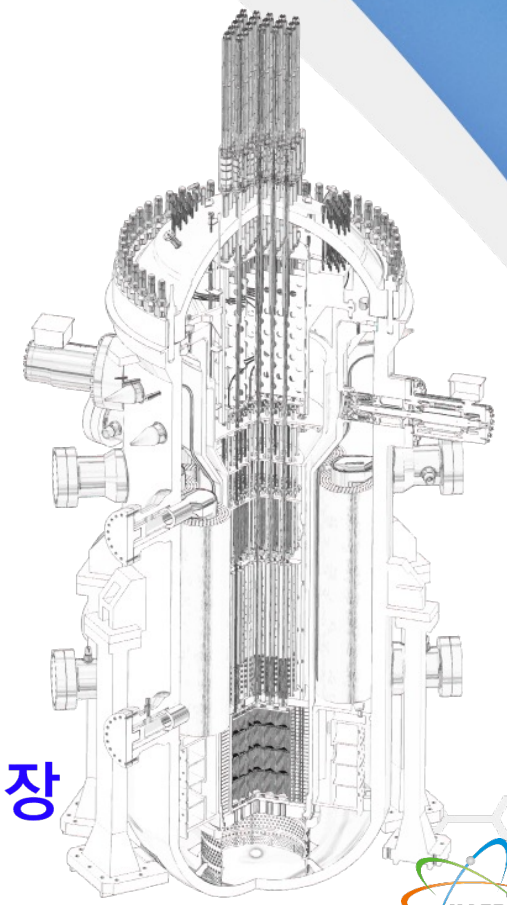
더 나은 세상을 위한 **원자력기술**

국민과 세계가 지지하는 한국원자력연구원

원자력 AI Transformation을 위한 산학연 협력 전략

10/26/23



한국원자력연구원 인공지능응용연구실장

유용균

한국원자력연구원 Korea Atomic Energy Research Institute

00 개인소개

- 한국원자력연구원 인공지능응용연구실장
- 사단법인/커뮤니티 AI프렌즈 대표
- UST 인공지능전공 전임교원
- Nvidia 앰배서더 / DLI Instructor
- 대한기계학회 인공지능머신연구회 이사

유용균 (Yonggyun Yu)
최적설계/바이올린/딥러닝에 관심있는 애아빠.
(주의) 딥러닝 전문가 아님.덕후. https://goo.gl/YehHTD
AI 프렌즈 운영진/기계학회 인공지능머신연구회 이사
수정

"연구는 덕질이다"…'딴짓'만 골라서 하는 별난 과학자

박성민 기자 | sungmin8497@hellodd.com | 입력 2018.09.03 15:44 | 수정 2019.10.28 17:53 | 댓글 7

[과학청년 부탁해 ㉞]유용균 원자력연 박사…AI 기반 최적설계 연구
보다 많은 사람이 AI 활용하는 플랫폼 개발 목표…"따뜻한 기술로 사회공헌"

유용균 박사는 젊은 과학을 '덕질'이라고 표현했다. 자신이 좋아하는 분야에 심취하다 보면 그 어떤 것보다 아까울 것이 없다는 이야기다.<사진=박성민 기자>

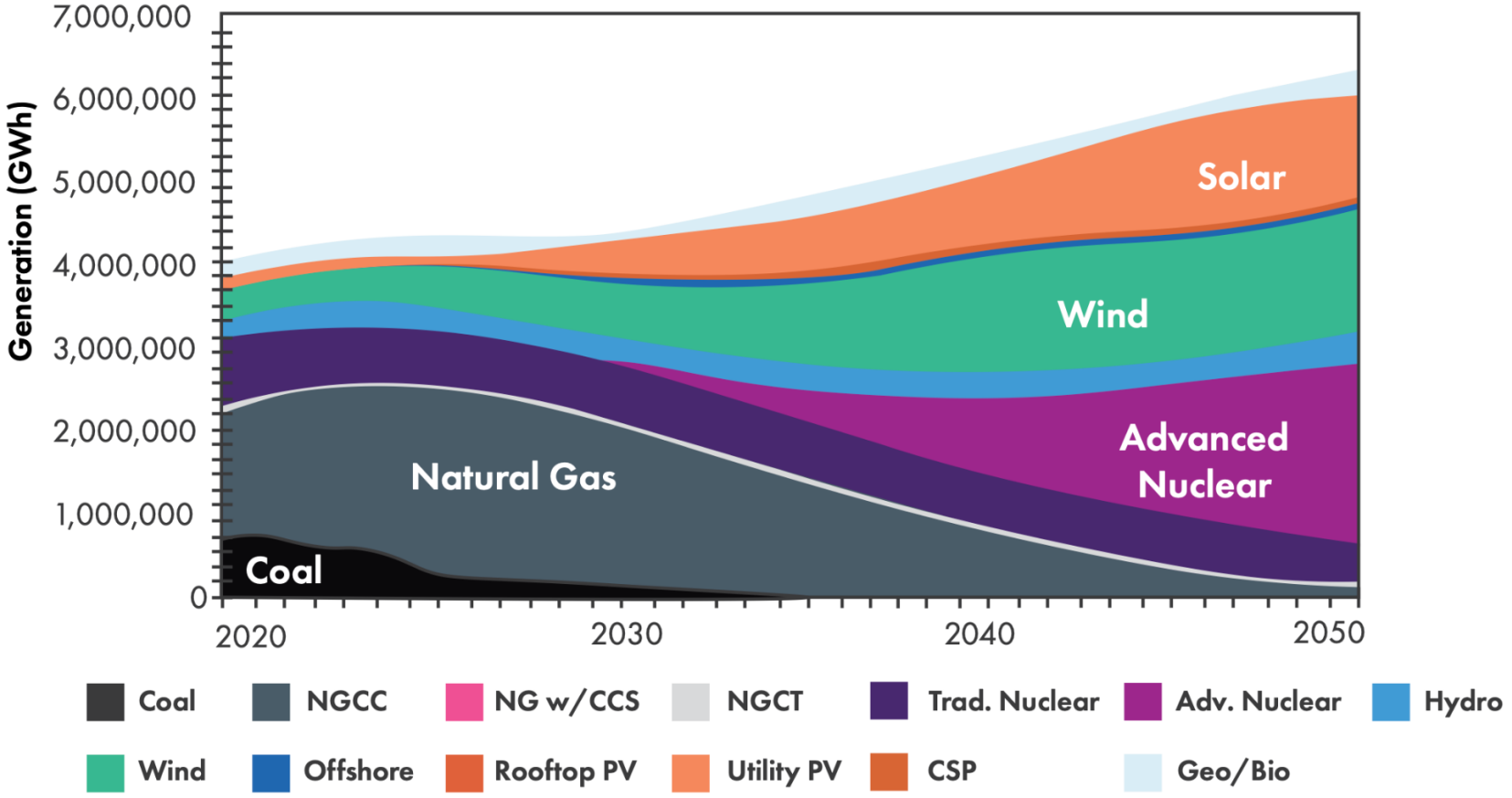
01

원자력분야의 인공지능 필요성

한국원자력연구원
Korea Atomic Energy Research Institute

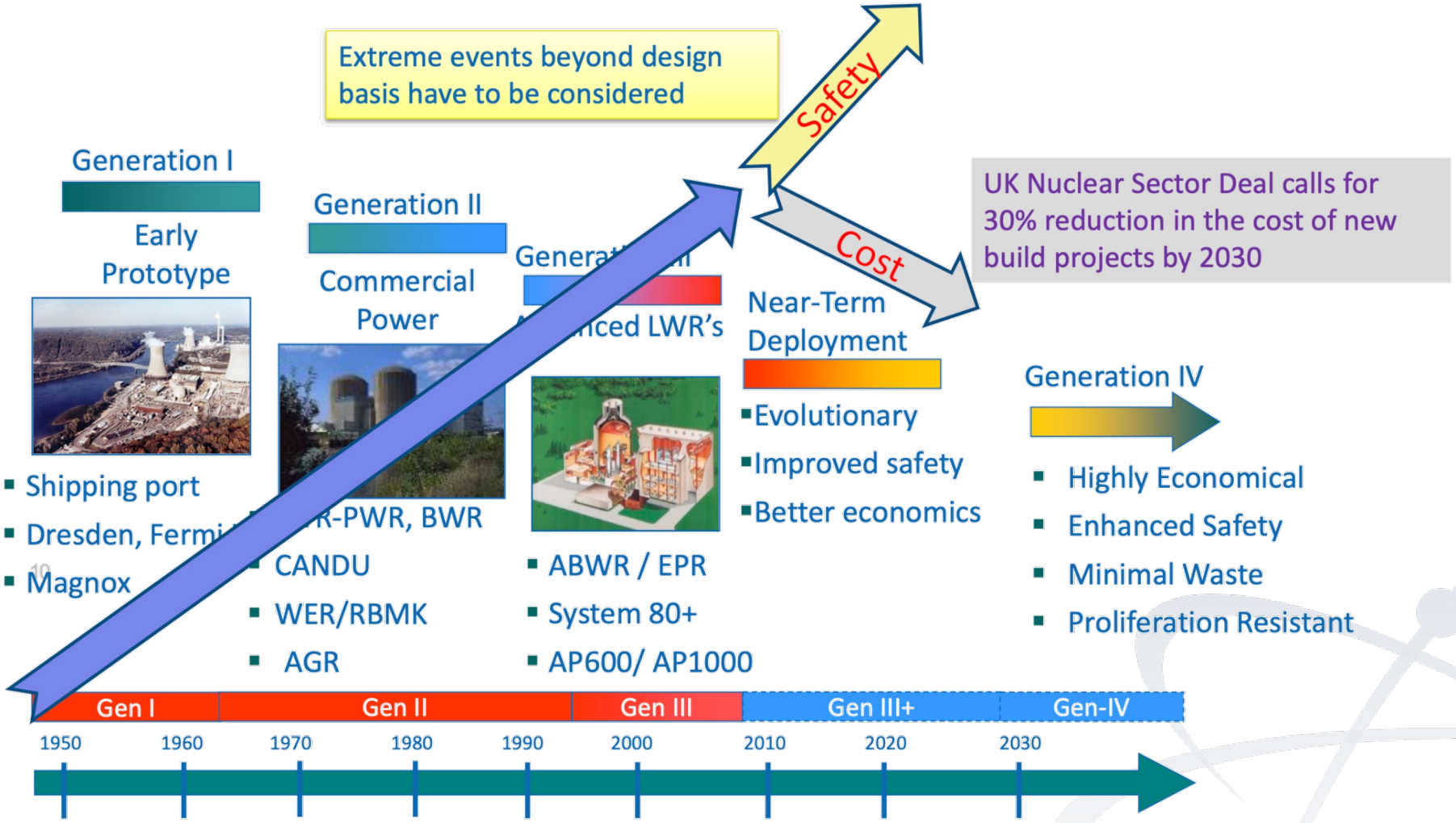
탄소중립에 따른 에너지 계통 복잡도 상승

Figure 2. Potential future electricity generation landscape through 2050[1]

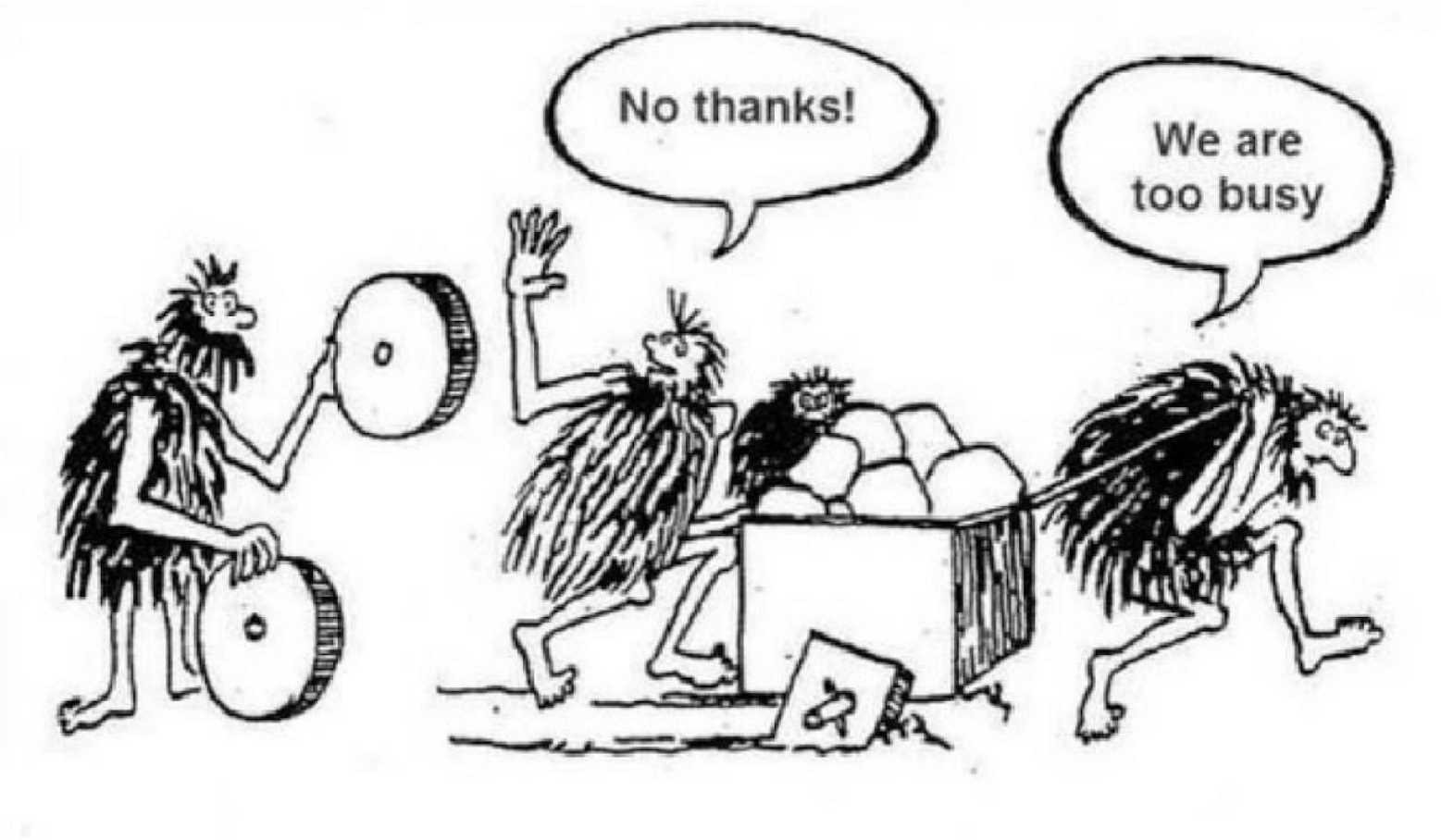


Vibrant Clean Energy LLC. "Role of Advanced Nuclear Technologies in Decarbonizing the U.S. Energy System." NEI Board of Directors presentation, Washington, D.C., May 2022.

차세대 원자력의 AI Transformation의 필요성



from Gen4 International forum, Artificial Intelligence in Support of the Nuclear Energy Sector


No thanks!
We are too busy

01 해외 원자력 디지털전환 관련 동향

- ARPA-E는 선진원자로 개발 및 원전의 유지보수(O&M) 시스템의 혁신적인 전환*을 위해 디지털 트윈 기술 등 혁신기술 개발을 목표로 하는 GEMINA 프로그램을 '19년부터 운영 중
- GEMINA는 인공지능(AI), 선진제어시스템, 예측 유지 보수 및 모델 기반 결함 감지와 같은 선진 센서 및 계측 기술을 적용한 O&M 솔루션 개발에 집중
- 미국 원전의 O&M 비용은 원자로 총 발전 비용의 약 80 %로 효과적이고 경제적인 원자로 O&M 기술의 필요성이 대두되었으며, ARPA-E는 이 프로그램을 통해 개발된 기술을 활용하여 원전의 고정 O&M 비용을 현재 13 $/MWh에서 2 $/MWh로 수준으로 줄이는 것을 목표로 하고 있음.

GEMINA 프로그램

NO	사업 내용	수행기관 (사업 기간)	지원금액 ($)
1	**•선진원자로의 예측 유지관리를 위한 AI 지원 디지털 트윈 개발** - GE Hitachi의 BWRX-300 SMR 설계를 기반으로, 원자로 시스템에 대한 지속적인 모니터링, 조기 경고, 진단 및 예후가 가능하도록 디지털 트윈을 개발, 알고리즘이 인식하지 못하는 상황이 발생할 경우 안전 작동 모드로 기본 설정되는 "Humble AI" 프레임워크를 개발하여 불확실성, 데이터 및 모델 보증, 디지털 트윈의 지속적인 학습을 보장	GE Research ('20.9.10.~ '23.9.9, 3년)	5,412,810
2	**•Build-to-Replace: 선진 원자로 O&M 비용 절감을 위한 새로운 패러다임 연구** - 원전 관리 체계를 기존 '유지 및 수리(maintain and repair)'에서 부품의 수명 예측을 통한'교체(replace and refurbish)' 접근 방식으로 전환하기 위한 연구를 수행하고, 또한 경제적 이점 측면에서 몇몇 선진 원자로의 설계를 평가	EPRI ('20.10.1.~ '23.9.30, 3년)	999,464
3	**•고정 O&M 비용 절감을 위한 Xe-100 디지털 트윈의 운영 및 유지관리 기술 개발** - X-100의 고정 O&M 비용을 $2/MWh로 줄이는 것을 목표로 개발 중인 디지털 트윈 프로젝트와 관련하여, 인적 요인 엔지니어링, 확률 위험 평가, 위험 분석, 보안 및 유지관리 평가 등을 통한 시스템 최적화 연구, 자동화, 로보틱스, 원격 및 중앙 집중식 유지보수, 모니터링 등 첨단 기술을 활용하여 원전 운영을 최적화 하는 혁신적인 방법을 개발 * 이를 위해 X-Energy는 종합 3D 모델과 디지털 트윈 모델링 등 두가지 모델링 체계를 개발 중	X-energy ('21.1.5.~ '23.1.4, 2년)	5,978,289
4	**•선진원자로 센서 및 구성품 유지관리 (MARS)** - ANL은 Kairos의 불화염 냉각 고온로의 O&M 비용 절감을 목표로, 구성품과 시스템 특성 분석을 위한 선진 분산 센서, 다기능 센서 및 데이터 수집 기술을 개발하고, 센서 데이터의 자동 분석을 위한 머신러닝 기반의 신호처리 알고리즘을 개발 (목표 O&M 비용: $2/MWh)	ANL ('20.9.28.~ '24.3.27, 3.5년)	2,199,009
5	**•디지털 트윈 기반 HTGR 냉각시스템의 성능 및 신뢰성 진단 기술 개발** - Framatom은 Metroscope* 소프트웨어와 함께 연계하여 HTGR의 냉각 시스템을 시뮬레이션하는 2개의 디지털 트윈을 개발 * 고장 라이이브러와 디지털 트윈을 연결시켜주고 문제가 되는 알고리즘을 감시하는 소프트웨어 패키지	Framatome ('20.10.1.~ '22.9.30, 2년)	809,701
6	**•BWRX-300 핵심 시스템용 고정확도 디지털 트윈스 개발** - MIT는 BWRX-300 SMR에 적용할 고정확도 디지털 트윈을 개발·검증하고, 예측 유지보수 접근법과 모델기반 결함 탐지 기술을 실증	MIT ('20.10.1.~ '23.9.30, 3년)	1,787,065
7	**•SSR-W(Stable Salt Reactor-Waste)에 적용할 다물리 디지털 트윈 개발** - SSR을 위한 다물리 디지털 트윈을 개발을 통해 SSR-W의 운용 및 유지보수 전략을 제시하고 가상의 테스트 환경을 제공	Moltex Energy ('21.4.1.~ '24.9.30, 3.5년)	4,618,868
8	**•용융염 원자로의 디지털 트윈을 위한 핵심 조사(Irradiation) 데이터 생산** - MIT는 연구용원자로를 이용하여 MSR 내에서 방사성 물질의 생성과 거동을 밝혀내기 위한 연구를 수행 (MSR의 디지털 트윈 개발을 위해서 이 데이터는 매우 중요)	MIT ('20.9.30.~ '22.9.29, 2년)	899,825
총합		**22,705,031**	

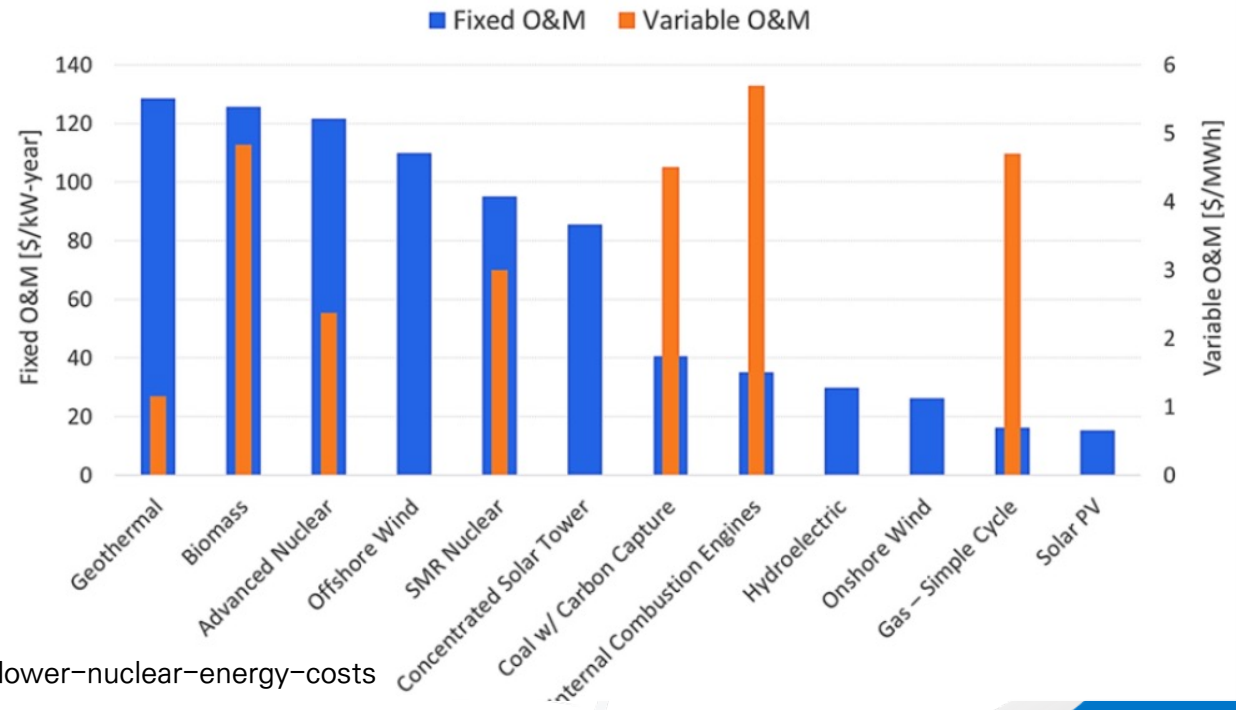
ROLE OF AUTONOMOUS OPERATION IN IMPROVING NUCLEAR UNITS PROFITABILITY

Impact of O&M costs on the economy of Nuclear Units

- Currently operated units are struggling to stay competitive in U.S. Deregulated markets. Significant impact of fixed O&M costs (largest portion goes to payroll for staffing).

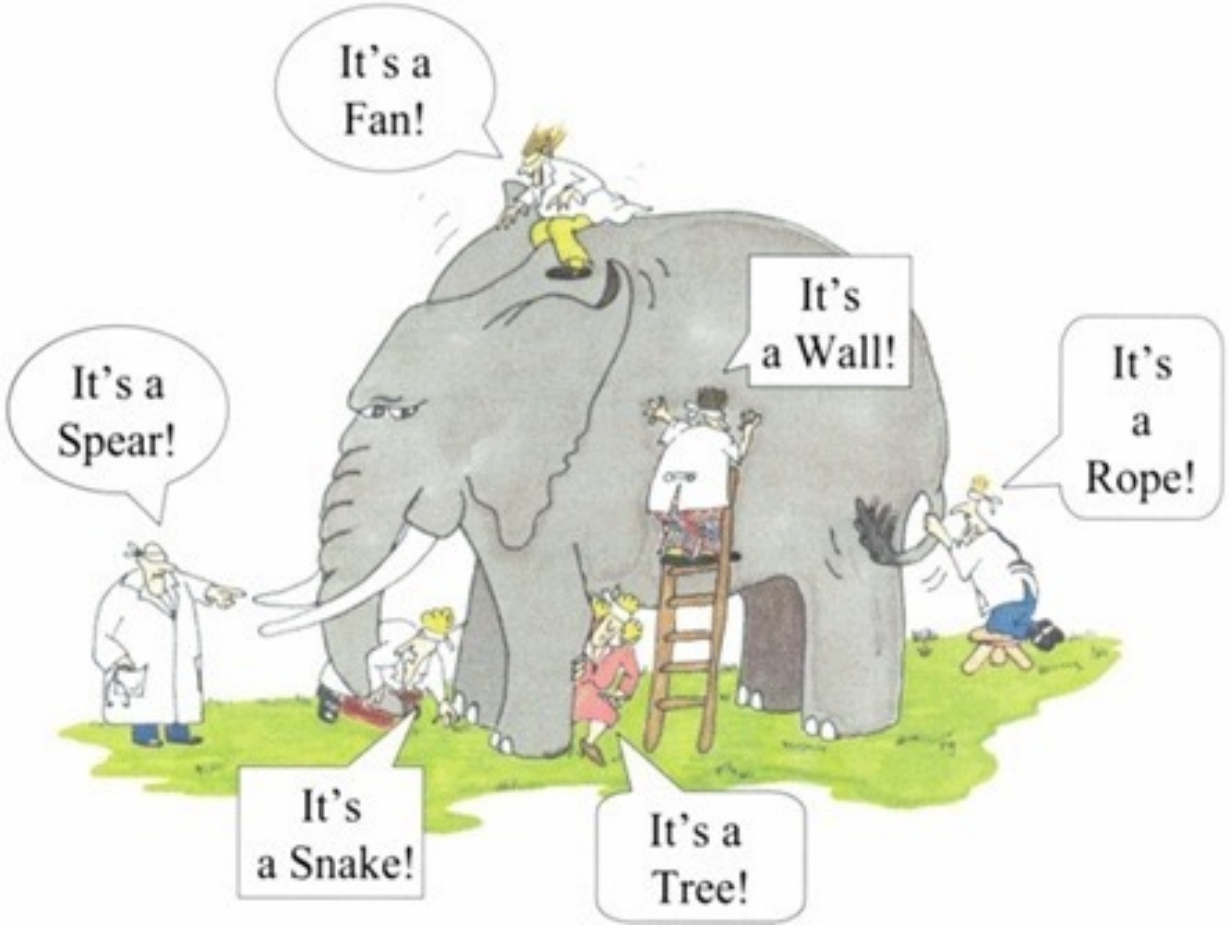
How can Autonomous Operation help saving on O&M costs?

- Limiting the number of MCR (Main Control Room) operators does not significantly reduce costs
- Most of the savings can be accomplished by optimizing the maintenance schedule
- Maintenance interventions can be less time-consuming, number of on-site technicians reduced



https://www.anl.gov/article/how-artificial-intelligence-could-lower-nuclear-energy-costs

인공지능의 정의는?

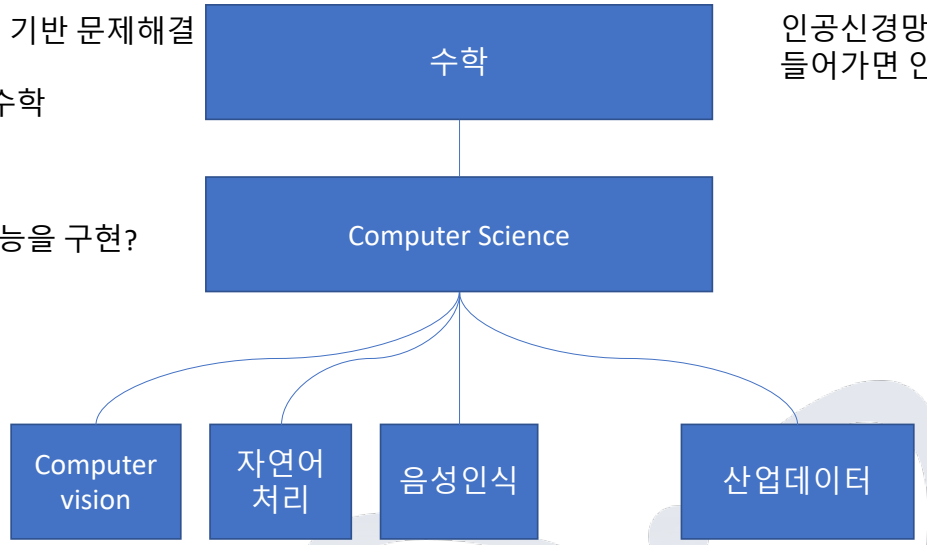


전체 시스템을 만드는 것?

데이터 기반 문제해결

통계, 수학

인공지능을 구현?



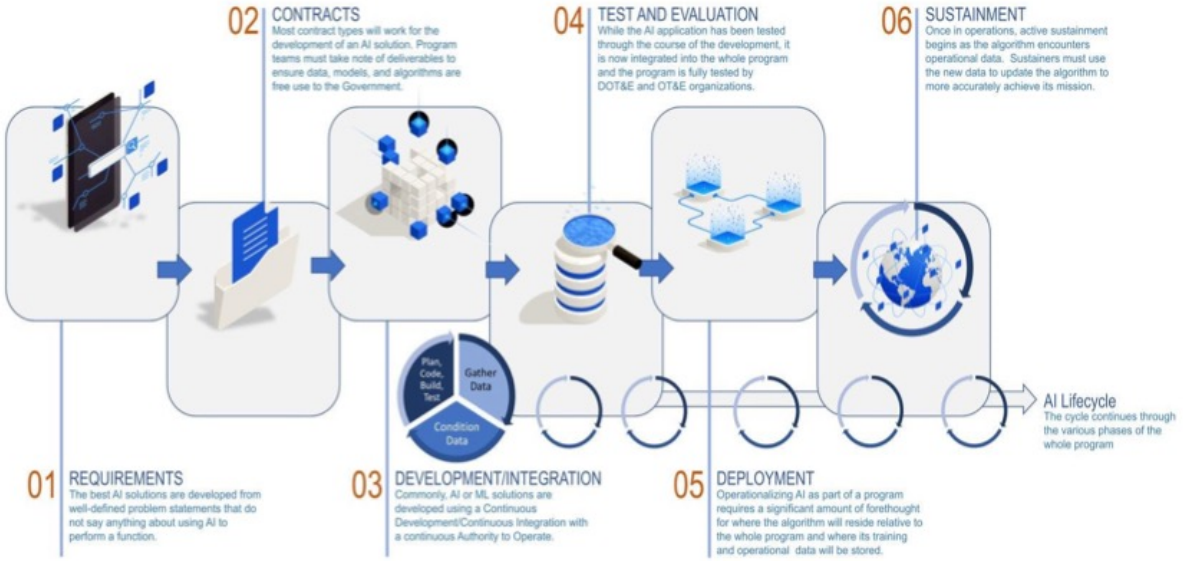
인공신경망이 들어가면 인공지능?

인간의 인식과 지능과 관련된 분야?

AI as the Program, AI as Part of the Program

Project Type 2. **AI is the Program**

Project Type 1. **AI as Part of the Program**



* MIT, Artificial Intelligence Accelerator

02

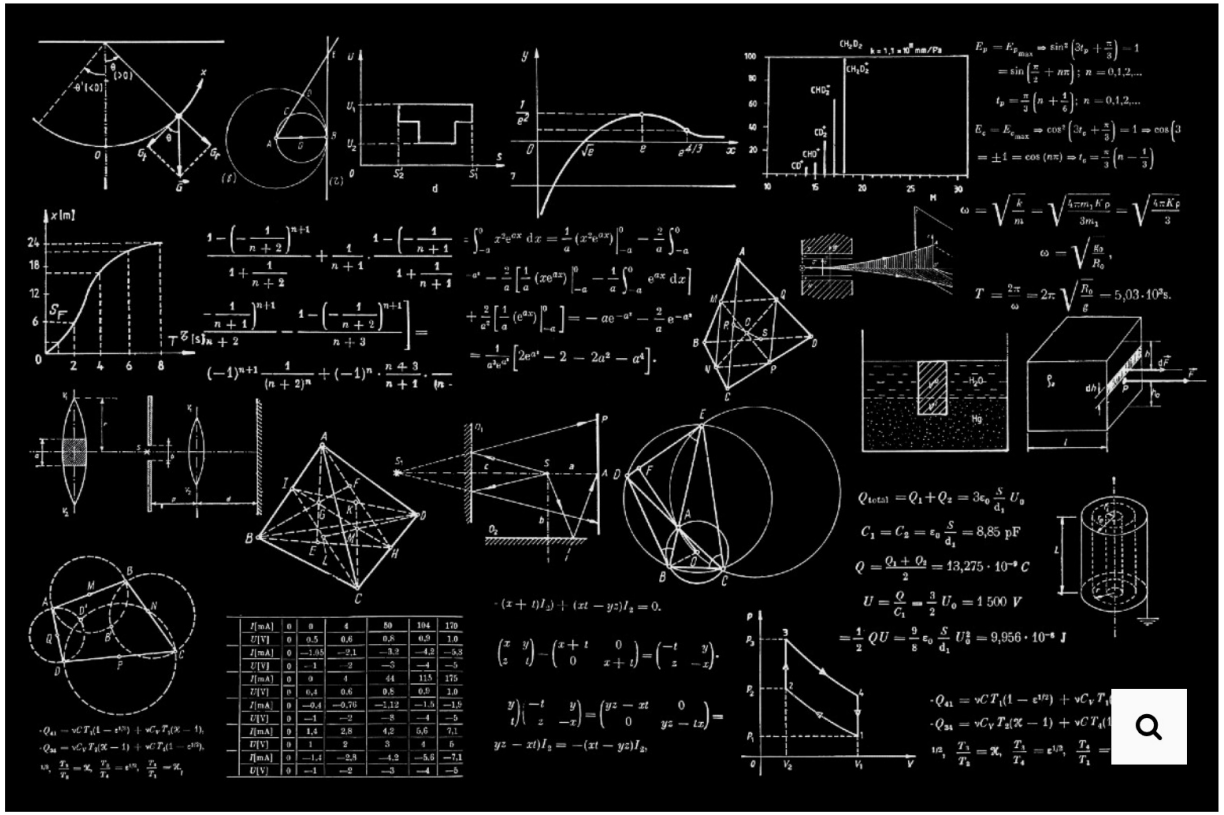
문제해결 방법론 측면의 AI

한국원자력연구원
Korea Atomic Energy Research Institute

하나의 관점: 인공지능은 새로운 수학이다

AI is the new math—and an enabler of humanity's potential

Math and AI become one to better the future of humanity.



Photograph: Dan Cristian Pădureț via Unsplash

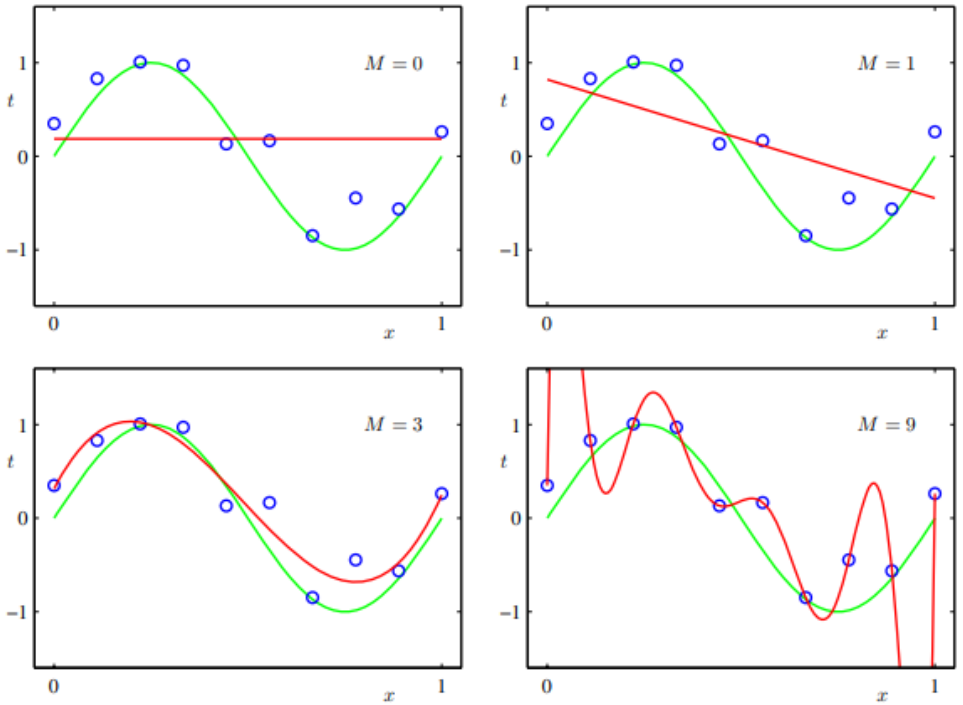


Figure 1.4 Plots of polynomials having various orders M, shown as red curves, fitted to the data set shown in Figure 1.2.

박사학위 주제 : 최적설계 함수 만들기

조건 $f(x)$

$$\underset{\rho}{\text{minimize}} \quad F = F(\mathbf{u}(\rho), \rho) = \int_\Omega f(\mathbf{u}(\rho), \rho) \mathrm{d}V$$

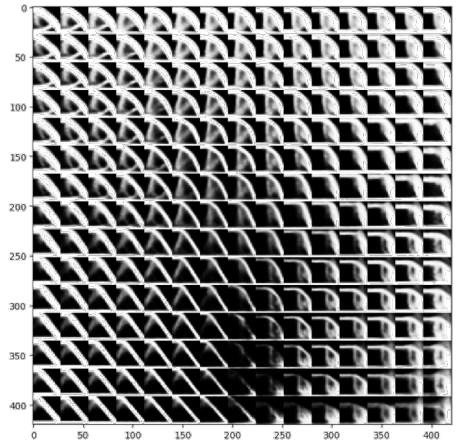
$$\text{subject to} \quad G_0(\rho) = \int_\Omega \rho \mathrm{d}V - V_0 \leq 0$$

$$G_j(\mathbf{u}(\rho), \rho) \leq 0 \text{ with } j = 1, \ldots, m$$

02 데이터 기반 최적설계

데이터로 설계하는 함수를 만들수 있지 않을까?

조건 $f(x)$



Structural and Multidisciplinary Optimization
https://doi.org/10.1007/s00158-018-2101-5

RESEARCH PAPER

CrossMark

Deep learning for determining a near-optimal topological design without any iteration

Yonggyun Yu[1] · Taeil Hur[2] · Jaeho Jung[1] · In Gwun Jang[3]

Received: 3 December 2017 / Revised: 17 August 2018 / Accepted: 17 September 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
In this study, we propose a novel deep learning-based method to predict an optimized structure for a given boundary condition and optimization setting without using any iterative scheme. For this purpose, first, using open-source topology optimization code, datasets of the optimized structures paired with the corresponding information on boundary conditions and optimization settings are generated at low (32 × 32) and high (128 × 128) resolutions. To construct the artificial neural network for the proposed method, a convolutional neural network (CNN)-based encoder and decoder network is trained using the training dataset generated at low resolution. Then, as a two-stage refinement, the conditional generative adversarial network (cGAN) is trained with the optimized structures paired at both low and high resolutions and is connected to the trained CNN-based encoder and decoder network. The performance evaluation results of the integrated network demonstrate that the proposed method can determine a near-optimal structure in terms of pixel values and compliance with negligible computational time.

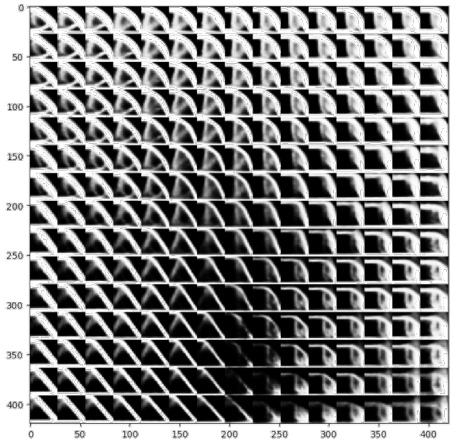
딥러닝과 최적설계의 융합

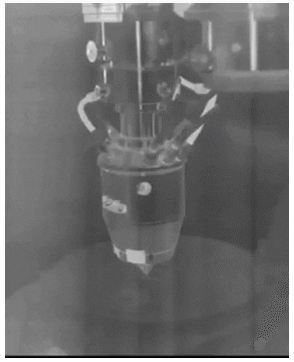
$$\underset{\rho}{\text{minimize}} \quad F = F(\mathbf{u}(\rho), \rho) = \int_{\Omega} f(\mathbf{u}(\rho), \rho) \mathrm{d}V$$

$$\text{subject to} \quad G_0(\rho) = \int_{\Omega} \rho \mathrm{d}V - V_0 \leq 0$$

$$G_j(\mathbf{u}(\rho), \rho) \leq 0 \text{ with } j = 1, \ldots, m$$

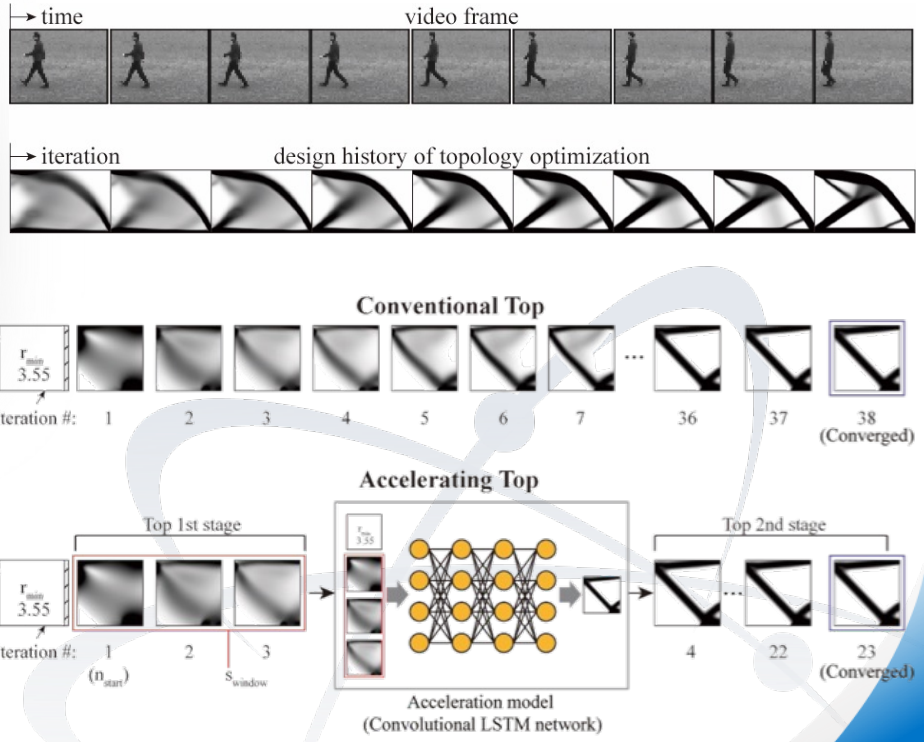
조건 $f(x)$





Unit module-based convergence acceleration for topology optimization using the spatiotemporal deep neural network

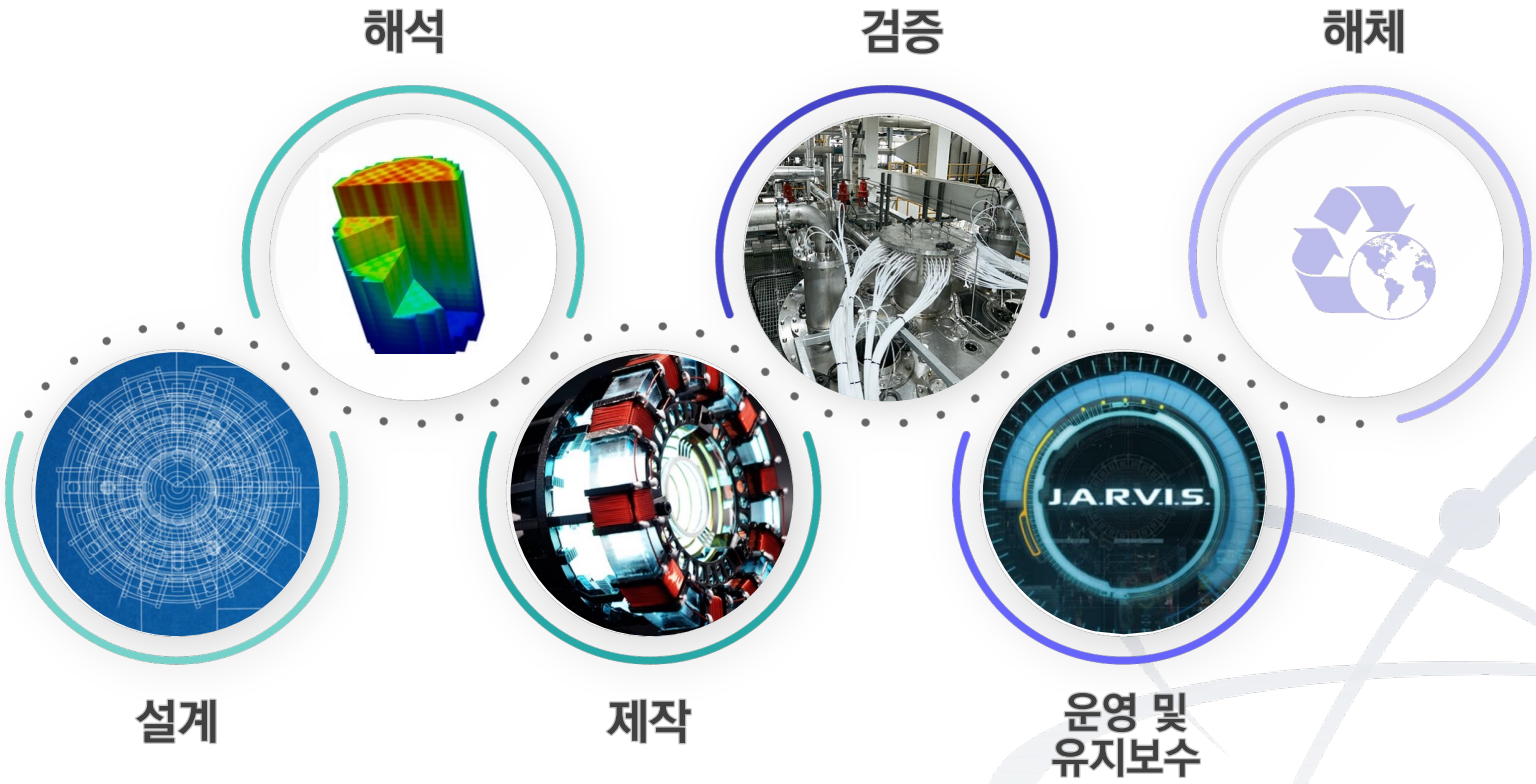
Younghwan Joo[1], Yonggyun Yu[2], and In Gwun Jang[3], Member, IEEE
[1]Korea Institute of Energy Research, 152 Gajeong-ro, Yuseong-gu, Daejeon 34129, Republic of Korea
[2]Korea Atomic Energy Research Institute, 111, Daedeok-daero 989 beon-gil, Yuseong-gu, Daejeon 34057, Republic of Korea
[3]Korea Advanced Institute of Science and Technology, Yuseong-gu, Daejeon 34141, Republic of Korea



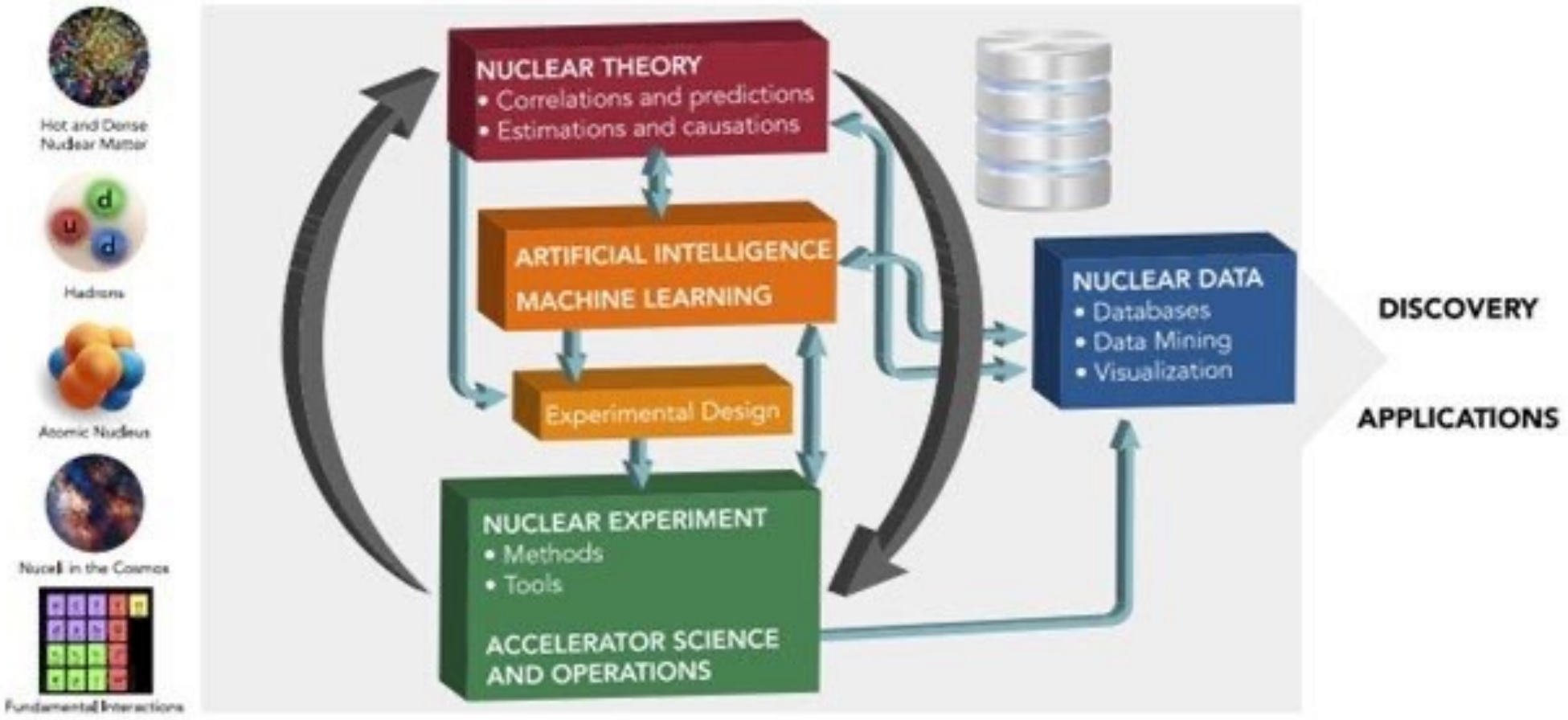
02 원자력분야의 인공지능 응용

함수를 만드는 일에 모두 적용할 수 있지 않을까?

$$f(x)$$

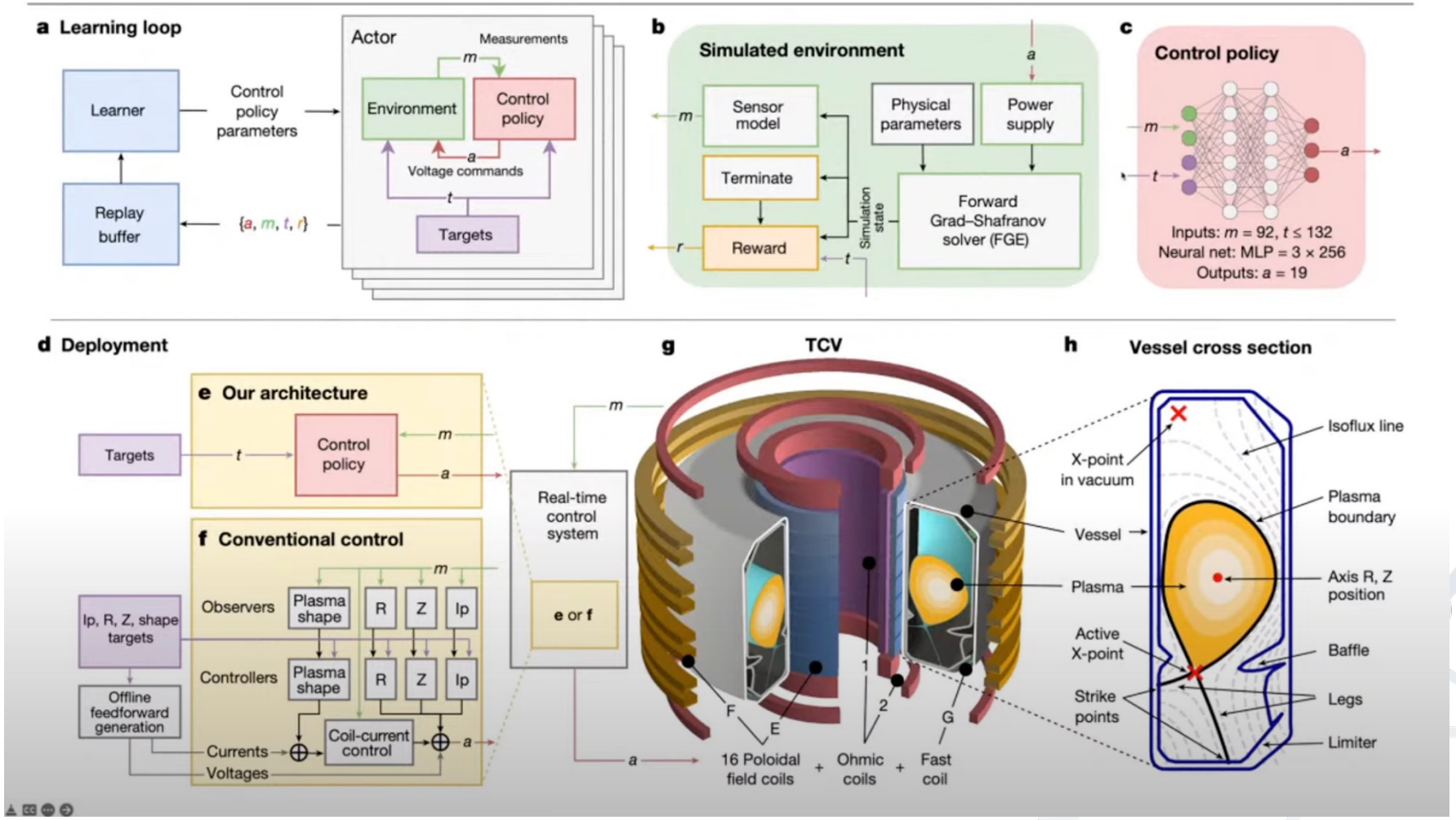


원자로 시뮬레이션 해석의 인공지능 적용



https://www.energy.gov/science/np/articles/machine-learning-takes-hold-nuclear-physics

인공지능을 활용한 핵융합 원자로 제어

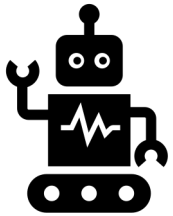


03

시스템 측면의 AI

한국원자력연구원
KAERI
Korea Atomic Energy Research Institute

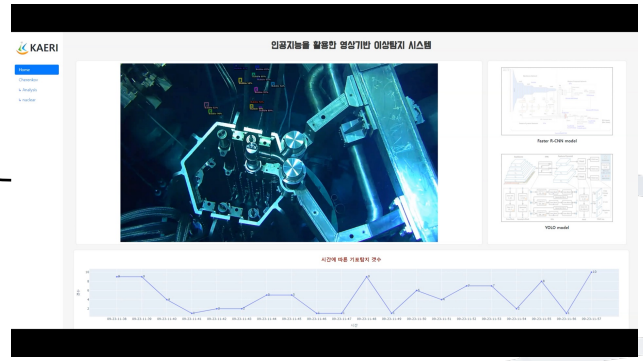
원자로 AI 운전원



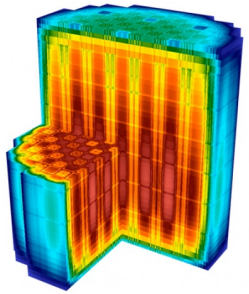
AI 운전원

Manual

Sensor Signal

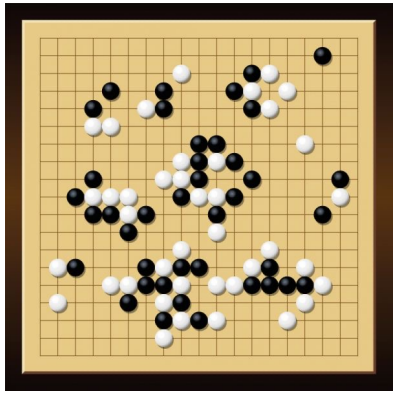


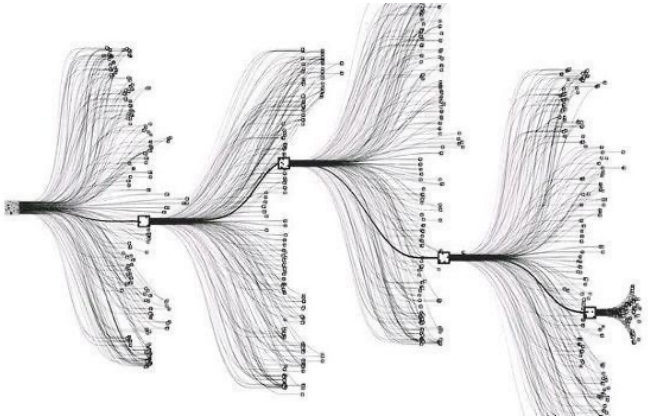
CCTV



Simulation

인공지능의 분류 (인식 vs 판단)

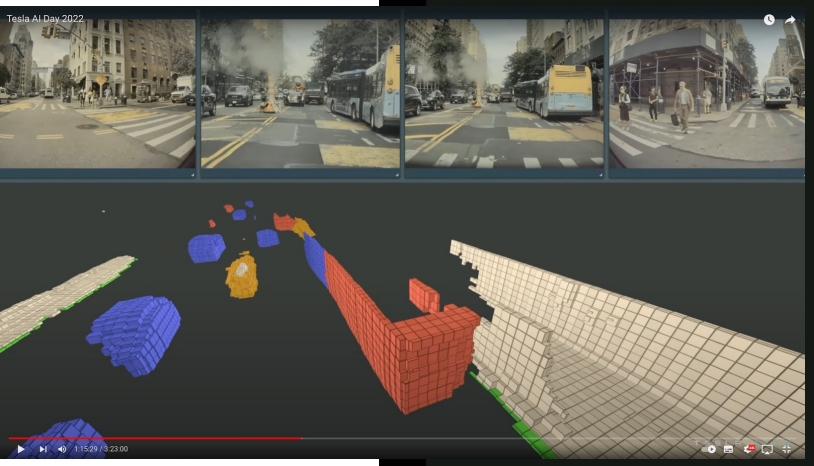


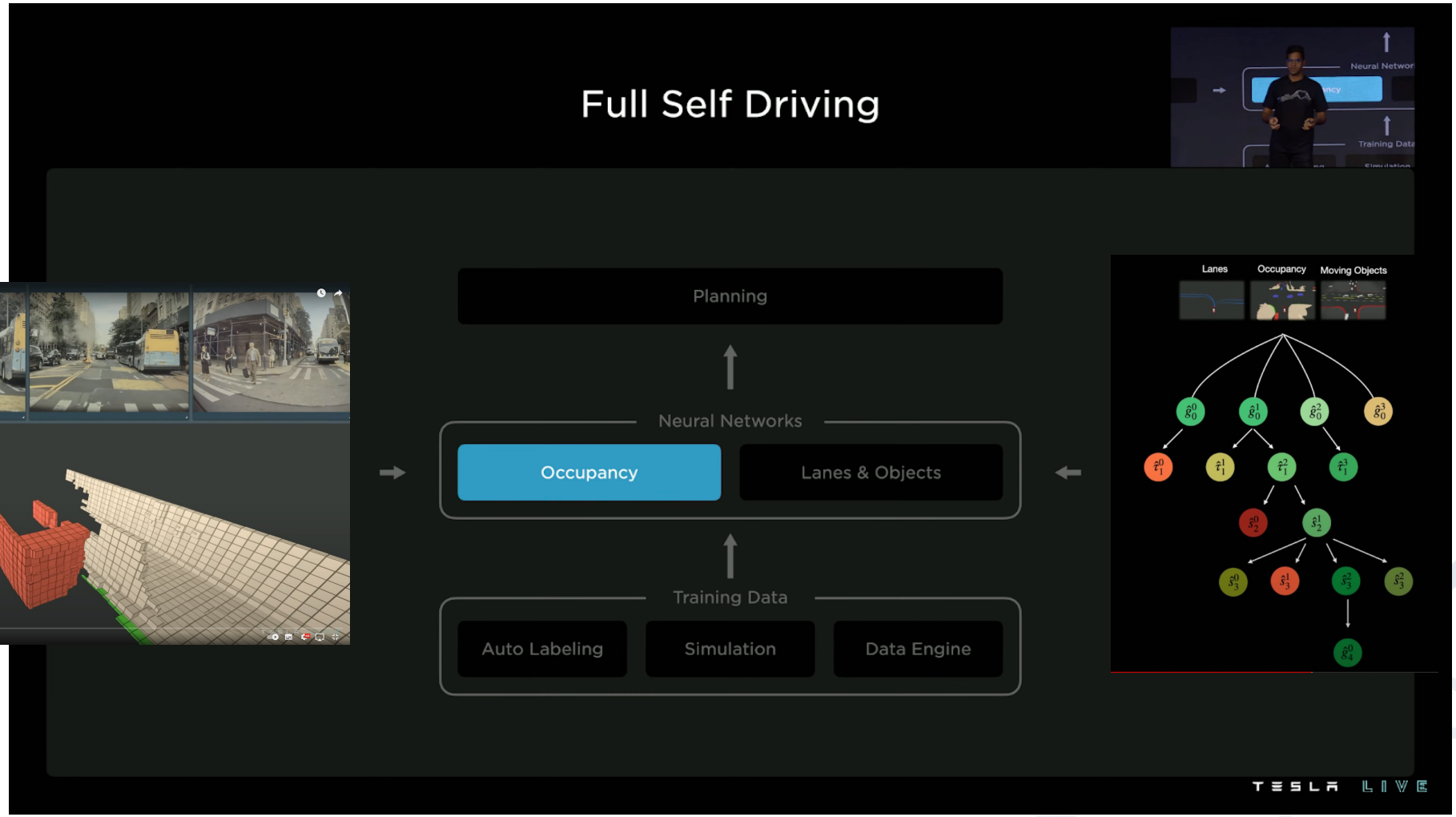


인식

판단

테슬라의 자율주행 기술





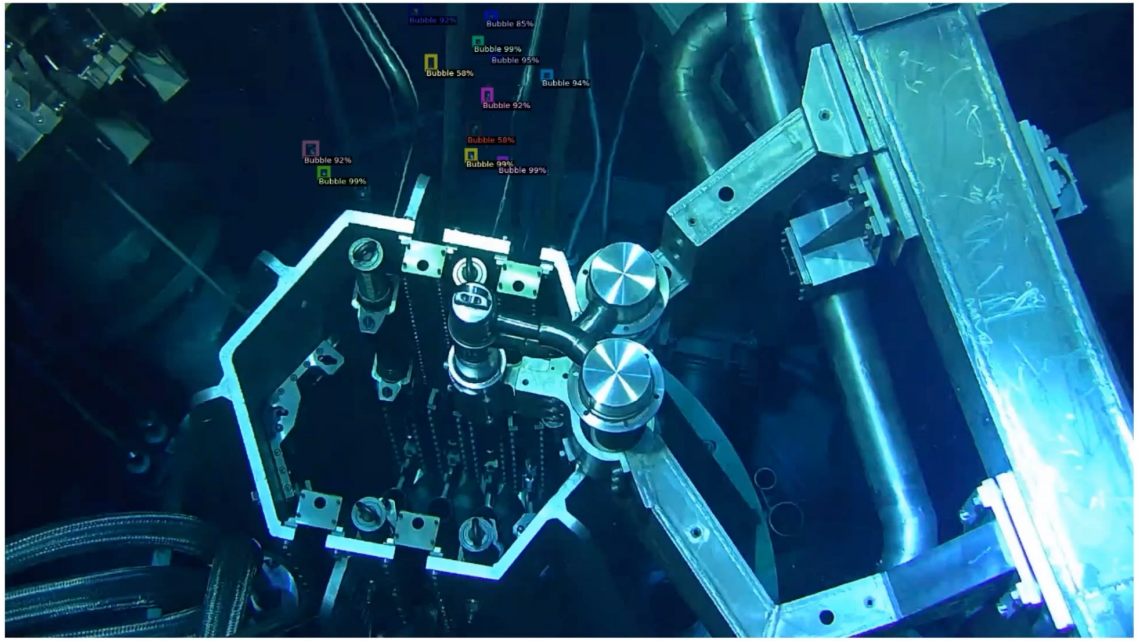
출처: 테슬라 AI Day 2

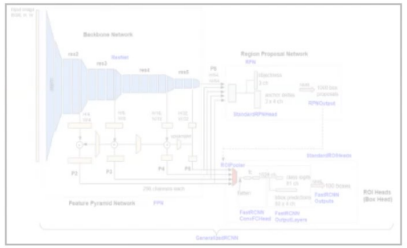
CCTV를 이용한 하나로 상태 감시

KAERI

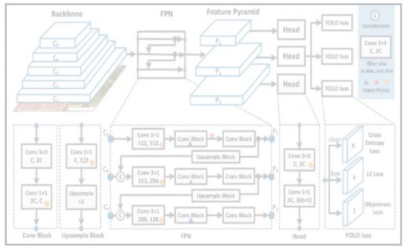
- Home
- Cherenkov
  - ↳ Analysis
  - ↳ nuclear

인공지능을 활용한 영상기반 이상탐지 시스템





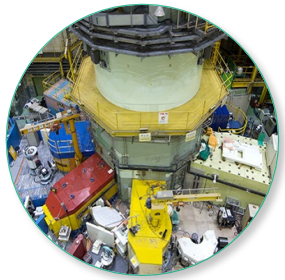
Faster R-CNN model



YOLO model

03 하나로 원자로의 데이터 기반 이상탐지

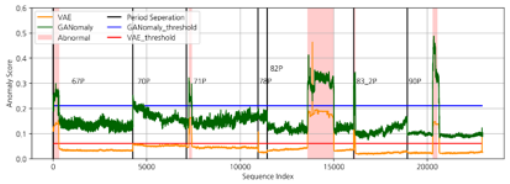
한국원자력연구원
Korea Atomic Energy Research Institute



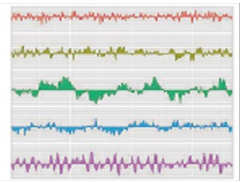
연구용 원자로 [하나로]

이상탐지 기술 동작 개요

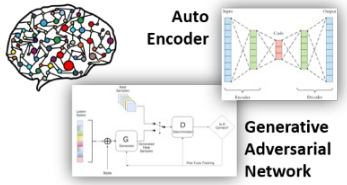
시계열 이상 탐지 및 예측



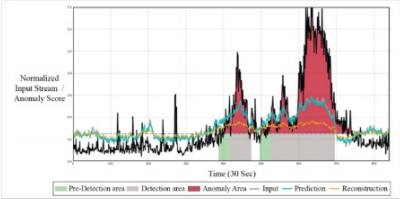
하나로 데이터 활용 이상탐지 결과



시계열 운전 데이터

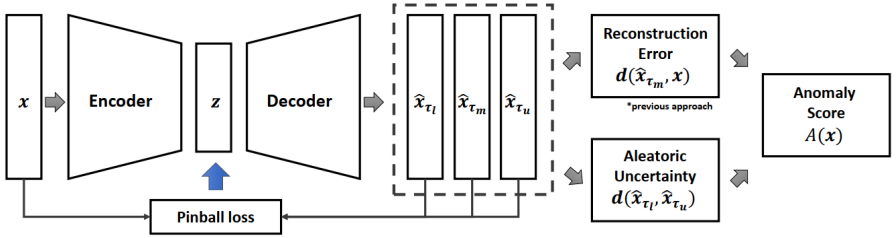


최적의 딥러닝 모델 탐색

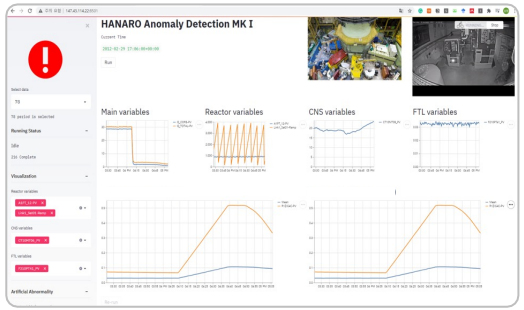


이상 탐지 및 예측

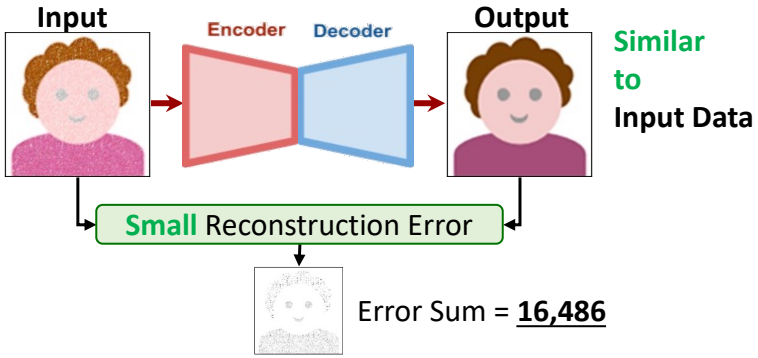
기술 개발 과정



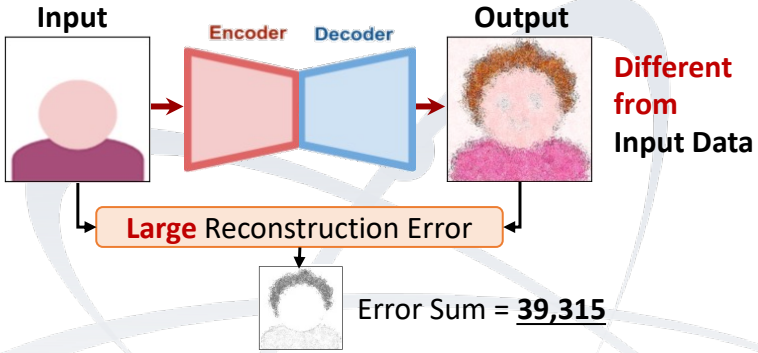
오토인코더 및 불확실성을 고려한 딥러닝 이상탐지 모델 구조



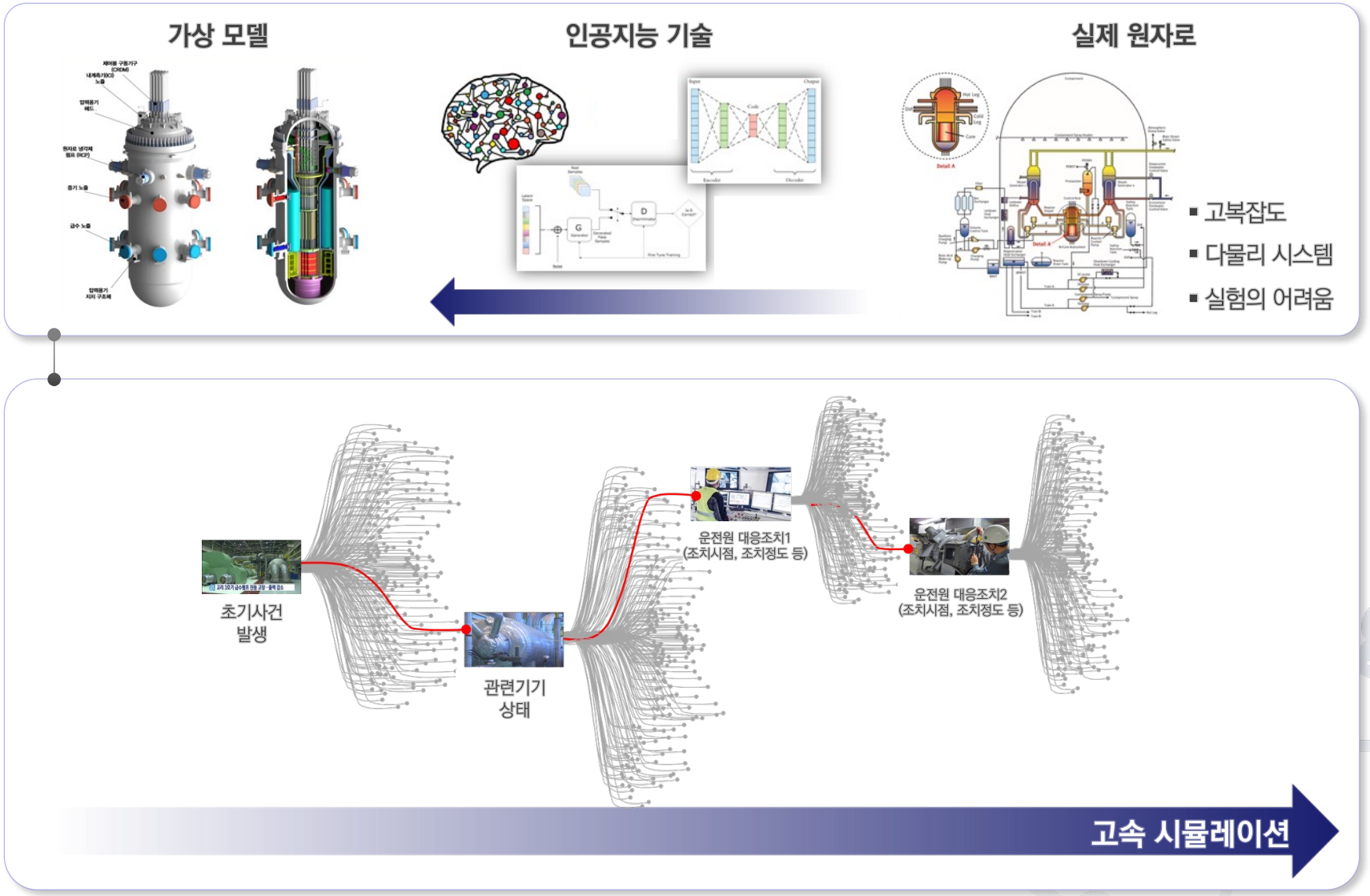
- 정상 데이터에 대한 모델 입출력 차이



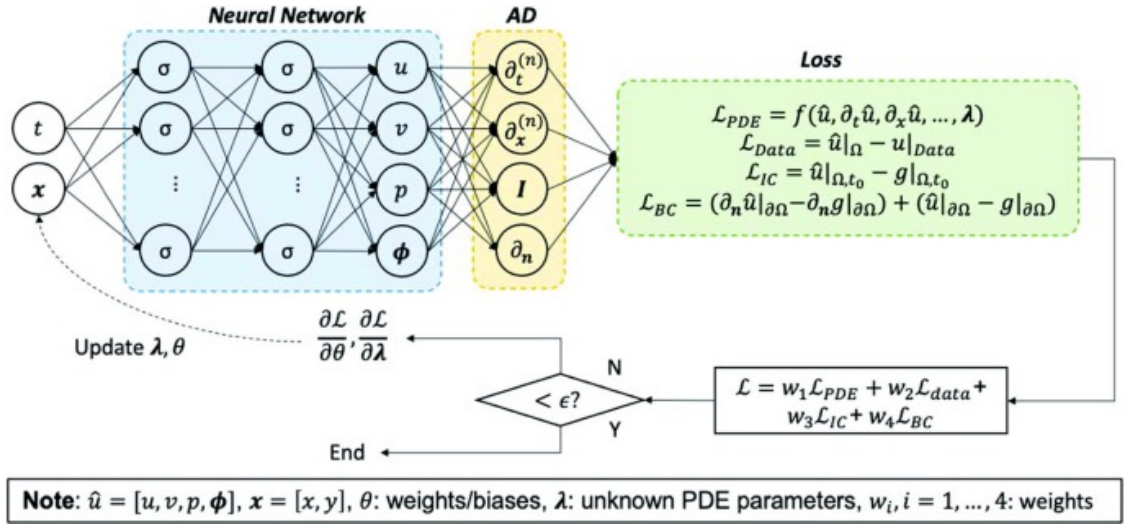
- 비정상 데이터에 대한 모델 입출력 차이



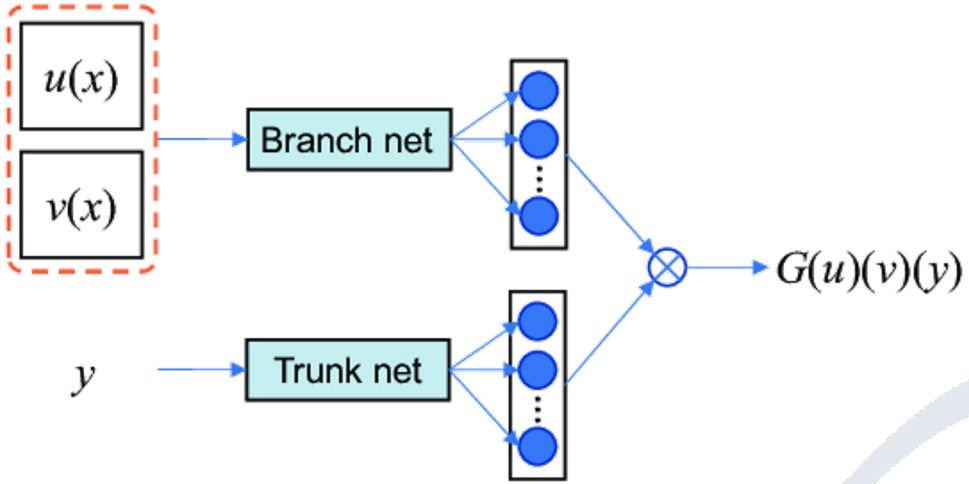
원자로 리스크 예측 및 사고대응



Physics Informed Machine Learning (PIML)



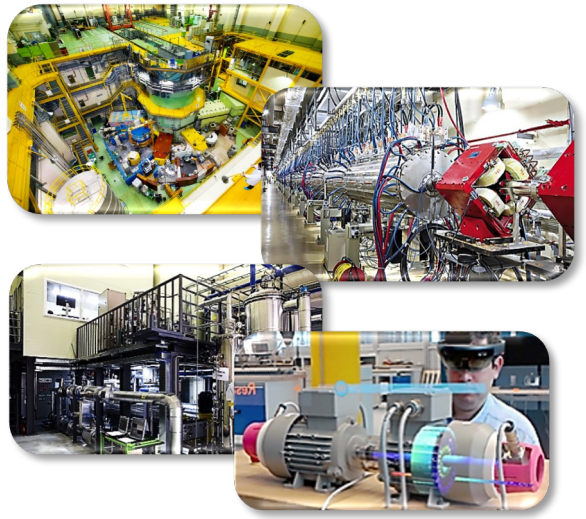
Physics Informed Neural Network

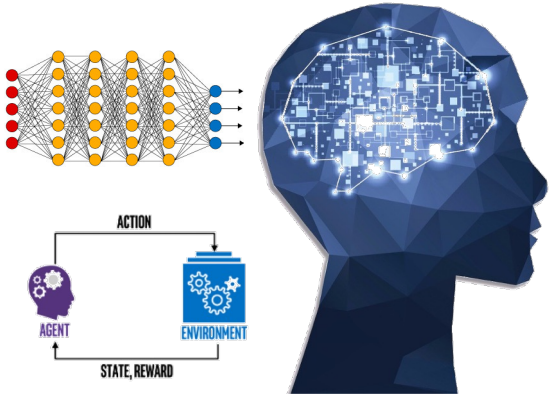


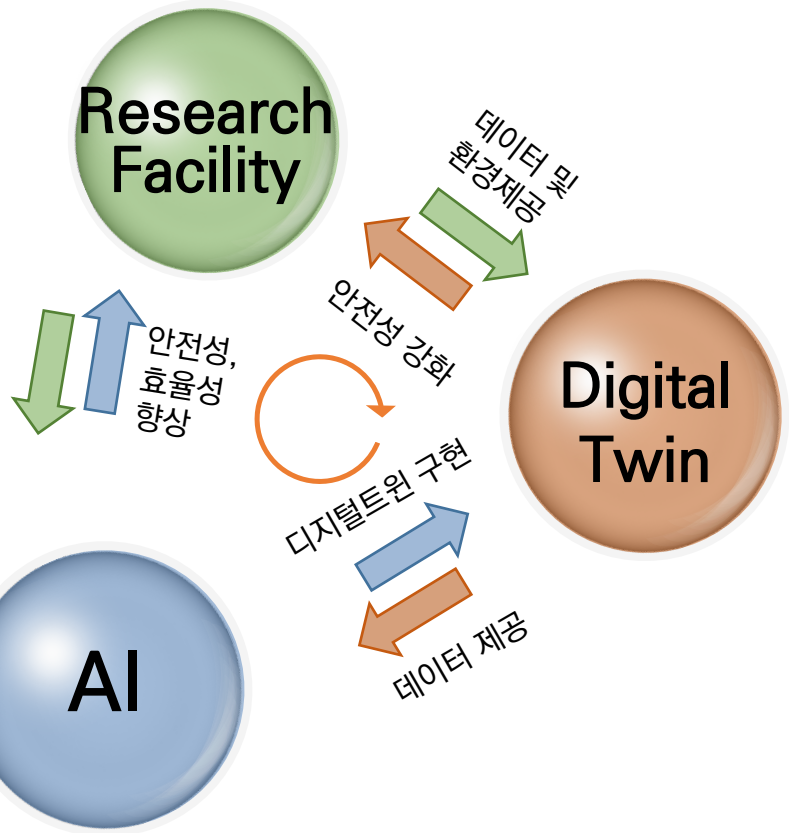
DeepONet

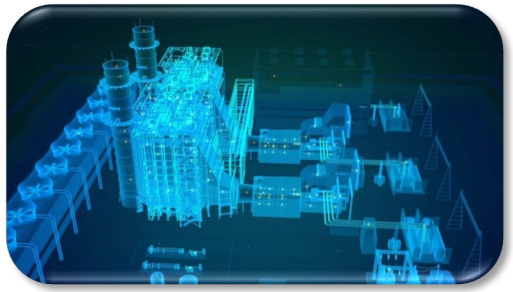
* Physics-informed neural networks (PINNs) for fluid mechanics: a review
** Enhanced DeepONet for Modeling Partial Differential Operators Considering Multiple Input Functions

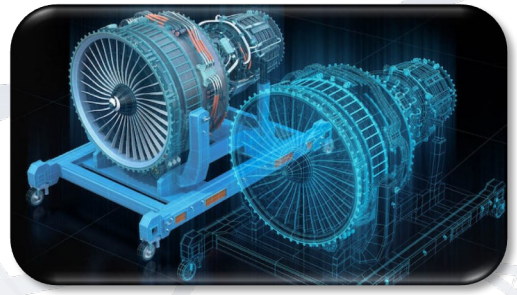
03 인공지능과 디지털트윈 관계 (AI+DT)





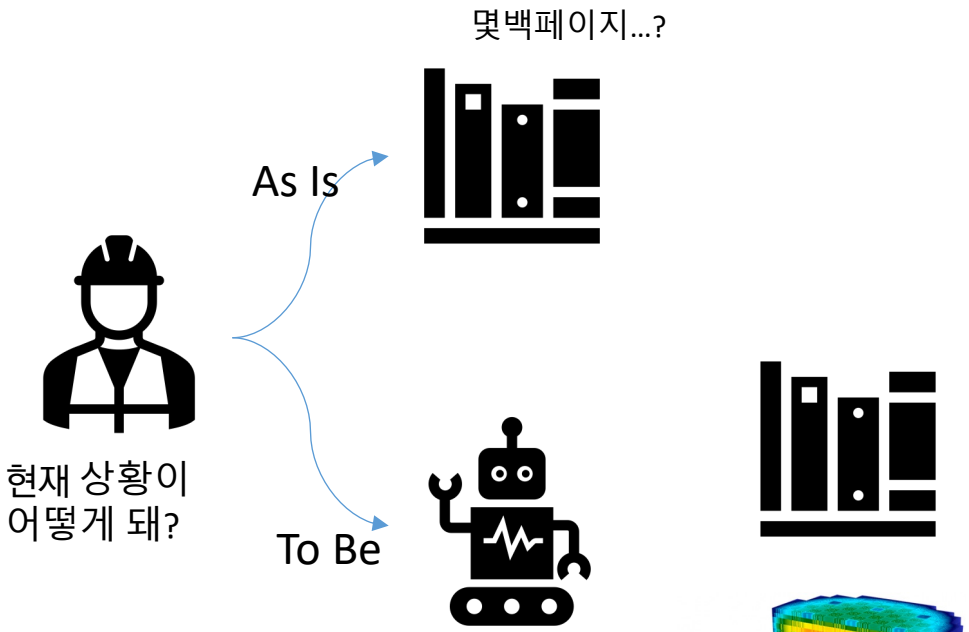


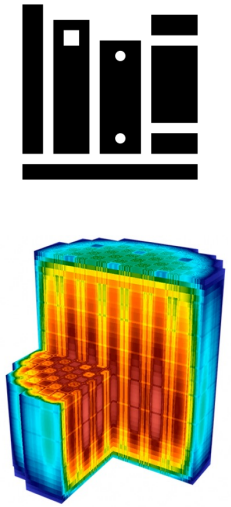




03 인공지능 기반 AI 운전원

여전히 풀리지 않는 숙제…





현재 상황이 유지되면
1시간 후에 핵연료 손상이
예상됩니다.
절차서 상으로 ~하하도록
대응하도록 지침이 정해져
있습니다.
조치 방안은 A, B, C가
있으며, 시뮬레이션 결과
A –> C -> B 순서로
작업을 수행하는 것이
가장 안전할 것으로
판단됩니다.

휴머노이드 파일럿 ‘파이봇’

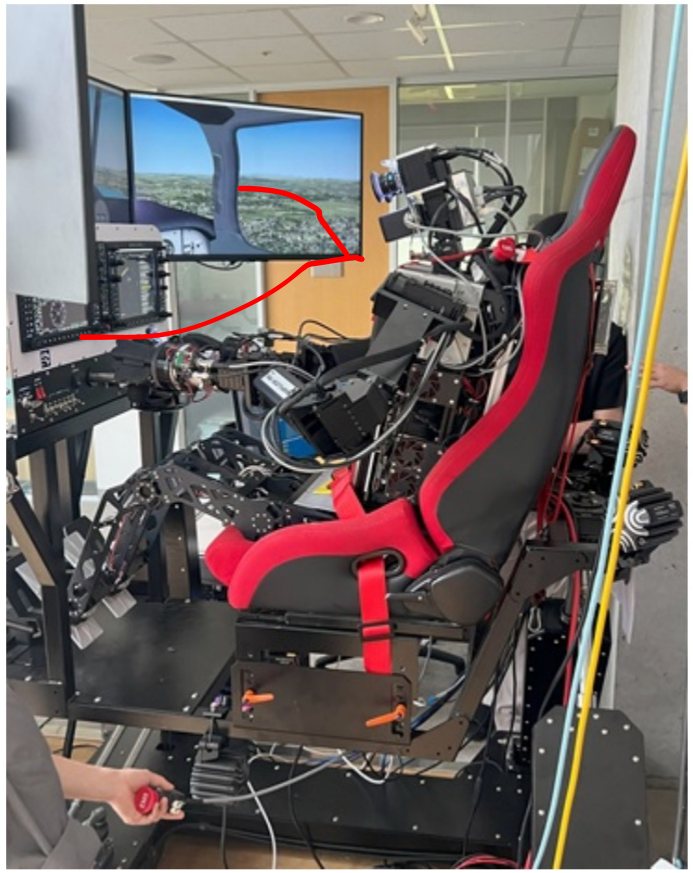
홈 > FOCUS > 기획

KAIST, 세계 최초 인공지능 자연어처리 기반 휴머노이드 파일럿 ‘파이봇’ 개발

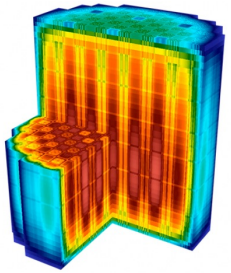
최광민 기자 | 입력 2023.07.19 08:15 | 댓글 0

가 가

> 전기및전자공학부 심현철 교수, 김재철AI대학원 주재걸 교수, 기계공학부 윤국진 교수, 전기및전자공학부 김민준 교수의 협업으로 로봇은 인간 조종사에게는 불가능한 전 세계 항공차트를 전부 기억하여 실수 없는 조종이 가능하며, 특히 최근 큰 이슈가 된 챗GPT를 적용해 항공기 조작 매뉴얼 및 비상 대처절차를 담은 자료(QRH)를 기억, 즉각적으로 대응

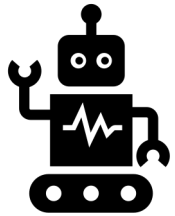


원자로 AI 운전원



Digital Twin

예측



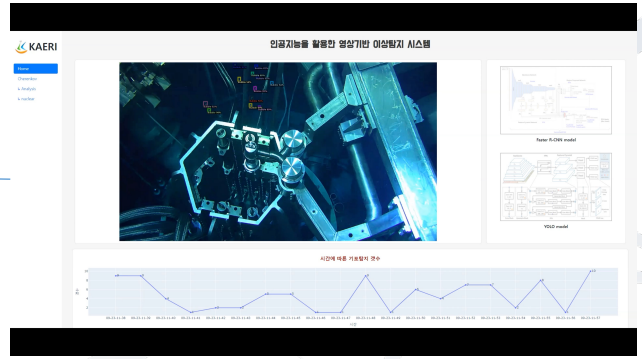
AI 운전원

인식

Sensor Signal

문서(복합정보)이해

Manual



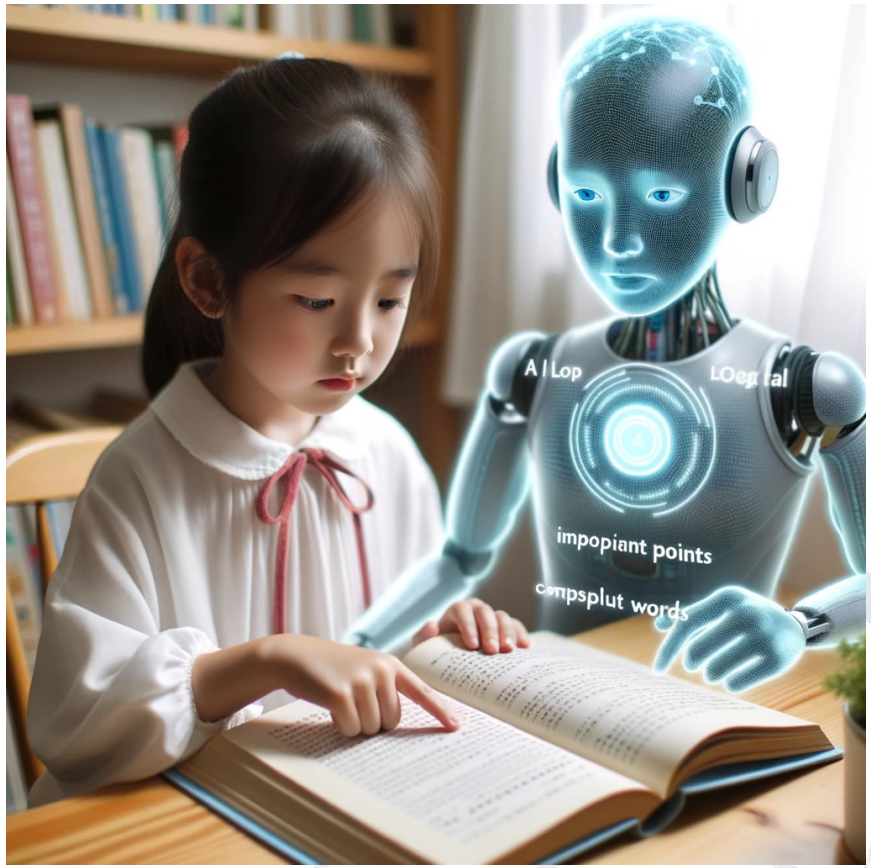
CCTV

04 Foundation Model for 원자력

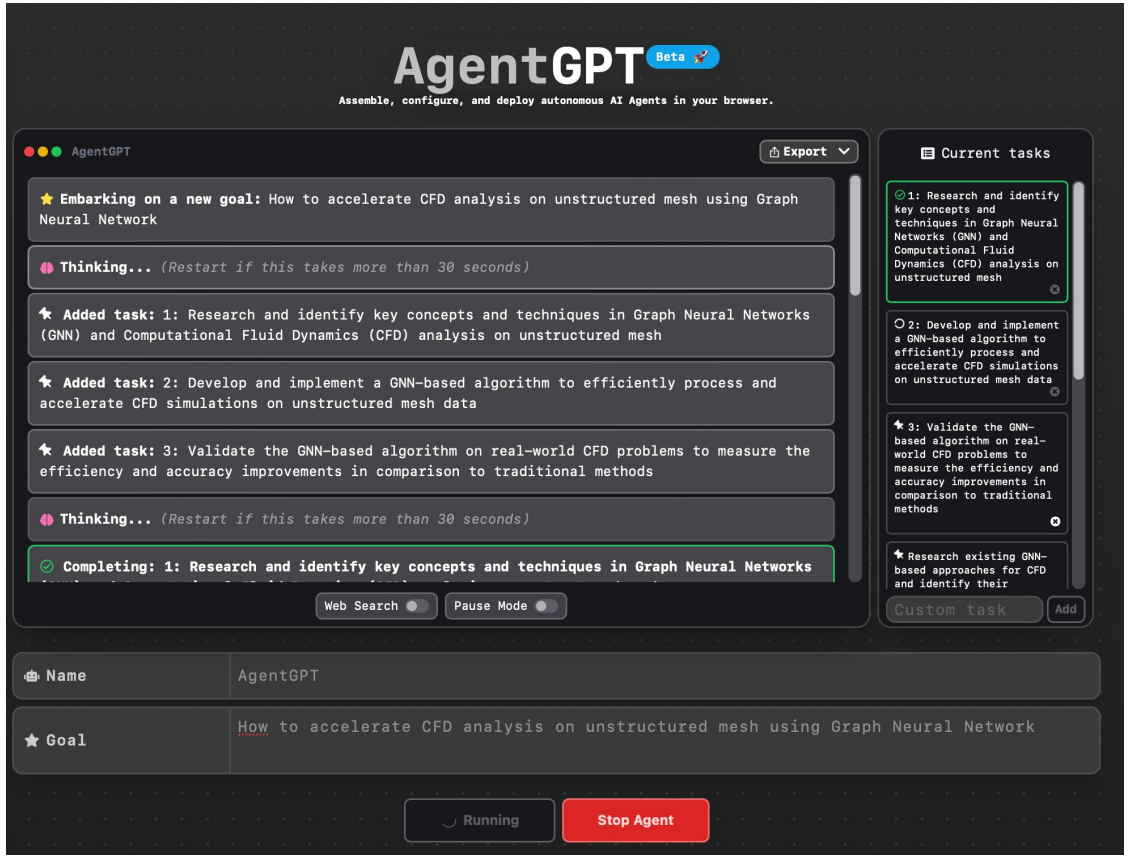
원자력 특화 언어모델 개발

- AtomGPT(?)
- **활용방안**
  - 원자력관련 지식 질의 응답
  - 규정 & 설계문서 질의 응답
  - 문서 검색
  - 설계 문서 자동 생성

원자력을 위한 멀티모달 인공지능



02 AgentGPT



Completing: 1: Research and identify key concepts and techniques in Graph Neural Networks (GNN) and Computational Fluid Dynamics (CFD) analysis on unstructured mesh

To achieve the overall objective of accelerating CFD analysis on unstructured mesh using Graph Neural Networks (GNN), we can start by completing the first sub-task, which involves conducting research on key concepts and techniques in GNN and CFD analysis on unstructured meshes. The following sections outline the key concepts and techniques involved in GNNs and CFD analysis on unstructured meshes.

Graph Neural Networks (GNN):

1. Graph Representation: Graphs are mathematical structures consisting of vertices (nodes) and edges. An edge connects two vertices, and it may have a weight assigned to it. Graphs can be represented by adjacency matrices or adjacency lists. Specifically, for GNN, graphs are represented as tensors to enable computational processing.

2. Graph Convolutional Networks (GCN): GCNs extend convolutional neural networks to work on graph-structured data. They use local neighborhood information to generate node embeddings. The key idea in GCNs is to propagate information from a node to its neighbors through convolutional operations. The update equation for a node embedding can be written as:

Completing: 2: Develop and implement a GNN-based algorithm to efficiently process and accelerate CFD simulations on unstructured mesh data

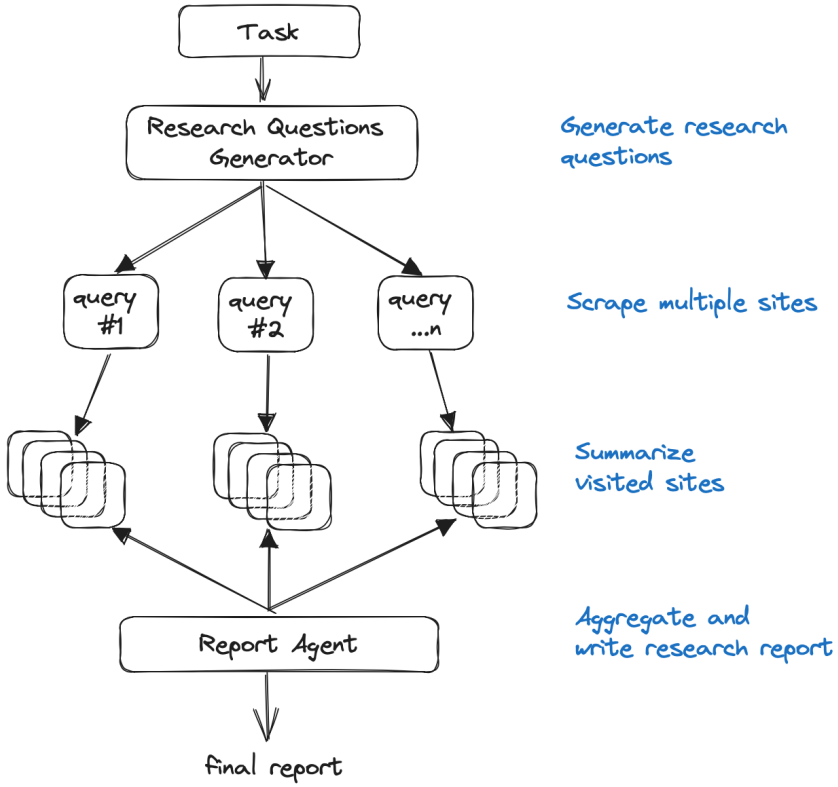
To perform the sub-task 2, we'll follow these steps to develop and implement a Graph Neural Network (GNN)-based algorithm to efficiently process and accelerate Computational Fluid Dynamics (CFD) simulations on unstructured mesh data.

1. **Understand the CFD problem and unstructured mesh data:** Before implementing the GNN-based algorithm, it is crucial to understand the CFD problem you want to solve and how the data is organized in unstructured mesh format. Unstructured mesh data consist of a set of nodes with their respective coordinates and their connectivity in the form of elements (i.e., how nodes are connected to form cells such as triangles or tetrahedrons). This information can be used to create a graph representation of the mesh.

2. **Choose a suitable GNN framework:** There are various GNN frameworks available, such as PyTorch Geometric, DGL (Deep Graph Library), or Spektral. Choose a suitable framework based on your requirements and familiarity with the programming language. For example, if you are comfortable with Python and PyTorch, PyTorch Geometric could be a suitable choice.

3. **Convert unstructured mesh data into graph representation:** Convert the unstructured mesh

02 ResearchGPT



U.S. Nuclear Industry's Strategy For Utilizing Artificial Intelligence

The United States nuclear industry is currently in a transitional phase. Amid the push towards implementing sustainable, clean sources of energy for the future, utilizing modern technologies such as artificial intelligence (AI) is increasingly vital. This report provides an in-depth look at the strategies employed by the U.S. nuclear industry in leveraging AI technologies.

Introduction

AI's potential applications to nuclear power and projected benefits deserve constant attention. Taking cues from other sectors that have utilized AI to streamline operations, reduce costs, and increase efficiencies, the U.S nuclear sector aims to integrate AI into its operations. However, this must be done strategically and maintaining a safety-first culture due to the high-stakes nature of nuclear operations.

AI in Nuclear Operations

One key area where AI can offer substantial contributions is in nuclear facilities' operations. This includes tasks such as asset management, maintenance, and even the handling of nuclear materials. AI-powered systems can be integrated to continuously monitor and manage assets, predicting potential breakdowns, and scheduling maintenance tasks efficiently to extend lifespan and reduce unplanned downtime. Such applications have already seen success in other energy sectors, hinting at similar potential benefits for the nuclear sector.

Further, AI-powered robotics can handle hazardous work in nuclear facilities, reducing risks for the human workforce. AI can also assist in the monitoring and detection of radiation, automatically alerting staff to potential leaks or escalation of radiation levels.

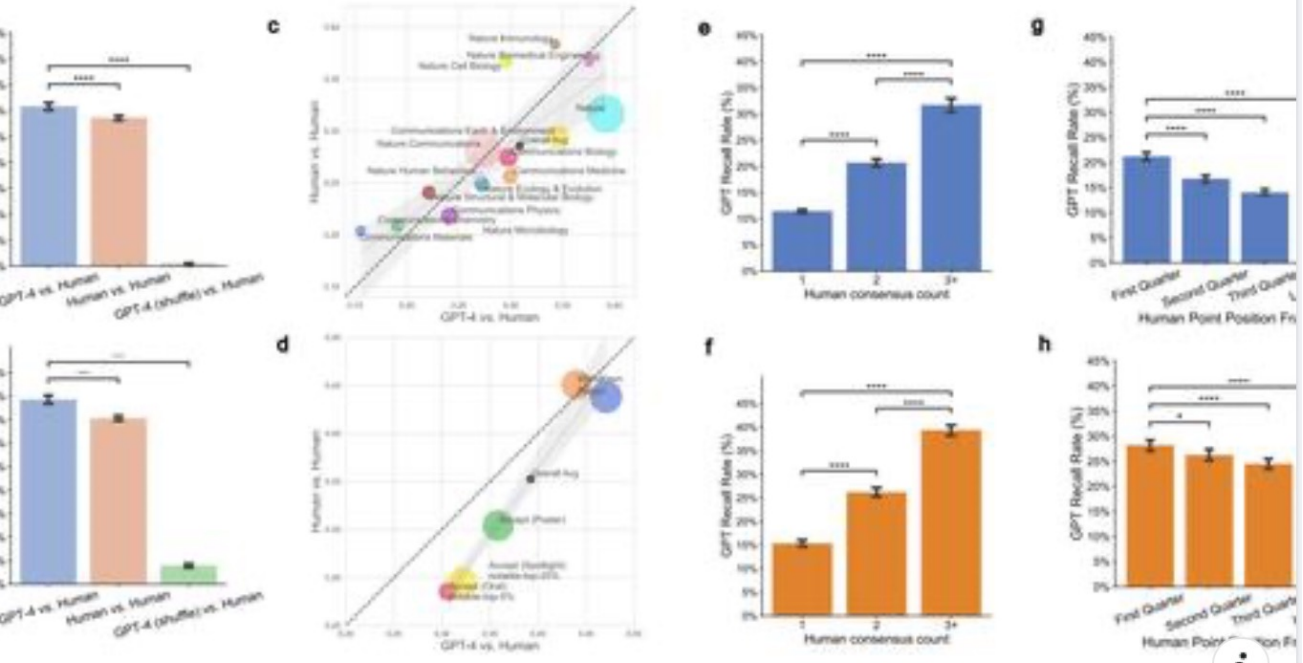
AI is also being explored for its potential applications in enriching nuclear materials for energy production. Machine learning algorithms can be utilized to enhance uranium enrichment processes, optimizing parameters for efficiency and reduced energy consumption.

AI in Nuclear Safety and Regulation

AI has promising applications to improve safety and regulation standards within the nuclear industry. AI can be employed to boost security through intelligent surveillance, potentially identifying unauthorized access. Through predictive modeling, AI can also assist in anticipating and proactively dealing with potential accidents, enhancing safety.

Jehyun Lee
4일 ·

● LLM 활용 연구 논문 리뷰 = 인간 리뷰어 동등 수준
- 언어모델이 리뷰를 할 수 있을만한 성능을 가지고 있는지에 대한 논문이 발표되었습니다.
- 스탠포드 대학에서 GPT-4로 리뷰한 결과를 실제 리뷰어들의 결과와 비교하는 방식으로 진행한 연구에서 매우 도움이 된다는 답변을 얻었습니다.
- Nature(3096편 8745 리뷰)와 ICLR(1709편 6505 리뷰) 논문을 대상으로 한 GPT-4와 인간 리뷰어의 일치도는 인간 리뷰어간 일치도와 유사한 수준입니다.
- Nature는 Communication 계열에서 일치도가 20% 이하로 낮았고, ICLR은 Withdrawn, Reject 판정에서 일치도가 45% 이상으로 높았습니다.
- 공개된 논문을 거대 언어 모델에 입력하는 데는 보안 및 윤리적 이슈가 없으므로, 여러분의 연구에 효과적으로 활용하시기를 바랍니다.
- 관련 기사: https://phys.org/.../2023-10-large-language-peer-review.html
- 해당 논문: https://arxiv.org/abs/2310.01783



PHYS.ORG
Large language models prove helpful in peer-review process
In an era plagued by malevolent sources flooding the internet with misrepresentations, dist...

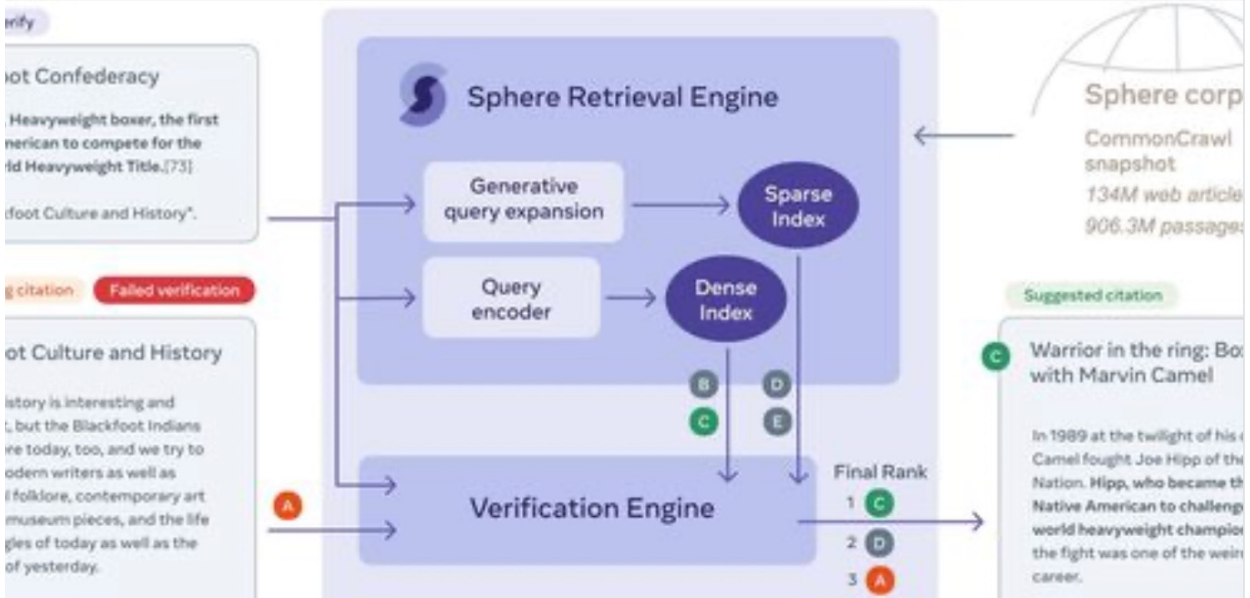
전종홍
3일 ·

어제 nature machine intelligence에 또 하나의 재미있는 논문이 나왔네요. 인간이 세계 지식을 탐색, 해석 및 검증하는 데 도움을 주는 기계를 구축하는 것인 동시에, AI를 이용하여 사람의 집단지성 검증에 도움을 주고자하는 시도네요. Meta의 연구 결과

제목: Improving Wikipedia verifiability with AI

요약: 검증 가능성은 Wikipedia의 핵심 콘텐츠 정책입니다. 즉, 주장은 인용에 의해 뒷받침되어야 합니다. Wikipedia 참조의 품질을 유지하고 개선하는 것은 중요한 과제이며 이러한 노력에 있어 인간을 지원할 더 나은 도구가 절실히 필요합니다. 우리는 정보 검색 시스템과 언어 모델을 기반으로 하는 인공 지능(AI)의 도움으로 참조 개선 프로세스를 해결할 수 있음을 보여줍니다. 우리가 SIDE라고 부르는 이 신경망 기반 시스템은 자신의 주장을 뒷받침할 가능성이 없는 Wikipedia 인용을 식별한 다음 웹에서 더 나은 인용을 추천할 수 있습니다. 우리는 기존 Wikipedia 참조를 바탕으로 이 모델을 훈련하므로 수천 명의 Wikipedia 편집자의 기여와 지혜를 결합하여 학습합니다. 크라우드소싱을 사용하여 우리는 시스템에서 검증할 수 없는 것으로 태그가 지정될 가능성이 가장 높은 상위 10% 인용에 대해 인간이 원래 인용된 참조와 비교하여 시스템에서 제안한 대안을 70% 더 선호한다는 사실을 관찰했습니다. 우리 시스템의 적용 가능성을 검증하기 위해 우리는 영어권 Wikipedia 커뮤니티에 참여하기 위한 데모를 구축했으며 SIDE의 첫 번째 인용 권장 사항이 다음과 같이 검증할 수 없을 가능성이 가장 높은 동일한 주장에 대해 기존 Wikipedia 인용보다 2배 더 자주 선호된다는 사실을 확인했습니다. 옆. 우리의 결과는 AI 기반 시스템이 인간과 협력하여 Wikipedia의 검증 가능성을 향상시키는 데 사용될 수 있음을 나타냅니다.

논문: https://www.nature.com/articles/s42256-023-00726-1



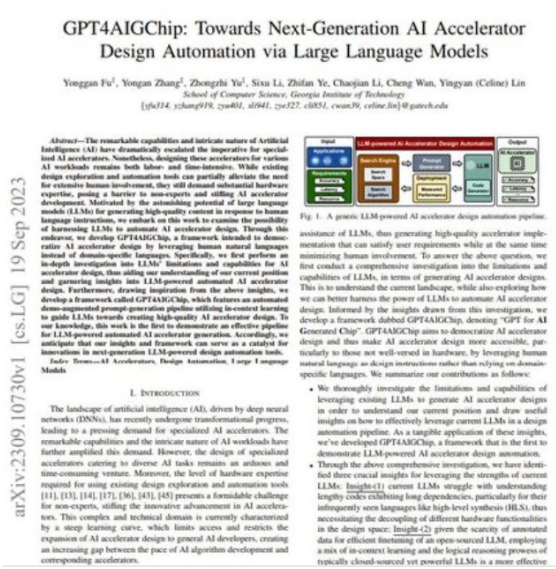
김성완
7시간 ·

GPT4AIGChip: Towards Next-Generation AI Accelerator Design Automation via Large Language Models

AI 가속 칩의 설계를 GPT4에게 시켜본 조지아텍의 연구랍니다. 아직은 인간 전문가 수준에는 못미치지만 기존 설계 탐색및 자동화 툴보다는 조금 나은 결과를 얻은 모양입니다.

논문 초록
인공지능(AI)의 놀라운 기능과 복잡한 특성으로 인해 전문 AI 가속기에 대한 필요성이 크게 증가했습니다. 그럼에도 불구하고 다양한 AI 워크로드에 맞게 이러한 가속기를 설계하는 데는 여전히 많은 인력과 시간이 소요됩니다. 기존의 설계 탐색 및 자동화 도구는 사람의 광범위한 개입의 필요성을 부분적으로 완화할 수 있지만, 여전히 상당한 하드웨어 전문 지식을 요구하기 때문에 비전문가에게는 장벽이 되고 AI 가속기 개발에 걸림돌이 됩니다. 인간의 언어 지시에 따라 고품질 콘텐츠를 생성하는 대규모 언어 모델(LLM)의 놀라운 잠재력에 주목한 우리는 LLM을 활용하여 AI 가속기 설계를 자동화할 수 있는 가능성을 검토하기 위해 이 작업에 착수했습니다. 이를 통해 우리는 도메인별 언어 대신 인간의 자연어를 활용하여 AI 가속기 설계를 대중화하기 위한 프레임워크인 GPT4AIGChip을 개발합니다. 특히, 먼저 AI 가속기 설계를 위한 LLM의 한계와 기능에 대한 심층 조사를 수행하여 현재 위치에 대한 이해를 돕고 LLM 기반의 자동화된 AI 가속기 설계에 대한 인사이트를 얻습니다. 또한, 위의 인사이트에서 영감을 얻어 상황별 학습을 활용한 자동화된 데모 증강 프롬프트 생성 파이프라인을 갖춘 GPT4AIGChip이라는 프레임워크를 개발하여 LLM이 고품질 AI 가속기 설계를 할 수 있도록 안내합니다. 저희가 알기로는 이 작업이 LLM 기반의 자동화된 AI 가속기 생성을 위한 효과적인 파이프라인을 입증한 최초의 사례입니다. 따라서 우리의 인사이트와 프레임워크가 차세대 LLM 기반 설계 자동화 도구의 혁신을 촉진하는 촉매제 역할을 할 수 있을 것으로 기대합니다.

논문: https://arxiv.org/abs/2309.10730



GPT4AIGChip: Towards Next-Generation AI Accelerator Design Automation via Large Language Models

Yonggan Fu[1], Yongan Zhang[1], Zhongzhi Yu[1], Sixu Li, Zhifan Ye, Chaojian Li, Cheng Wan, Yingyan (Celine) Lin
School of Computer Science, Georgia Institute of Technology
[yfu314, yzhang919, zyu401, sli941, zye327, cli851, cwan39, celine.lin]@gatech.edu

Abstract—The remarkable capabilities and intricate nature of Artificial Intelligence (AI) have dramatically escalated the imperative for specialized AI accelerators. Nonetheless, designing these accelerators for various AI workloads remains both labor- and time-intensive. While existing design exploration and automation tools can partially alleviate the need for extensive human involvement, they still demand substantial hardware expertise, posing a barrier to non-experts and stifling AI accelerator development. Motivated by the astonishing potential of large language models (LLMs) for generating high-quality content in response to human language instructions, we embark on this work to examine the possibility of harnessing LLMs to automate AI accelerator design. Through this endeavor, we develop GPT4AIGChip, a framework intended to democratize AI accelerator design by leveraging human natural languages instead of domain-specific languages. Specifically, we first perform an in-depth investigation into LLMs' limitations and capabilities for AI accelerator design, thus aiding our understanding of our current position and garnering insights into LLM-powered automated AI accelerator design. Furthermore, drawing inspiration from the above insights, we develop a framework called GPT4AIGChip, which features an automated demo-augmented prompt-generation pipeline utilizing in-context learning to guide LLMs towards creating high-quality AI accelerator design. To our knowledge, this work is the first to demonstrate an effective pipeline for LLM-powered automated AI accelerator generation. Accordingly, we anticipate that our insights and framework can serve as a catalyst for innovations in next-generation LLM-powered design automation tools.

Index Terms—AI Accelerators, Design Automation, Large Language Models

I. Introduction

The landscape of artificial intelligence (AI), driven by deep neural networks (DNNs), has recently undergone transformational progress, leading to a pressing demand for specialized AI accelerators. The remarkable capabilities and the intricate nature of AI workloads have further amplified this demand. However, the design of specialized accelerators catering to diverse AI tasks remains an arduous and time-consuming venture. Moreover, the level of hardware expertise required for using existing design exploration and automation tools [11], [13], [14], [17], [36], [43], [45] presents a formidable challenge for non-experts, stifling the innovative advancement in AI accelerators. This complex and technical domain is currently characterized by a steep learning curve, which limits access and restricts the expansion of AI accelerator design to general AI developers, creating an increasing gap between the pace of AI algorithm development and corresponding accelerators.

Fig. 1. A generic LLM-powered AI accelerator design automation pipeline.

assistance of LLMs, thus generating high-quality accelerator implementation that can satisfy user requirements while at the same time minimizing human involvement. To answer the above question, we first conduct a comprehensive investigation into the limitations and capabilities of LLMs, in terms of generating AI accelerator designs. This is to understand the current landscape, while also exploring how we can better harness the power of LLMs to automate AI accelerator design. Informed by the insights drawn from this investigation, we develop a framework dubbed GPT4AIGChip, denoting "GPT for AI Generated Chip". GPT4AIGChip aims to democratize AI accelerator design and thus make AI accelerator design more accessible, particularly to those not well-versed in hardware, by leveraging human natural language as design instructions rather than relying on domain-specific languages. We summarize our contributions as follows:

- We thoroughly investigate the limitations and capabilities of leveraging existing LLMs to generate AI accelerator designs in order to understand our current position and draw useful insights on how to effectively leverage current LLMs in a design automation pipeline. As a tangible application of these insights, we've developed GPT4AIGChip, a framework that is the first to demonstrate LLM-powered AI accelerator design automation.
- Through the above comprehensive investigation, we have identified three crucial insights for leveraging the strengths of current LLMs: Insight-(1) current LLMs struggle with understanding lengthy codes exhibiting long dependencies, particularly for their infrequently seen languages like high-level synthesis (HLS), thus necessitating the decoupling of different hardware functionalities in the design space; Insight-(2) given the scarcity of annotated data for efficient finetuning of an open-sourced LLM, employing a mix of in-context learning and the logical reasoning prowess of typically closed-sourced yet powerful LLMs is a more effective

arXiv:2309.10730v1 [cs.LG] 19 Sep 2023

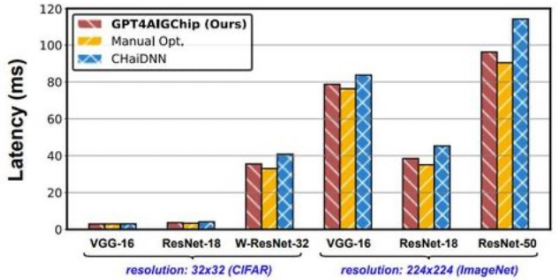
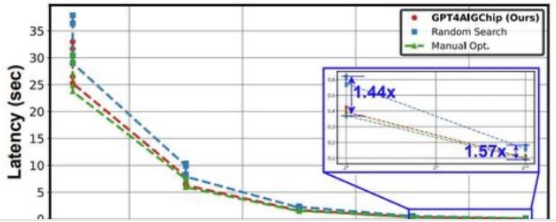


Fig. 6. Benchmark GPT4AIGChip's generated designs with CHaiDNN's [40] generated ones and manually optimized ones by experts on six networks.



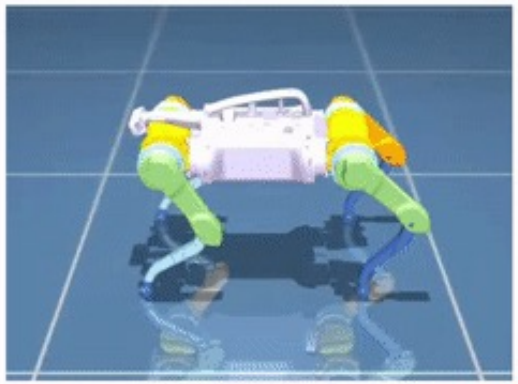
BLOG ›

Language to rewards for robotic skill synthesis

TUESDAY, AUGUST 22, 2023

Posted by Wenhao Yu and Fei Xia, Research Scientists, Google

User



04 AI의 한계

설명력

재현성

비용

데이터 편향

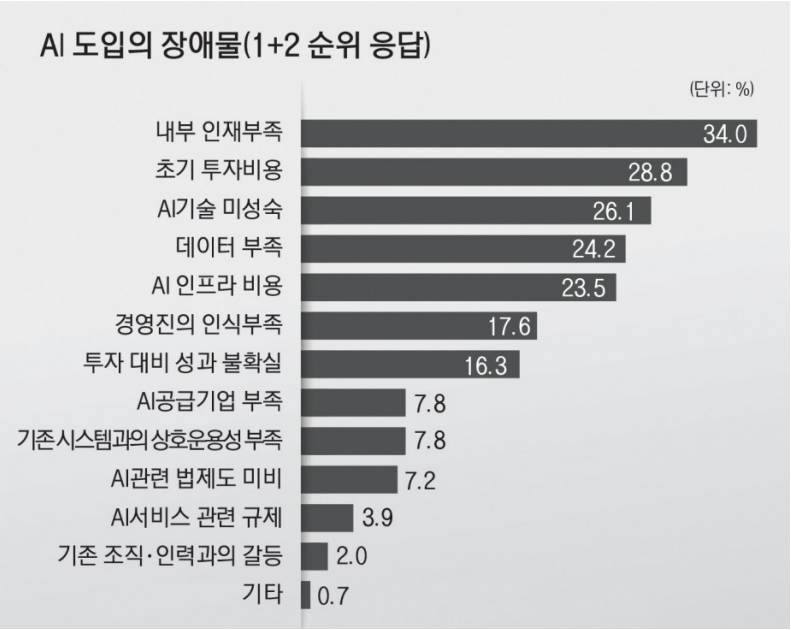
COMPUTER SCIENCE

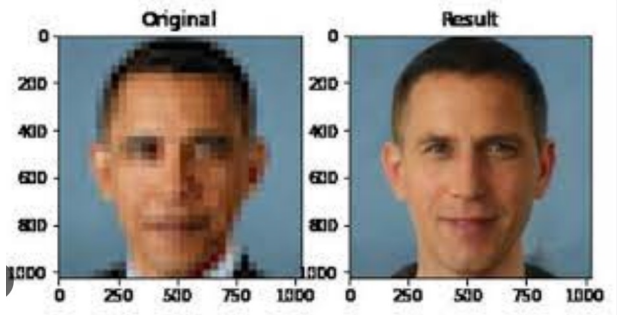
Artificial intelligence faces reproducibility crisis

Unpublished code and sensitivity to training conditions make many claims hard to verify

By **Matthew Hutson**

(AAAI) in New Orleans, Louisiana, r





원자력 AI의 어려움

데이터 부족

- 비정상데이터 부족
- 데이터 확보의 어려움

규제

- 원자력 AI와 관련된 규제 존재 X
- 실제 환경에 적용하기 위한 복잡다단한 인허가 과정

신뢰성

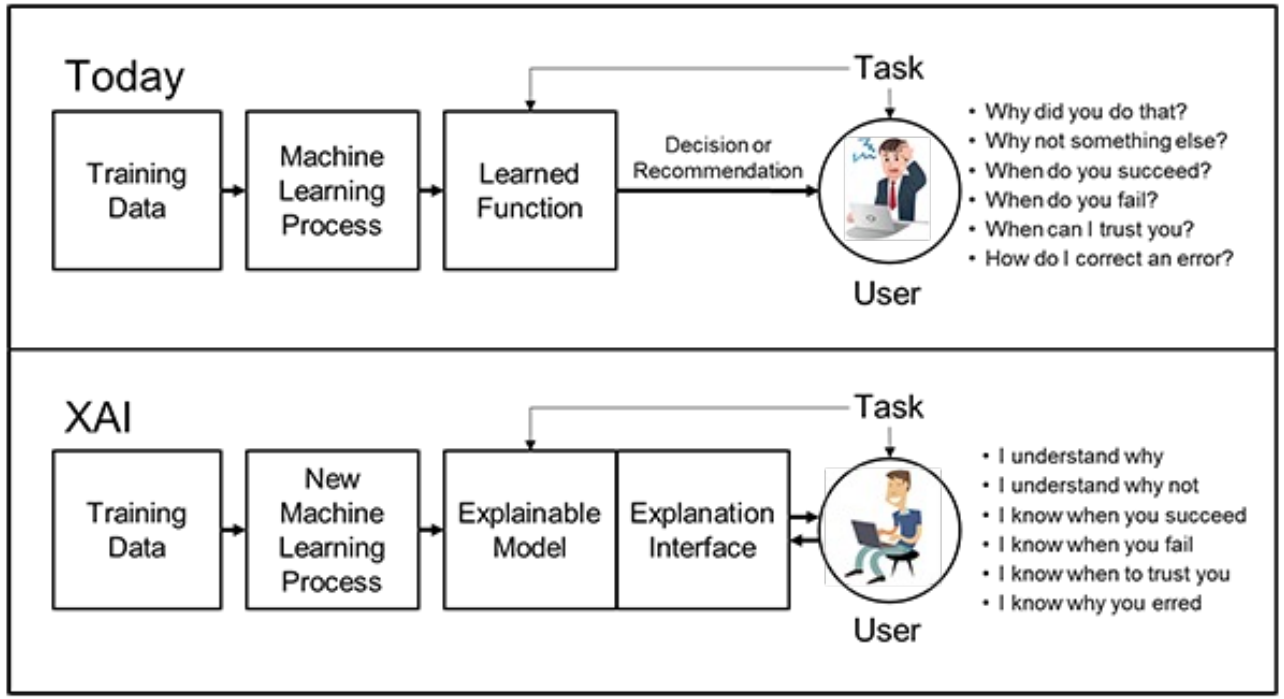
- AI의 경우 블랙박스이기 때문에 결과에 대한 설명불가

보안

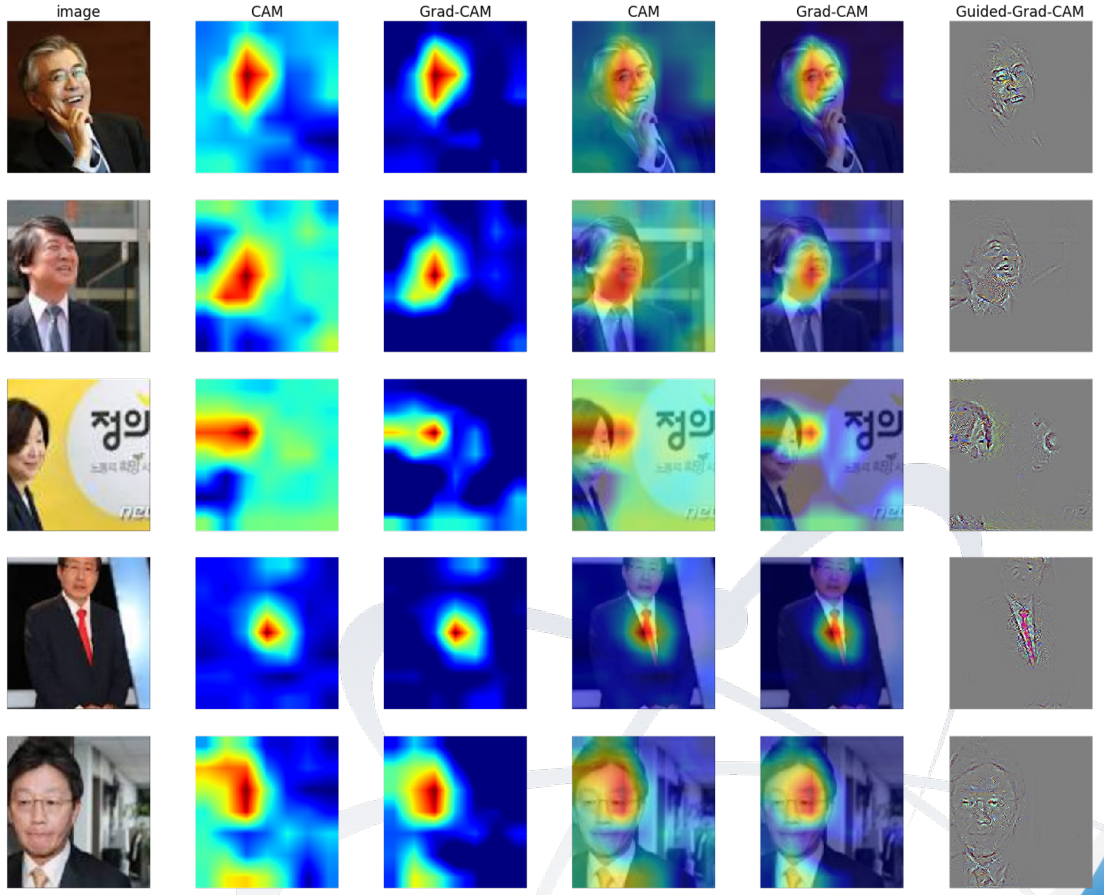
- 원전 국가 1급보안시설
- 한수원, 한국원자력연구원 등 데이터공유 어려움

Trustworthy AI

설명가능한 인공지능 (eXplainable AI, XAI)

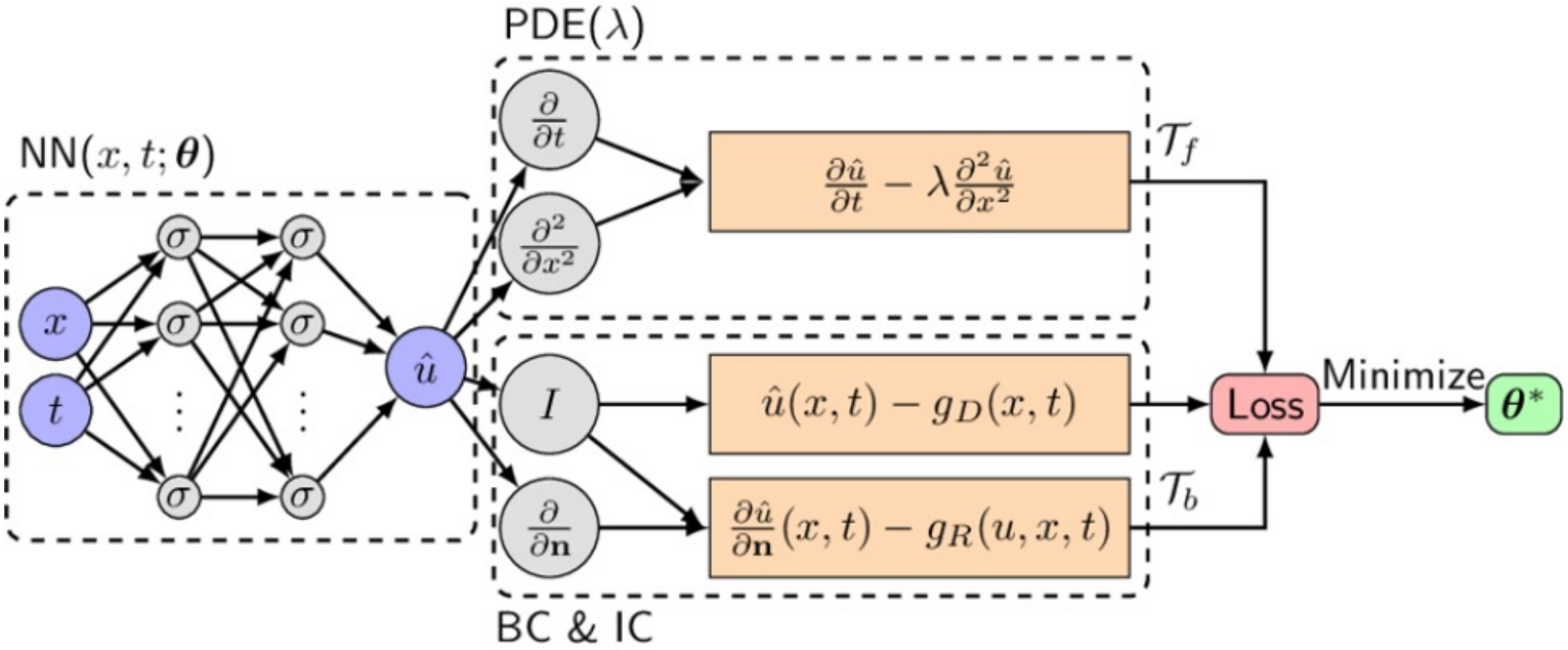


방위고등연구계획국(DARPA, Defense Advanced Research Projects Agency)



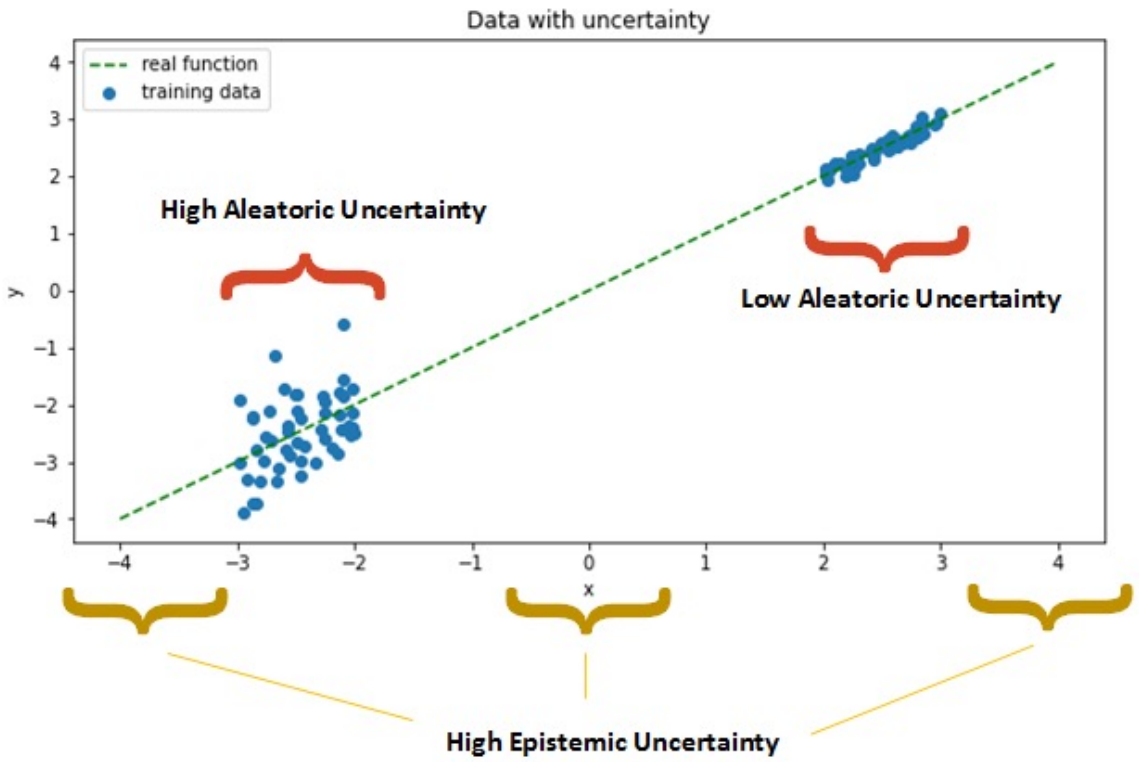
Trustworthy AI

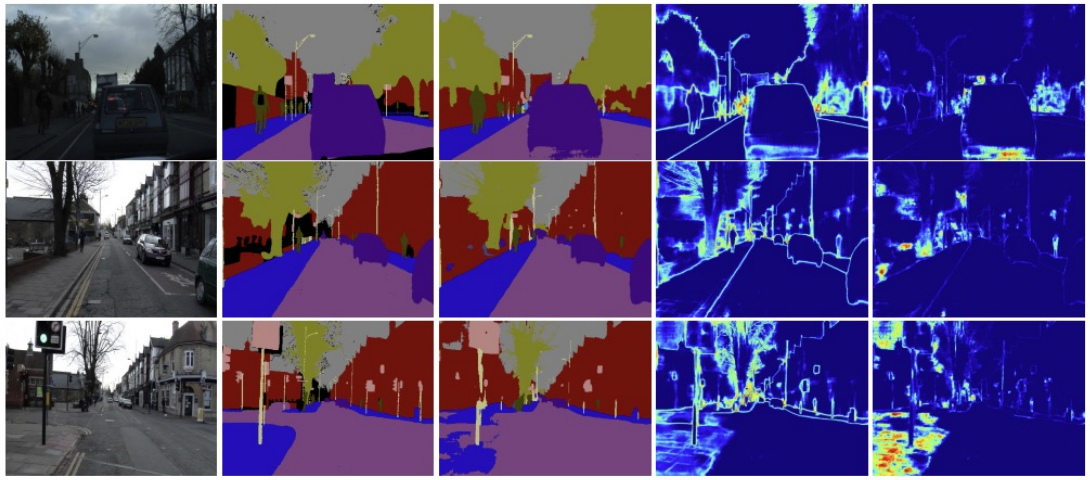
Physics Informed Neural Network(PINN)



https://www.youtube.com/watch?v=vFghJ0r39K4&list=PL8VpspSED2HynhsEbbHmFnsYd_BBUYDff

불확도 예측 (Uncertainty Quantification)



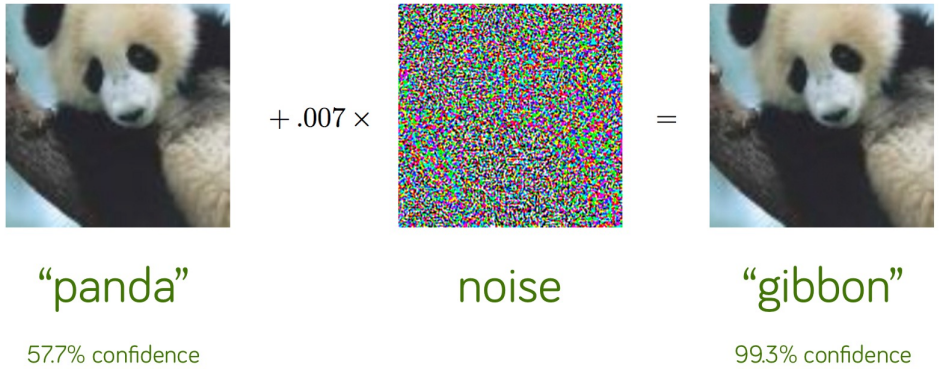


(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

What Uncertainty Do We need in Bayesian Deep Learning for Computer Vision?_ Yarin Gal

Trustworthy AI

Adversarial Attack



https://techxplore.com/news/2019-04-adversarial-patch-ai.html

부분자율시스템 관련 법률

자동차의 현재 속도		전방최소안전거리(m)
(km/h)	(m/s)	
7.2 이하	2.00 이하	2.0
10	2.78	3.1
20	5.56	6.7
30	8.33	10.8
40	11.11	15.6
50	13.89	20.8
60	16.67	26.7
70	19.44	33.1
80	22.22	40.0
90	25.00	47.5
100	27.78	55.6
110	30.56	61.1

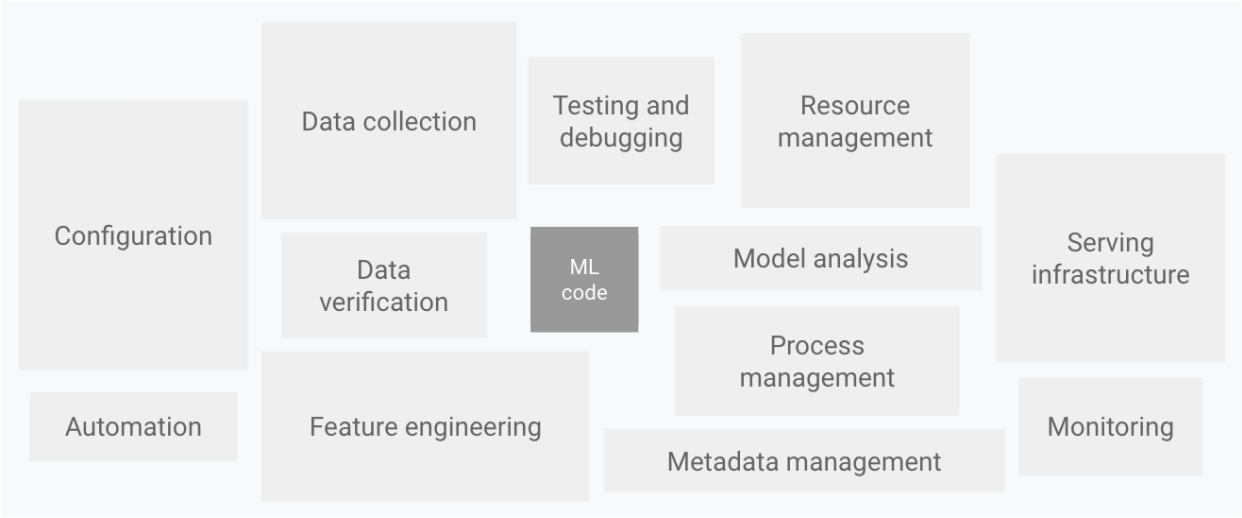
[부분 자율주행시스템의 감지범위]

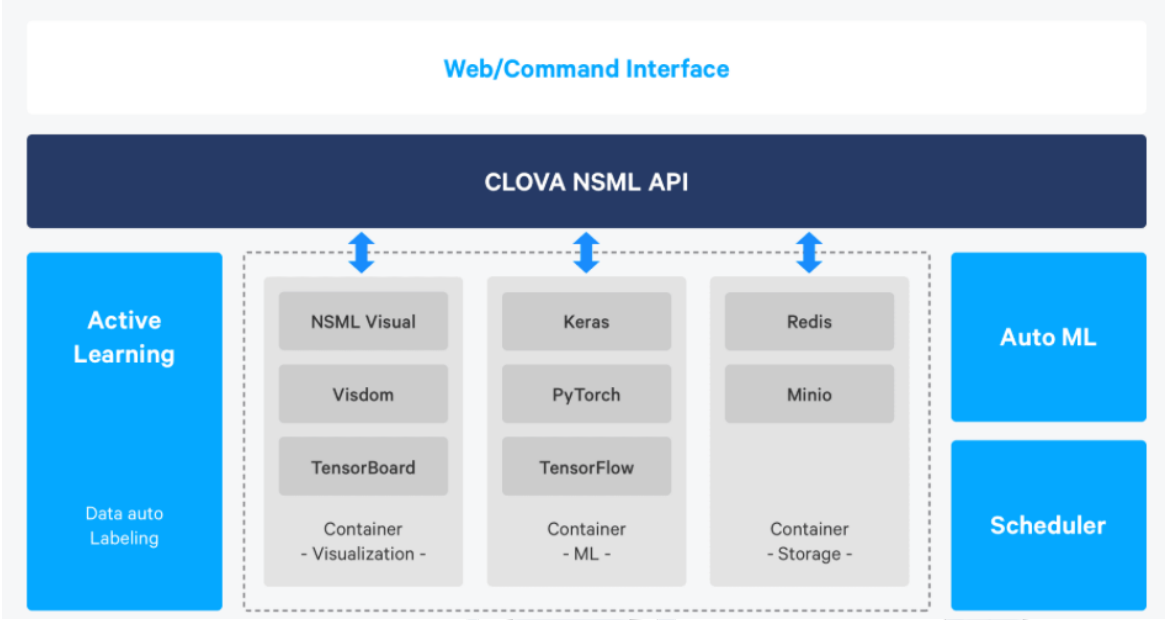
원자력의 AI 또한 명확한 계산을 통한, 구체적인 조건이 필요하다.

이를 위해 많은 데이터가 필요할 것으로 예상됨

AI Platform

MLOPS (ML + Devops)

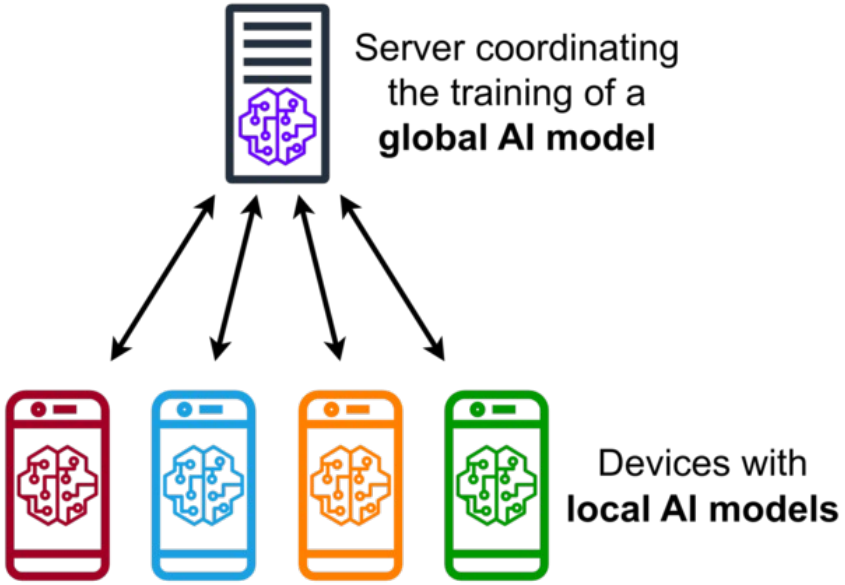


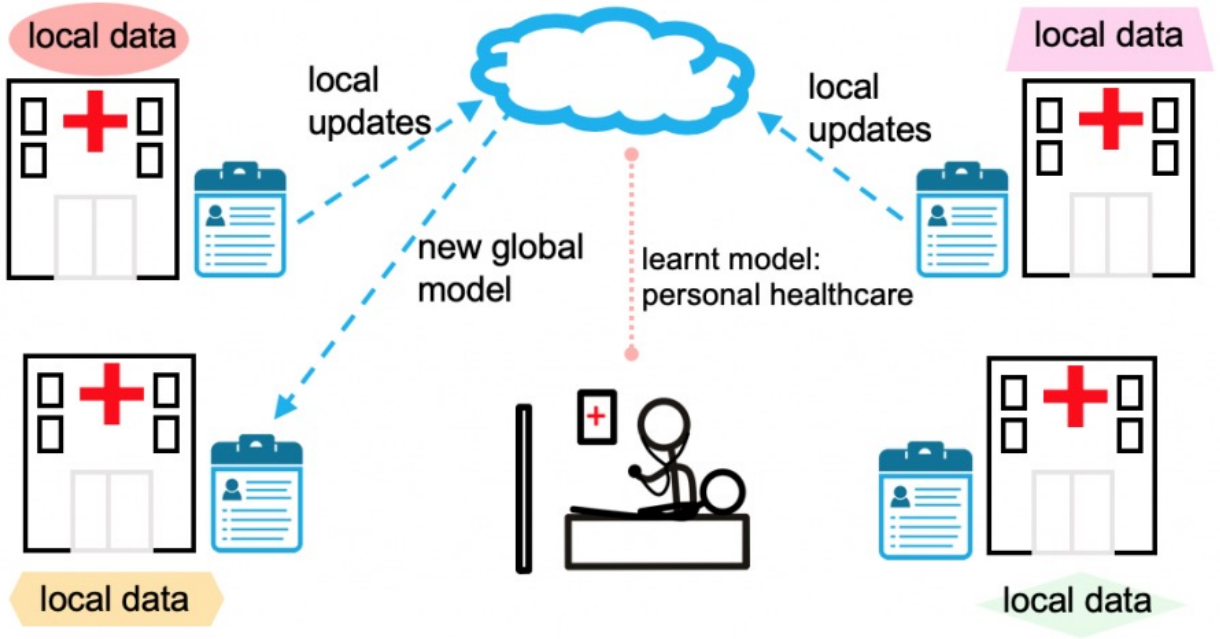


https://en.wikipedia.org/wiki/Federated_learning

05 AI Platform

Federated Learning



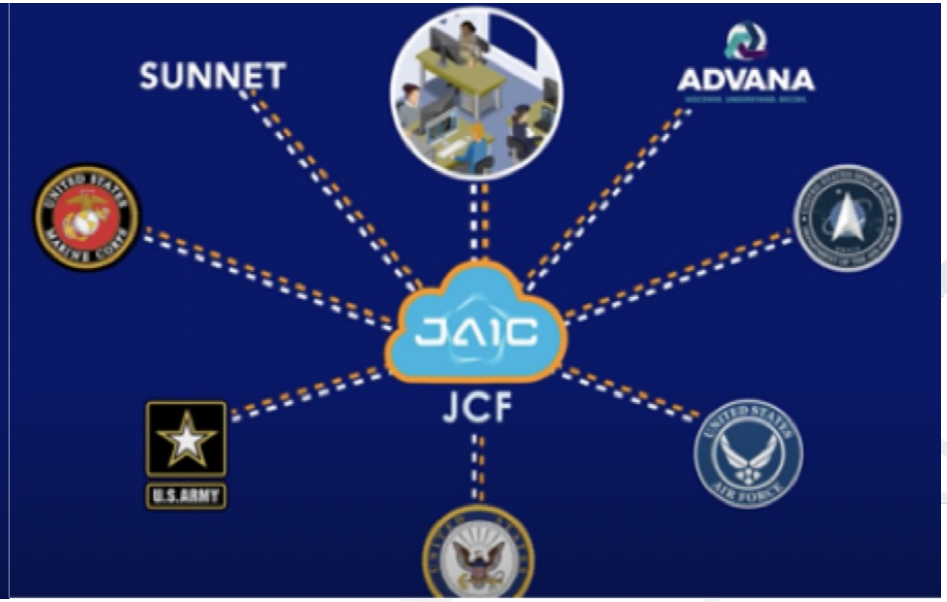


https://en.wikipedia.org/wiki/Federated_learning

https://blog.ml.cmu.edu/2019/11/12/federated-learning-challenges-methods-and-future-directions/

05 국방 AI 플랫폼 – Joint Common Foundation

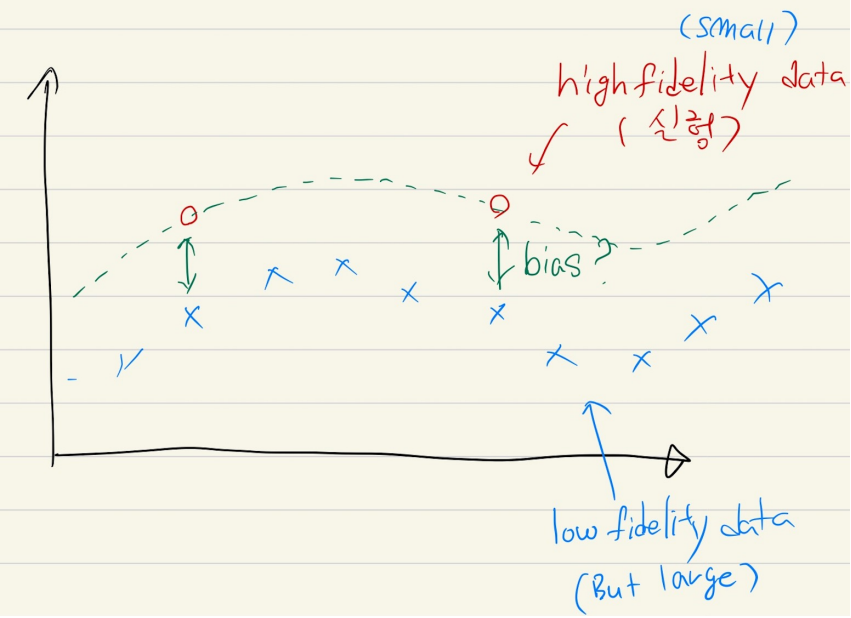
- 미 국방성이 군사 전반에 AI 애플리케이션을 개발하고 배포할 수 있도록 개발한 안전한 클라우드 기반 AI 플랫폼
- 국방부의 다양한 기관과의 협업을 개선하고 군사 전반에 걸쳐 인공지능 개발에 필요한 프로세스를 간소화하는 것을 목표로 함



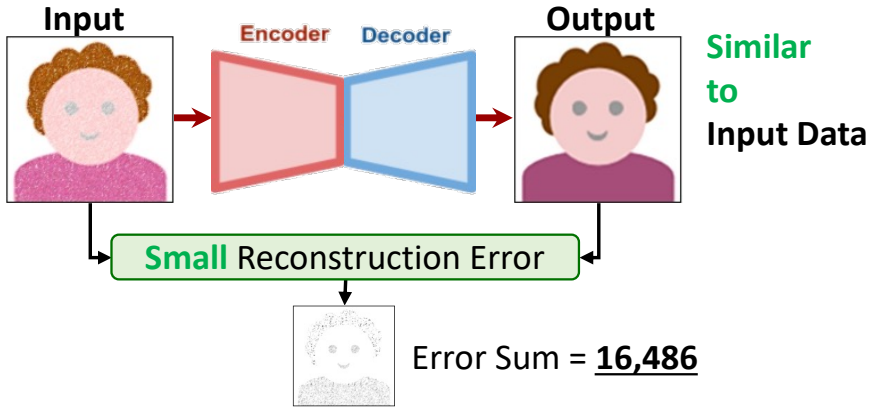
05 데이터 문제

데이터 부족 & 불균형

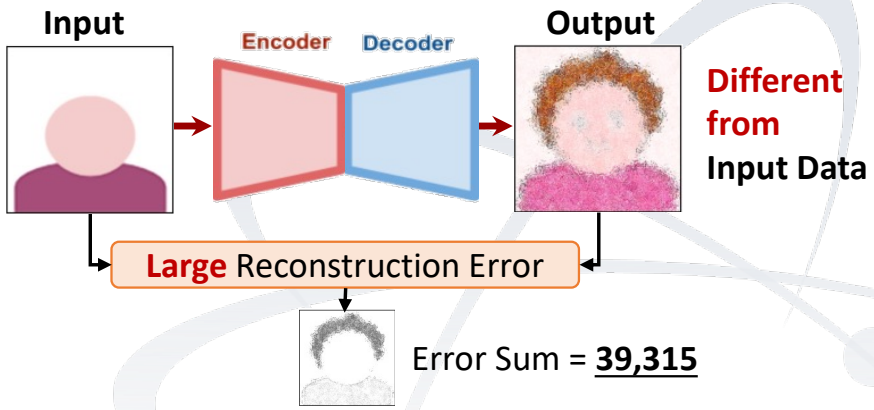
- Anomaly detection
- Synthetic Data
- Domain adaptation
- Multi-fidelity Modeling



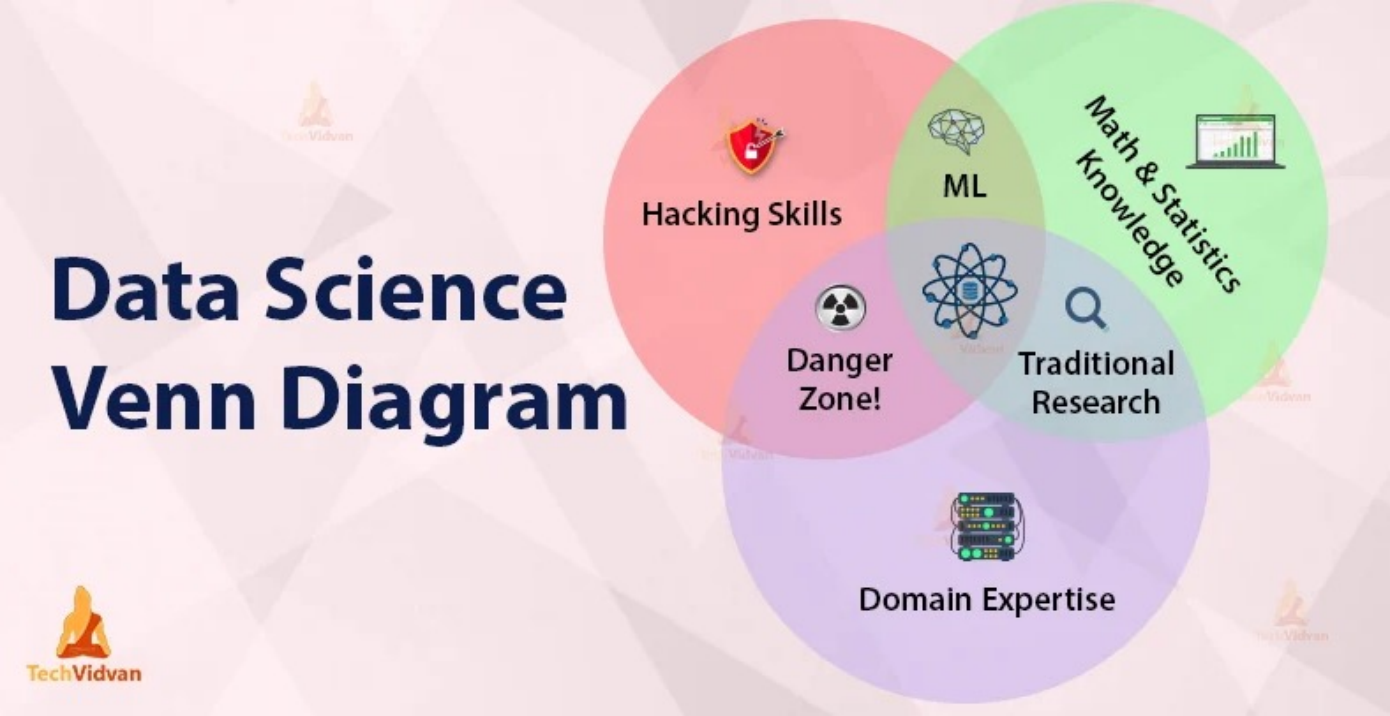
▪ 정상 데이터에 대한 모델 입출력 차이



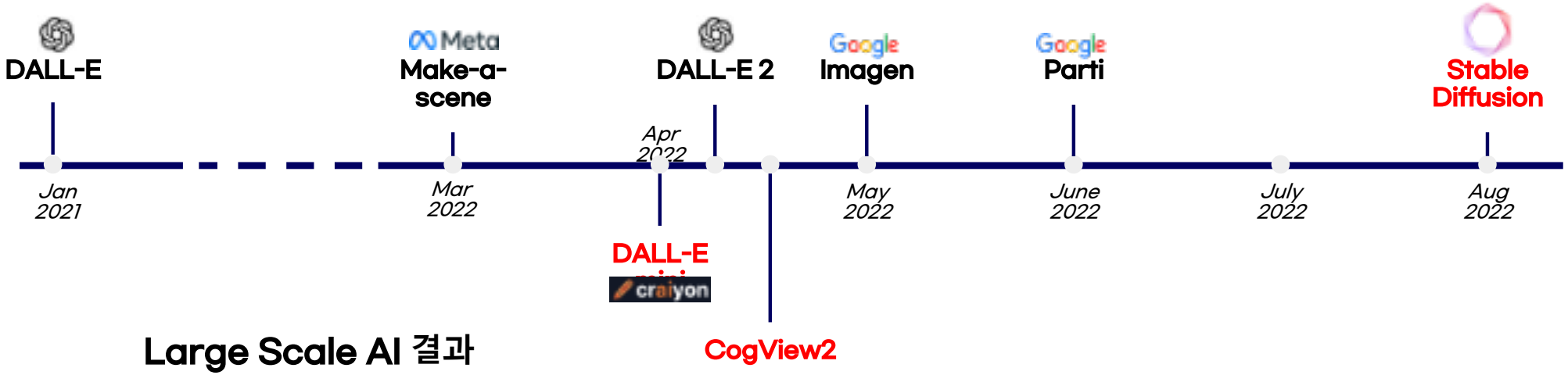
▪ 비정상 데이터에 대한 모델 입출력 차이



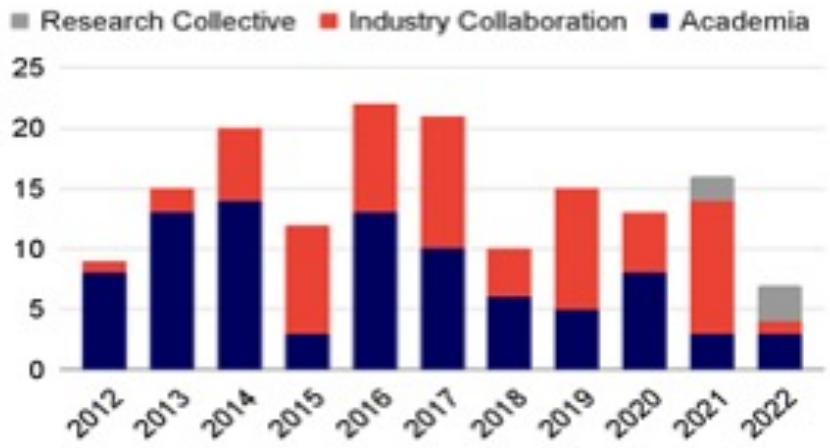
원자력 인공지능 인력


Data Science
Venn Diagram
Hacking Skills
ML
Math & Statistics
Knowledge
Danger
Zone!
Traditional
Research
Domain Expertise
TechVidvan

03 Decentralized Research



Large Scale AI 결과



출처: stateof.ai 2022

05 원자력연구원의 인공지능조직(안)

한국원자력연구원
Korea Atomic Energy Research Institute

디지털플랫폼실

- 디지털플랫폼 구축
  - 정부 디지털플랫폼 정책 대응
  - Data Lake & MLOPS 구축
- 연구&행정 자동화(RPA)
- 원자력국가전략수립을 위한 데이터 분석
  - 연구동향, 전력수요예측, 전력그리드 분석 등

인공지능응용연구실

- 원자력 핵심 AI 연구
  - XAI, UQ
- 디지털트윈 연구
  - Data Driven Simulation
- 원내 부서 AI 연구 지원
- 컨설팅 및 Pilot 프로젝트
- 내부 인공지능 교육

(KAERI-대학)
원자력디지털인력양성센터

- SMR 등 신규 원자로의 디지털전환을 위한 인력양성
- 대학과 원자력을위한 디지털(AI, DT) 인력양성 프로그램 운영
- 대학-KAERI 공동연구
- 인턴쉽 및 인력교류
- 설명자료(링크)

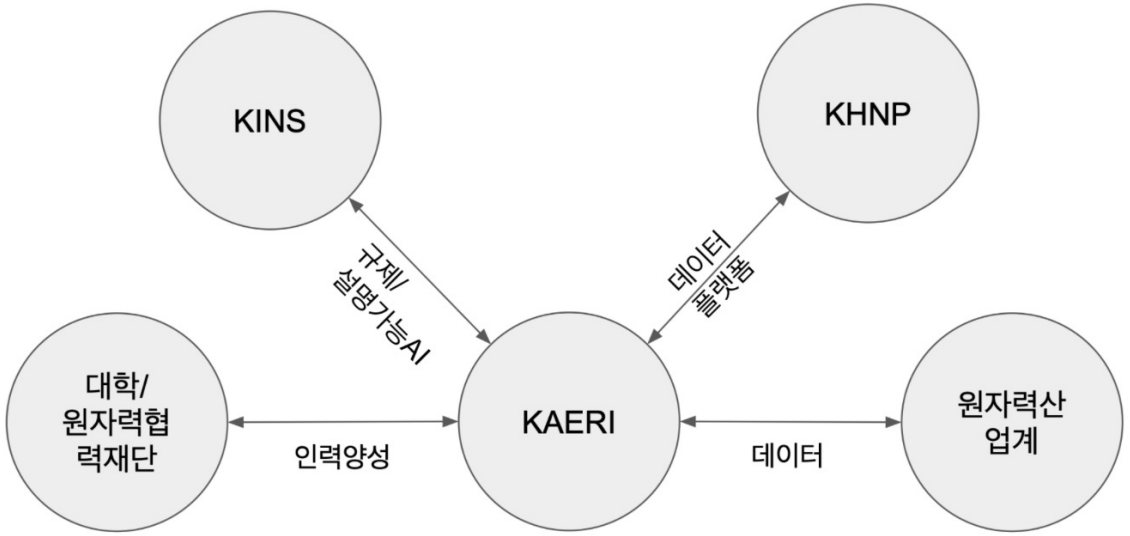
05 원자력 AI Transformation 위한 산학연 협력

원자력의 AI Transformation

- 비용, 효율성!
  - 연구, 개발 프로세스 효율화
  - 알고리즘 효율화
  - 운영 효율화

원자력인공지능 한계극복

- 개발자, 커뮤니티 문화 도입
- 산학연 협력 네트워크 강화
- 원자력인공지능 인력양성
- 설명가능한 인공지능, 신뢰성 강화



원자력 인공지능 허브