말한다. 풍스 교수는 프랑스 파리 고등사범학교ENS-PSL  
컴퓨터과학과 교수이자 뉴욕대NYU 쿠란트Courant수리과학  
연구소 및 데이터사이언스센터 글로벌 석좌교수로 있으며,  
파리 AI 융합연구소 PR[AI]RIE-PSAI의 과학디렉터이자 스  
타트업 인핸스 랩Enhance Lab의 CEO를 맡고 있다. 서울에서  
열린 ‘글로벌 AI 프론티어 심포지엄 2025’에서 발표를 마친  
그와의 인터뷰를 통해, 컴퓨터 비전의 철학, 자기지도학습  
의 구조, 그리고 다음 세대 AI 연구의 방향성을 들어봤다.

자기지도학습을 이해하는 새로운 시선을 선보이다

풍스 교수는 ‘Dual Perspectives on Non-Contrastive  
Self-Supervised Learning’이라는 강연을 통해, 딥러닝의  
작동 원리를 새롭게 해석할 수 있는 틀을 제시했다. 그는  
딥러닝이 실제로 놀라울 정도로 좋은 성능을 보이지만, 그  
성능의 근본 원인을 이론적으로 설명하는 데에는 여전히  
한계가 있다는 문제의식에서 출발했다.

특히 그는 자기지도학습self-supervised learning 분야에서 사  
용되는 대표적인 기법들, 예를 들어 정지 그래디언트stop-  
gradient나 지수이동평균EMA, Exponential Moving Average 기법이  
어떻게 모델의 표현 붕괴representation collapse를 방지하고 학  
습 성능을 향상시키는지를 주목했다. 지금까지 연구자  
들이 주로 이러한 현상을 최적화 이론optimization perspective  
의 관점에서 설명해왔다면, 풍스 교수는 동역학 시스템  
dynamical systems이라는 새로운 시각에서 분석을 시도했다.  
이것이 바로 그가 말하는 ‘이중 관점dual perspectives’이다.

실험 결과도 주목할 만하다. 그는 선형 모델linear setting에  
서는 이들 기법이 안정적으로 수렴converge하는 반면, 비선

형 모델non-linear setting에서는 그렇지 않다는 사실을 확인했  
다. 이 발견은 딥러닝이 어떤 상황에서 잘 작동하고, 또 언  
제 잘 작동하지 않는지를 이해하는 데 핵심적인 실마리를  
제공한다.

“컴퓨터 비전은 기본 물리와 기하학에서 출발”

“기초를 모르면 진짜 실력을 쌓을 수 없다”고 말하는  
풍스 교수는 최신 인공지능 기술인 합성곱 신경망CNN이  
나 트랜스포머를 배우고 싶어하는 학생들에게 먼저 ‘이  
미지는 어떻게 만들어질까? 카메라는 어떻게 작동할까?’  
라고 묻는다고 한다. 그가 저술한 대표 교재 『Computer  
Vision: A Modern Approach』는 지난 20년간 컴퓨터 비  
전을 배우는 학생들에게 기본기의 중요성을 강조해온 교  
과서다. 그는 이미지가 형성되는 물리적 원리, 카메라의  
수학적 구조, 인간 시각의 기하학적 해석 등을 먼저 이해  
해야 한다고 말했다. 기술이 아무리 발전해도, 기초 이론  
이 먼저 갖춰져야 한다는 철학이다.

최근 AI 기술이 쉽게 쓸 수 있는 도구가 되면서, 누구나  
딥러닝 모델을 사용할 수 있게 된 점에 대해 그는 긍정적으  
로 평가하다. 하지만 한편으로 그는 “이론적 깊이를 갖춘  
연구자가 점점 줄어드는 것 같다”며 걱정도 함께 내비쳤다.

AI 분야는 매일같이 새로운 논문과 기법이 쏟아지는 시  
대지만, 풍스 교수는 이런 흐름 속에서도 “시간이 모든 것  
을 증명해준다”고 강조하며 흔들리지 않는다. 10년 후에  
도 인용되는 논문, 여전히 널리 쓰이는 개념이 진짜 원천  
기술이라는 뜻이다. 예를 들어 그는 CNN이 보여준 ‘평  
행이동 불변성shift invariance’이 얼마나 강력한 ‘귀납적 편향  
inductive bias’이었는지, 트랜스포머 역시 사실은 과거의 ‘비  
국소 평균 필터non-local means filter’와 유사한 아이디어를 재  
발견한 것이라고 강조하며 “좋은 개념은 결국 다시 돌아  
온다”고 말했다. 그는 “최근 논문만 쫓지 말고, 고전 논문  
도 읽으라”며 “이론의 맥락 속에서 깊이 생각하는 법을 잊  
지 말라”는 조언도 잊지 않았다.

## 연구 여정은 ‘한순간’이 아니라 ‘꾸준한 대화’를 통해 열려

폰스 교수는 연구 여정에서 특별한 한순간이 인생을 바꾸었다는 식의 이야기를 하지 않았다. 대신 그는 “연구를 해오면서 끊임없이 호기심을 품고, 동료들과 대화하는 과정에서 연구의 방향이 자연스럽게 열렸다”고 말했다. 즉, 반짝이는 한순간의 아이디어보다 오래도록 꾸준히 생각하고 대화한 결과가 지금의 자신을 만들었다는 뜻이다.

그는 특히 네덜란드 위트레흐트대의 안 쿤더링크 Jan Koenderink 명예교수 같은 학자의 영향을 많이 받았다고 한다. 쿤더링크 교수는 인간의 시각 지각과 기하학을 연결해서, 복잡한 시각 현상을 단순한 수학 원리로 설명하려고 했던 세계적 연구자다. 이런 접근은 폰스 교수에게 큰 철학적 영감을 주었고, 이후 폰스 교수의 연구에서도 물리 기반의 시각 이해가 중요한 테마로 자리 잡게 됐다.

프랑스로 돌아온 후, 폰스 교수는 같은 학교 ENS-PSL의 프란시스 바흐 Francis Bach 교수와의 협업을 통해 머신러닝이라는 새로운 분야에 더 깊이 발을 들였다. 바흐 교수는 통계적 학습 이론과 최적화에 밝은 연구자인데, 이 협업 덕분에 폰스 교수는 기하학 중심의 시각 연구와 머신러닝의 수학적 기법을 융합할 수 있었다.

## AI가 세상을 이해하려면 언어 외에 시각과 세계 모델 필요

폰스 교수는 자율 AI 로봇이 현실 세계에서 안정적으로 작동하려면, 단순히 언어에 의존하는 것이 아니라 ‘세계를 이해하는 모델 World Model’이 필요하다고 강조했다. 그는 현재의 비전-언어 모델 VLM, Vision-Language Model이 이미지보다 텍스트에 과도하게 의존하는 경향이 있다고 비판했다. 그는 “이미지를 제거해도 모델 성능이 크게 바뀌지 않는 경우가 있다”며 “이는 시각 정보가 제대로 활용되지 않고 있다는 신호”라고 설명했다.

그는 로봇이 신뢰성 있게 움직이려면 단순히 눈앞의 이미지를 해석하는 것 이상이 필요하다고 밝혔다. 폰스 교수는 “로봇은 보이지 않는 공간도 예측하고, 목표에 도달하기 위한 최적의 경로를 계획해야 한다”며 “이는 탐색과 추론, 즉 글로벌 탐색 global exploration 능력이 필요하다는 뜻”이라고 말했다. 그는 현재 미국 뉴욕대 NYU와 메타 Meta의 얀 르쿤 Yann LeCun이 개발 중인 JEPA Joint Embedding Predictive Architecture는 이러한 세계 모델 문제를 해결하려는 대표적 시도 중 하나로 언급했다.

또 폰스 교수는 시종일관 “세상은 본질적으로 3차원 3D”이라며 “AI가 세상을 잘 이해하려면 3D 표현을 고려해야



‘글로벌 AI 프론티어 심포지엄 2025’에서 발표를 마친 뒤 인터뷰에 응하고 있는 장 폰스 교수.

한다”고 강조했다. 그는 굳이 3D 모델링 데이터를 쓰지 않더라도, 여러 각도에서 촬영한 이미지들만으로도 암묵적으로 implicitly 3D 구조를 추론할 수 있다고 설명했다.

특히 단단한 물체 rigid shape는 비교적 다루기 쉽지만, 옷이나 단백질처럼 유연한 형태 non-rigid shape는 훨씬 다루기 어려워 여전히 도전 과제다. 그럼에도 그는 “컴퓨터 비전 기술이 복잡한 비정형 대상을 다룰 수 있다면 과학적 탐색에도 크게 기여할 수 있다”고 밝혔다. 예를 들어 천문학에서 외계 행성을 탐지할 때 여러 차례의 관측 데이터를 종합해 행성의 존재나 궤도를 추론하거나 생물학(분자과학)에서 단백질 접힘, DNA 구조 변화 등을 최신 비전 모델로 모델링하면 신약 개발이나 질병 메커니즘 분석에 중요한 단서를 제공할 수 있다.

## “연결하고 질문하라”, 그가 말하는 AI 연구의 본질

폰스 교수는 기하학 기반의 컴퓨터 비전 geometry-based computer vision이 앞으로 더욱 중요해질 분야로 디지털 헬스케어, 정밀 제조, 로봇 공학 등을 꼽았다. 이 분야들에서는 정확한 포즈 인식 pose estimation, 정밀한 3D 재구성 3D reconstruction, 공간적 위치 추정 spatial localization 등이 필수적

이다. 단 1mm의 오차가 생명을 좌우하거나 생산 라인을 멈추게 할 수 있는 산업에서는 수학적 정합성과 물리 기반의 비전 기술이 매우 큰 차이를 만든다는 것이 그의 설명이다.

그는 특히 “한국 연구자들과의 협업 경험이 매우 긍정적이었다”고 회고하며, 프랑스의 AI 연구 거점인 PRAIRIE-PSAI 및 ENS-PSL과 한국의 국가AI연구거점 NAIRL의 국제 공동연구 가능성에도 기대를 표했다. 이는 단순한 기술 교류를 넘어, 기초 이론과 실용 응용을 아우르는 협력 플랫폼이 될 수 있다는 점에서 주목할 만하다.

인터뷰 말미에 폰스 교수는 후속 세대에게 한 가지 조언을 남겼다. “좋은 아이디어는 혼자 고민한다고 나오지 않습니다. 아이디어를 주고받고, 반응을 듣고, 때로는 반박을 당하는 과정에서 진짜 통찰이 나옵니다.”

요즘 시가 모든 분야에 퍼지고 있다. 그는 “미래는 불확실하지만 기회는 많다”며 “지금이야말로 기초를 갖춘 연구자에게 최고의 기회가 열리는 시기”라고 강조했다. 급변하는 논문 경향을 쫓는 것도 중요하지만, 그 속에서 이론의 뿌리와 역사적 맥락을 함께 보는 안목이야말로 앞으로 더 빛날 자산이 될 수 있다는 말이다.

세계적인 석학인 폰스 교수는 단지 최신 모델을 개발하거나 성능 지표를 높이는 데에만 관심을 두지 않는다. 그는 늘 ‘왜 이 기술이 작동하는가?’라고 자문한다고 한다. 기계학습 모델의 정체와 구조를 물리, 기하학, 심지어 인간 시지각 이론과 연결시키려는 그의 연구는 지금의 딥러닝 흐름 속에서 드물게 만나는 깊은 사유의 결과다.

“AI가 자동화되고 추상화될수록, 오히려 기본 물리와 기하학, 시지각의 본질에서 출발하는 연구가 더 중요한 시대가 될 것입니다.”

지금, 우리에게 필요한 연구란 그런 것이 아닐까.

NAIRL

## 산업 AI 혁신을 선도하는 AI 스타트업 플레이오니

### AI로 제조에서 로봇틱스까지 도전한다

지난 10월 20일 국가AI연구거점의 파트너 기업 중 하나인 플레이오니<sup>Pleiony</sup>를 찾아갔다. 서울 서초대로에 자리한 플레이오니에서 우제학 대표를 만나 창업 6년 차 인공지능<sup>AI</sup> 스타트업을 운영하는 어려움과 함께 미래 비전을 전해 들었다.

#### ‘관심’에서 ‘업종’으로… 딥러닝과의 첫 만남

“AI는 그저 유행이 아닙니다. 패러다임을 바꾸는 근본 기술이라고 생각합니다.”

딥러닝, 강화학습, AI 로봇틱스까지 다루는 인공지능 스타트업 플레이오니는 산업용 AI 솔루션을 개발할 뿐만 아니라 피지컬 AI, 임베디드 AI를 포함하는 차세대 로봇 지능에 도전하려고 준비하고 있다. 창업자이자 대표인 우제학 대표는 강화학습 기반 제어 시스템과 AI 로봇의 현실적 구현을 위해 동분서주하고 있다. 플레이오니는 딥러닝 및 대형언어모델<sup>LLM</sup> 기반의 인공지능 기술 및 응용 솔루션을 연구개발하는 R&D 중심 기업이다.

플레이오니는 국가AI연구거점 참여 기업으로서도 활동하고 있다. 우 대표는 KAIST 김재철AI대학원 CAIO<sup>Chief AI Officer</sup> 과정 5기로 세미나 방식의 강의를 들으면서 강화학습에 대한 관심을 갖게 됐고, 현 국가AI연구거점 센터장인 김기응 교수도 만났다. 플레이오니는 김 교수가 국가AI연구거점에 도전할 때 참여 기업으로 요청하자 이에 적극적으로 응했다고 한다.

우 대표가 인공지능에 본격적인 관심을 갖게 된 것은 2014년. 국내 정보과학회 컨퍼런스에서 한 발표자가 소개한 신경망 컨퍼런스(현 NeurIPS)가 계기가 됐다. “딥러닝 같은 신경망이 세상을 바꿀 수도 있겠다는 생각이 들었어요. 2017년에 실제 오픈 모델을 써보면서 가능성을 체감했고, 1년간 미국 대학에 방문 학자로 머물며 컨퍼런스에 많이 참여하면서 AI 연구를 직접 들여다봤죠.” 그동안 IT 회사를 운영하던 그는 2019년 AI 스타트업 플레이오니를 설립하며 AI 시장에 뛰어들었다.

#### 산업 AI 혁신을 선도하는 AI 기술 전문 기업

회사명 ‘플레이오니’는 그리스 신화의 요정 ‘플레이오네<sup>Pleione</sup>’란 이름에서 따왔다. 플레이오네와 아틀라스 사이에서 태어난 일곱 자매는 밤하늘의 플레이아데스성단으로 유명하다. “플레이오니의 딸이 마이아인데, MAIA를

‘Multimodal AI Application’의 약자로 해석해 회사의 방향성까지도 맞췄어요.” 덕분에 회사는 이름도, 철학도 남다르다.

플레이오니는 산업 AI 혁신을 선도하는 AI 기술 전문 기업을 표방하고 있다. 초기 기술은 이미지 인식 기반의 딥러닝<sup>CNN</sup>으로, 제조 현장의 머신 비전 문제를 해결하는 데 초점을 맞췄다. 회사는 초창기에 AI 기술을 갖고 산업 또는 제조 분야에 쓸 수 있는 솔루션을 개발하며 AI 제조 쪽을 하다가 AI 모니터링과 관련된 전반적인 문제를 해결하는 데 나섰으며, 이후 강화학습 응용 분야인 AI 에너지 최적화 문제를 다루었고, 요즘 AI 로봇틱스 쪽으로 분야를 확장하고 있다.



KAIST 김재철AI대학원 CAIO 과정에서 강의를 들으며 강화학습에 관심을 갖게 됐다는 우제학 대표.

특히 플레이오니는 다양한 산업 현장에 AI 기반 모니터링 및 예측 모델도 적용해 왔다. 우 대표는 “연구 위주로 실제 현장에 적용한 사례가 많지 않다”며 “개념 증명<sup>PoC, Proof of Concept</sup> 정도로만 개발한 경우가 조금 있다”고 밝혔다. 개념 증명이란 아이디어, 기술, 시스템이 이론이 아니라 현실에서도 작동하는지 사전에 검증하는 과정을 의미



우제학 대표는 플레이오니 직원들과 함께 매주 AI와 관련된 세미나와 논문 스터디를 꾸준히 하고 있다.

한다. 예를 들어 AI 제조 분야에서 결함 검사의 일환으로 생산된 알루미늄 폴대의 표면에서 결함을 감지한 사례, 정 부 과제로는 CNC 기계에서 진동 데이터나 모터 전류 값을 센싱해 실시간으로 충동을 감지하고 비상 정지시키는 솔루션 개발 사례, 냉장고 부품의 용접 단계에서 수집한 전류 데이터로 공정 불량을 찾아내는 프로젝트 등이 대표 적이다.

#### 강화학습 기반 제어에서 스마트 팩토리까지

최근에는 LLM을 포함한 딥러닝과 함께, 강화학습 기반 제어 시스템은 물론이고 AI 로봇틱스, 피지컬 AI 분야에도 집중하고 있다. 예를 들어 강화학습 기반의 OIS(Optical Image Stabilization) 시스템 개발이 대표적이다. OIS 칩은 카메라에서 손 떨림 보정을 구현하는 칩으로 보통 전통적인 로봇 제어 기술인 PID(Proportional Integral Derivative) 기술이 들어간다. “기존 PID 제어 방식의 소프트웨어를 강화학습으로 대체

해달라는 요청을 받았어요. 우리가 만든 모델은 실제 칩에 들어갈 정도의 성능을 보였고, 향후 카메라 모듈 자동 세팅에까지 적용될 수 있을 거라 봅니다.”

또 회사는 스마트 팩토리 분야에서도 AI 기술을 활용하고 있다. 자주 활용하는 기술 중 하나가 원-클래스 이상 탐지(One-Class Anomaly Detection)다. 우 대표는 “이상 상황 데이터는 일반적으로 확보하기 어렵기 때문에, 정상 상태 데이터만 학습해 이상 특징 벡터를 분리하는 방식으로 구현한다”며 “딥러닝의 벡터화 표현 모델을 기반으로 하고 특징 벡터 공간에서의 분리를 활용해 이상 여부를 판별하는 구조”라고 설명했다

현재는 현대제철의 ‘공정 데이터 활용 배관 막힘 진단 서비스 개발’ 과제를 진행하고 있다. 즉 부생가스(COG)라는 고부가가치 연료가스를 공급하는 배관에서 가끔 막히는 현상이 발생하는데, 공정 데이터를 바탕으로 배관 막힘을 사전에 탐지하는 AI 모델을 개발하는 중이다. 우 대표는

“타르가 점차 쌓이며 배관이 막히는 걸 감지하는 AI 모델을 만들고 있다”며 “완성되면 실제로 현장에서 설치해 쓸 수 있을 것”이라고 설명했다.

#### AI 로봇 제어 기술에 박차

앞으로 플레이오니는 AI 로봇 제어 기술에 박차를 가한다. 우 대표는 “아직 실적은 많지 않지만, 강화학습 개념으로 접근하되 비전을 적용해 로봇 매니폴레이션을 잘할 수 있는 기술을 개발하고자 한다”며 “이는 피지컬 AI, 임베디드 AI 분야와 관련된 것”이라고 설명했다. 예를 들어 로봇 팔에 카메라를 달아 PC나 자동차에서 부품 삽입 및 조립, 병뚜껑 닫기 같은 정밀 작업을 수행하는 모델을 개발하고 있다.

또 로봇 파운데이션 모델 중 대표적 기술인 VLA(Vision Language Action)와 강화학습 기반 제어 기술을 결합하려는 연구도 이어가고 있다. VLA는 로봇의 주변 환경 인지, 자연어 이해 능력, 그리고 구체화 액션을 하나의 컴퓨팅 체계로 통합하기 위한 일종의 로봇 파운데이션 모델이다. 플레이오니는 이 기술을 산업 로봇에 적용하기 위한 다양한 실험적 시도를 하고 있다.

아울러 인간 참여형 강화학습인 HART(Human-Augmented Robot Training)를 활용하기 위해 노력하고 있다. HART는 인간의 직관적인 시연과 실시간 교정 데이터를 강화학습 알고리즘에 결합하는 기술인데, 이 혁신적인 접근은 기존 강화학습의 비효율성을 극복하고 단 몇 시간 만에 인간 전문가 수준을 뛰어넘는 로봇 제어 정책을 학습할 수 있게 한다.

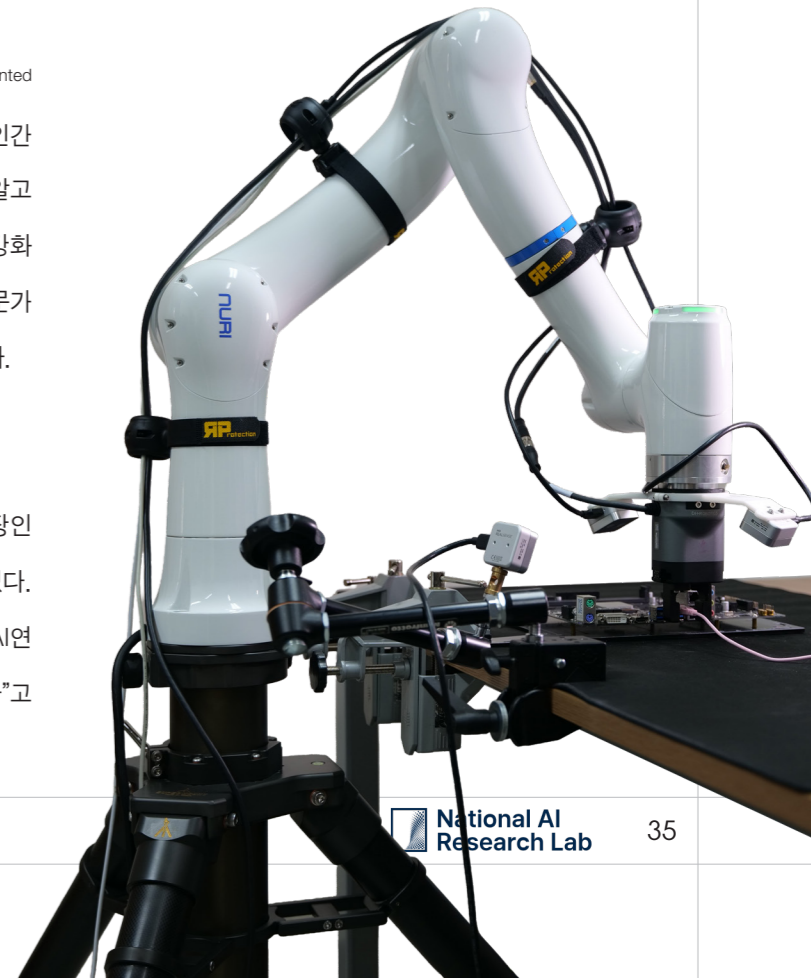
#### “소규모 조직임에도 AI 트렌드 따라가며 노력해”

플레이오니는 국가AI연구거점 참여사로서 센터장인 KAIST 김기웅 교수 등과 세미나·상담을 이어오고 있다. 우 대표는 “향후 로봇 파운데이션 모델 분야에서 국가AI연구거점과 공동으로 할 만한 프로젝트도 탐색하고 있다”고 밝혔다.

우 대표는 스타트업 경영자로서 국가 단위의 AI 거버넌스 통합의 필요성도 강조했다. “로봇은 산자부가 맡고, AI는 과기정통부, 중기부가 맡고 있는데, 각 부처마다 과제 기준이 달라 비효율이 큼니다. 통합된 지휘 체계가 필요하다고 봅니다.”

직원 10명 내외의 소규모 조직임에도, 플레이오니는 매주 세미나와 논문 스터디를 꾸준히 이어간다. “대표이자 CTO로서 아직도 배우는 자세로 임하고 있습니다. AI는 세계적인 석학들이 이끄는 분야니까, 따라가기만 해도 고된 여정이죠. 하지만 스타트업 분위기 속에서 트렌드를 안 놓치려고 노력하고 있습니다.” 우 대표의 말에서 AI 기술을 향한 욕심이 느껴진다. **NAIRL**

최근 플레이오니는 강화학습 기반으로 비전을 적용해 로봇 매니폴레이션을 잘할 수 있는 기술을 개발하고 있다.





## NAIRL Frontier Letter | 2025 | Vol.02

발간인 국가AI연구거점 김기웅 센터장  
발행일 2025년 12월  
발행처 서울특별시 서초구 태봉로 108 서울 AI허브 별관 2층  
기획 국가AI연구거점 연구전략팀 나현대 연구조교수  
제작 (주)동아에스앤씨

국가 AI 연구거점은 4개 대학(KAIST, 고려대, 연세대, POSTECH) 컨소시엄 외 미국, 캐나다, 프랑스, UAE의 총 14개 국제공동연구기관, 국내 12개 협력기업, 서울특별시 및 서초구청 지자체 협력으로 구성된 역동적인 산·학·연·관 협력 생태계입니다.

