



[K-Cloud ^인공모과제 성과공유회]

# 인공지능 기반 노심 이상상황 조기진단 기술 실적용을 위한 설명가능한 인공지능기술 개발



2025. 10. 29(수)

과제책임자 : 울산과학기술원 이덕중 교수  
발표자: 울산과학기술원 실무담당자 윤동민

## 발표순서

01 연구개요

02 연구목표  
계획대비  
실적

03 연구개발 성과

04 연구 수행  
노력도

05 기술의 완성도  
및 활용계획

06 결론



# 1. 연구 개요

❖ **과제명** : 인공지능 기반 노심 이상상황 조기진단 기술  
실적용을 위한 설명가능한 인공지능기술 개발

❖ **연구목표**

- 설명가능한 인공지능(XAI) 기술을 활용한 신뢰할 수 있는 원자로 노심 이상상황 조기진단 인공지능 기술개발

❖ **연구기간** : 2022.10 ~ 2024.9 (24개월)

❖ **연구과제비** : 8.00 억원

(단위 : 억원)

구분	1차년도	2차년도	소계
울산과기원	3.37	3.63	7.00
이엔유주식회사	0.63	0.37	1.00
합계	4.00	4.00	8.00

(과제 시행계획서 최종본 기준)

❖ **PD부서** : 한수원 중앙연구원 노심해석그룹 (담당PD 문상래)

❖ **주관기관** : 울산과학기술원 연구책임자 이덕중

❖ **참여기관** : 이엔유주식회사 연구책임자 이지우

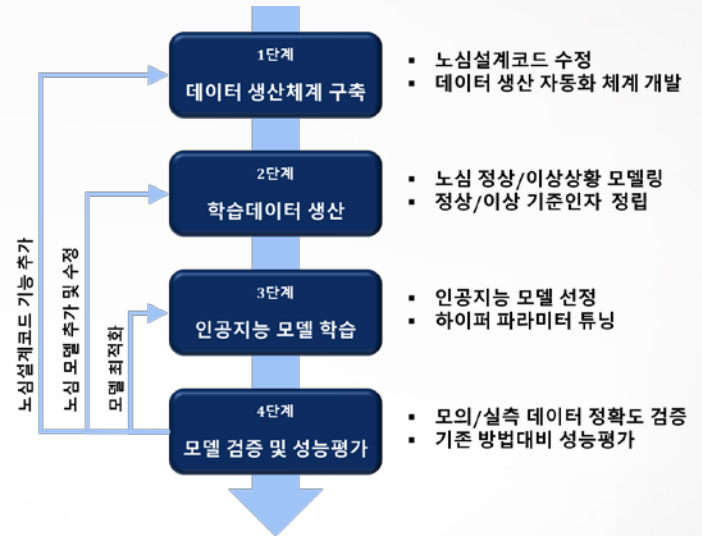
# 1. 연구 개요

## ❖ 연구배경

- 국내외 원전 노후화/장주기에 따라 AOA 등 노심 이상상황 다수 발생
- 빅데이터 시스템을 통한 노심이상진단 기술의 필요성 인식
- “인공지능 기반 노심 이상상황 조기진단 기술개발” (18K) 과제 수행

## ❖ 선행 연구과제

- 학습데이터 생산용 노심설계코드 개선
- 데이터생산 자동화 모듈 개발
- 노심 비정상 운전 모델링 방법 개발
- 이상진단용 인공지능 모델 선정
- 노심 이상진단 인공지능 모델 최적화
- 이상진단 인공지능 모델 성능평가
- 노심 이상진단 인공지능 GUI 개발



## ❖ 선행 연구과제 평가의견

- 발전소 시스템 활용 가능하게 추가 기획 연구 필요
- 실용성 향상을 위한 운전관련 예측 필요
- 한수원 연구과제에 활용 및 한수원 중장기 연구과제로 후속연구 필요
- GUI 활용한 운전원 교육 훈련 및 실제데이터 활용한 후속연구 필요

# 1. 연구내용 요약

## ❖ 설명가능한 인공지능 모델로 해결하고자 하는 노심 이상상황

1. 크러드 발생으로 인한 AOA 상황 (CIPS)
2. 인도치 않은 제어봉 위치 오정렬
3. 제어봉 위치지시기 (CEAC) 오지시
4. 노내계측기 고장으로 인한 신호 오류
5. 주기 초, 노내계측기 교차 설치로 인한 신호 오류
6. 출력 제어 중, 인도치 않은 PDIL 위반 상황
7. 냉각수 펌프 이상으로 인한 비대칭 온도를 가진 냉각수 유입

## ❖ 노심 이상상황 진단 및 실적용 방법

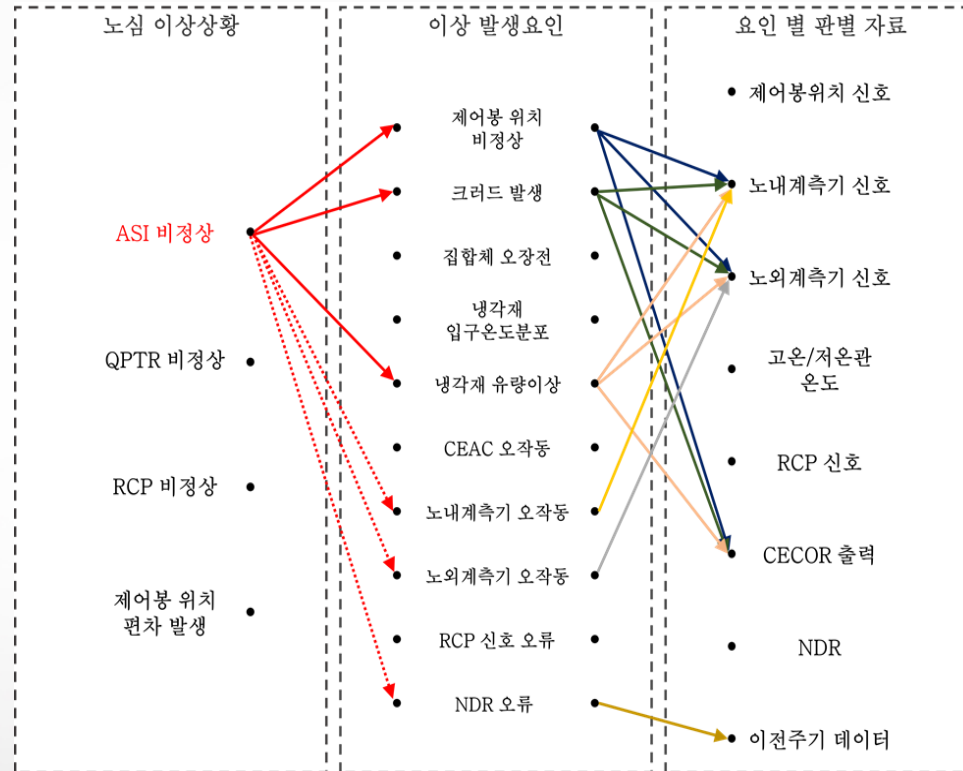
1. 실제 운전이력을 포함하여 이상상황을 모사한 RAST-K 데이터 생산
2. 노심 이상상황의 특성을 고려한 인공지능 모델 구현 및 학습
3. 학습된 모델의 GUI 연계 인터페이스 구축 (참여기관)
4. 모델의 진단 결과 가시화
5. 모델의 진단 근거 및 과정을 설명하는 XAI 가시화
6. 웹기반 통합 노심감시 체계(디지털트윈) 과제와 연동

# 1. 연구 개요

## ❖ 노심 이상상황 상관관계도

### 1. 크러드 발생으로 인한 AOA 상황 (CIPS)

- 주기 초부터 서서히 발생하는 시간 의존 데이터셋 → Time-series



AOA 발생 요인 및 판별자료 상관관계도

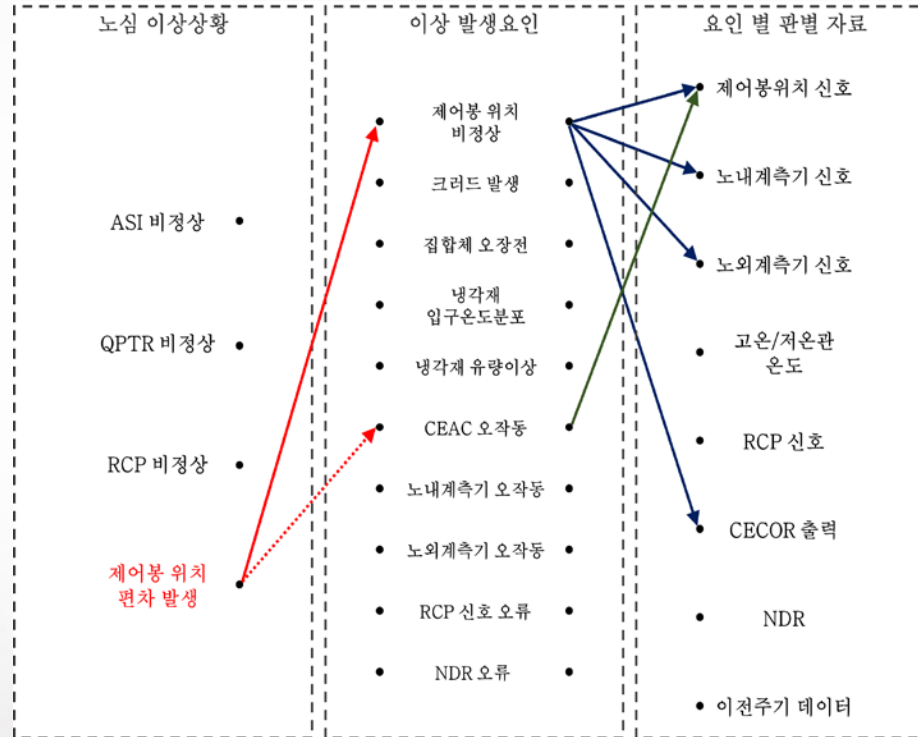
# 1. 연구 개요

## ❖ 노심 이상상황 상관관계도

### 2. 인도치 않은 제어봉 위치 오정렬

### 3. 제어봉 위치지시기 (CEAC) 오지시

– Signal failure 판별 데이터셋 → Classification



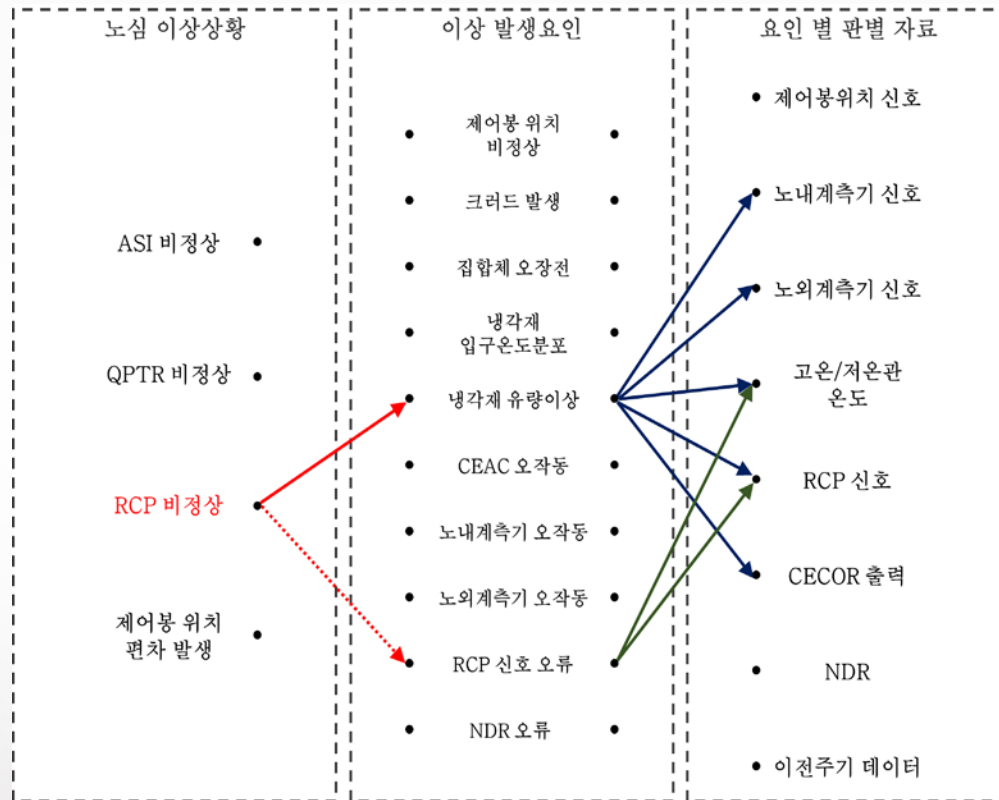
## 제어봉 위치편차 요인 및 판별자료 상관관계도

# 1. 연구 개요

## ❖ 노심 이상상황 상관관계도

### 7. 냉각수 펌프 이상으로 인한 비대칭 온도를 가진 냉각수 유입

- 냉각수 입구온도 비대칭 영향 확산 데이터셋 → Locality



냉각재 펌프 비정상 요인 및 판별자료 상관관계도

## 2. 연구계획 목표대비 실적

### ❖ 개발기술의 평가방법

1. 이상진단 AI 분류성능지표 95% 달성
2. 설명가능 AI 기술 선정의 적정성
3. 노심 이상진단 XAI 판단 근거 및 설명 타당성
4. 노심 이상진단 XAI 판단 근거자료 5개 시각화
5. 원자로 노심 상태감시 GUI 30개 시각화

## 2. 연구계획 목표대비 실적

구분	항목	단위	계획	실적	
지적 재산권	특허(출원)	건	2	2	
	프로그램(등록)	건	4	6	
논문	SCI(E)	건	3	3	
	비SCI(E)	국외게재	건	0	0
		국내게재	건	0	0
		국외발표	건	0	3
		국내발표	건	0	2
기타(S/W, 기술보고서, 시작품 등)			1	1	
합계				19	

## 2. 연구계획 목표대비 실적

성과물명	형태	완료일	비고
"원자로의 비정상 축방향 출력 편차 예측 방법 및 이를 수행 하기 위한 컴퓨팅 장치" 특허 출원	특허	2023. 12. 11.	주관기관
"인공지능을 이용한 원자로의 냉각재 입구온도의 비대칭 예측방법, 이를 수행하는 전자 장치 및 컴퓨터 프로그램 " 특허 출원	특허	2024. 09. 03.	주관기관
노심 제어봉 위치 비정상 진단 인공지능 학습데이터 생산 프로그램	프로그램	2023. 09.	주관기관
XAI 기반 원전노심 이상진단 인공지능 GUI	프로그램	2024. 08.	참여기관
CIPS 조기진단용 트랜스포머 모델	프로그램	2024. 09.	주관기관
냉각수 입구온도 예측 인공지능 학습모델	프로그램	2024. 09.	주관기관
ICI 집합체 비정상 설치 판별 인공지능 학습모델	프로그램	2024. 09.	주관기관
노내계측기 신호 비정상 판별 인공지능 학습모델	프로그램	2024. 09.	주관기관
국제 저널 논문 1건 게재	논문	2024. 05.	주관기관
국내외 학회 발표 논문 5건 게재	논문	2024. 09.	주관기관
설명가능 인공지능 분석 기술보고서	기술보고서	2023. 10.	주관기관

### 3. 연구개발 성과

#### ❖ 성과목표 ① : 원자로 노심 이상진단 AI 분류성능지표 95% 달성

- 7개 노심 이상상황 학습데이터 및 진단 모델 개발
- 성능 평가 방법 : 학습 데이터와 같은 규칙으로 생성된 독립된 검증데이터로 오차 계산 및 RMS 합

노심 이상상황 진단	회귀모델	성능 지표		오차	
제어봉 오정렬	소프트맥스 다층 퍼셉트론 모델	CR 위치 상대오차 RMS		2%	
PDIL 위반					
CIPS 조기 진단 (120일 뒤 ASI 예측)	비전 트랜스포머 모델	ASI 절대오차 RMS		0.0074*	
냉각수 입구온도 비대칭	자동회귀 디코더 모델	온도 상대오차 RMS		0.42%	
PI 신호 오류 진단	진단 모델	정확도	정밀도	Recall	F1.
노내계측기 신호 Fail	자가회귀 오토인코더 모델	99%	100%	99%	99%
노내계측기 교차 설치	자가회귀 오토인코더 모델	100%	91%	99%	95%
제어봉 위치지시기 오지시	자가회귀 오토인코더 모델	100%	100%	90%	94%

\*최종평가 당시 수준으로, 사후 연구를 통해 ASI 절대오차 0.000264까지 달성함.

### 3. 연구개발 성과

- ❖ 노심 이상상황 진단 모델 학습데이터 생산 기술
- ❖ 크러드 발생 : RAST-K 크러드 발생 모듈 활용 연소계산 모사
  - ❖ CRUD mass balance equation

$$\frac{dC_{CRUD}(t')}{dt'} = \frac{1}{M_{RCS}} \left( S_{CRUD}^{RCS} + S_{CRUD}^{reload}(t') - \sum_{i=1}^I \frac{dM_{CRUD,i}(t')}{dt'} - \epsilon \dot{M}_{letdown} C_{CRUD}(t') \right)$$

1. 제어봉 오정렬 : 임의 시점, 임의 제어봉 위치 오정렬하여 연소 계산
  2. 냉각수 입구온도 비대칭 : 임의 비대칭 온도 분포로, 임의 시점부터 적용하여 연소 계산
  3. 장기 저출력 모사 : 임의 기간 (총 10일간), 출력 감발하여 연소 계산
    - ❖ 제어봉 삽입 및 보론 주입 방법을 임의 비율로 설정하여 출력 감발
- ❖ 상기 1~4번의 방법을 조합하여 다양한 데이터셋 구성
    - ❖ 계산된 데이터를 재가공하여 ICI 고장 등 True-Negative\* 데이터 생성

True-Negative\* : 실제 노심은 정상운전상태이나, 신호체계의 고장으로 비정상신호가 발생한 경우

# 3. 연구개발 성과

## ❖ 노심 이상상황 진단 모델 학습데이터 생산 목록

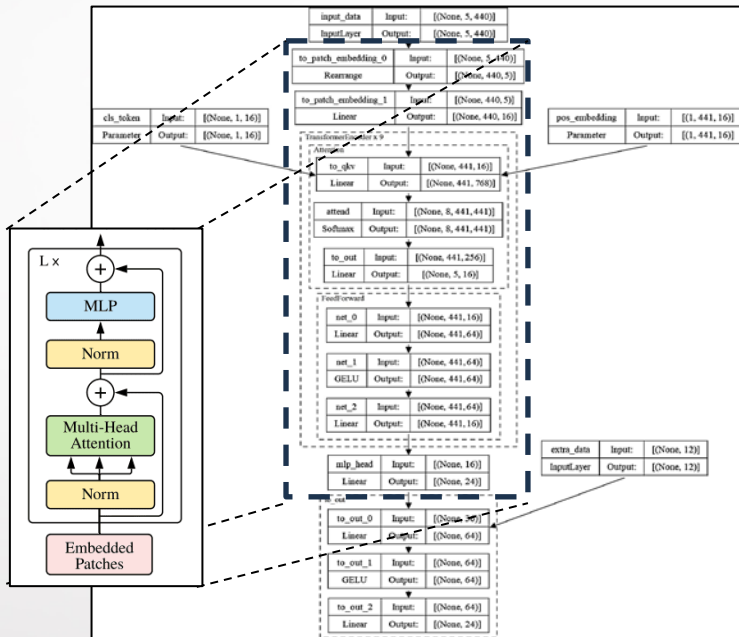
- 다양한 운전이력과 **노심 이상상황 모사** 경우의 수를 포함한 빅데이터 구축
- **132만개** 시나리오 경우의 수, 각 52개의 시계열 연소 스텝 보유 데이터 생산

데이터셋	모사 적용						데이터 수
	크리드 침적	제어봉 오정렬	냉각수 온도 비대칭	냉각수 유량 비대칭	장기 저출력 경험	N주기	
A	X	O	X	X	X	5	80000
B	X	X	O	X	X	5	80000
C	X	X	O	O	X	5	80000
D	O	X	X	X	X	5	129000
E	O	X	O	O	X	5	80000
F	O	O	X	X	X	5	80000
G	O	O	O	O	80~100%	5	64000
1	X	X	O	X	80%	5	2252
2	X	X	O	X	80%	5	7212
3	X	X	O	X	80%	5	7208
4	O	X	O	X	80%	5	4477
5	O	O	O	X	80%	5	4488
6	X	O	O	X	80%	5	7548
7	X	O	O	X	80%	5	5338
8	X	O	O	X	80%	5	5363
9	X	O	O	X	X	5	703
10	X	O	O	X	70%	5	8000
11	X	O	O	X	X	5	8000
12	X	O	O	X	X	5	8000
13	X	O	O	X	70%	5	7955

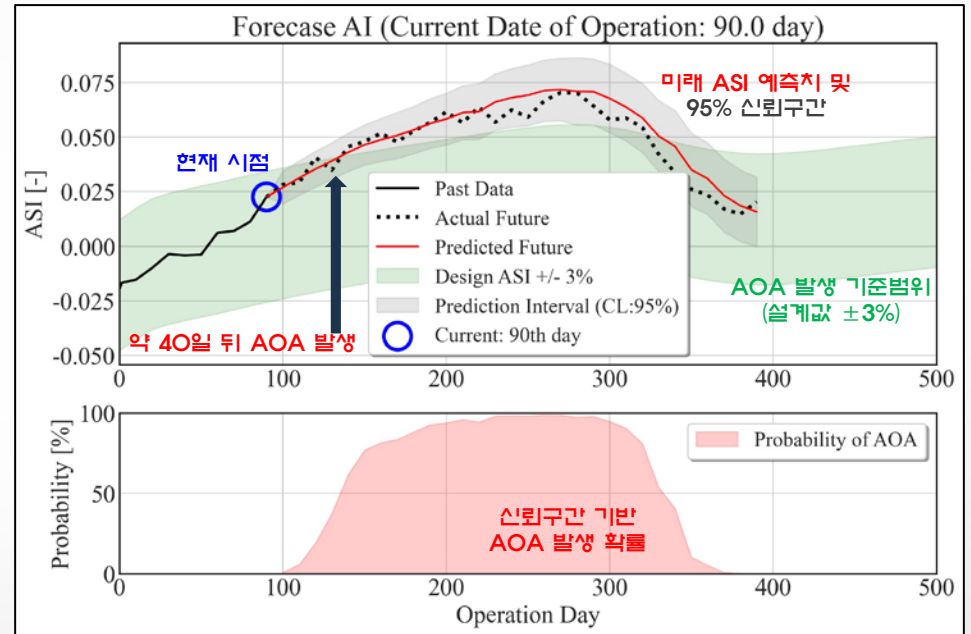
데이터셋	모사 적용						데이터 수
	크리드 침적	제어봉 오정렬	냉각수 온도 비대칭	냉각수 유량 비대칭	장기 저출력 경험	N주기	
14	X	O	O	X	X	4	7772
15	O	O	O	X	X	4	7421
16	O	X	O	X	50~100%	5	8400
17	O	O	O	X	50~100%	5	8400
18	X	O	O	X	X	5	8398
19	O	O	O	O	50~100%	5	8391
20	O	X	O	O	50~100%	5	8397
21	X	X	O	O	90~100%	5	8400
22	O	X	X	O	50~100%	5	8392
23	O	O	O	O	50~100%	5	48000
24	O	O	X	X	50~100%	5	148600
25	O	X	O	X	50~100%	5	160000
27	X	X	X	O	X	5	66472
28	O	X	X	O	X	5	37048
29	X	O	X	X	50~100%	5	61600
30	X	X	X	X	50~100%	5	58400
<b>총 시나리오 수</b>							<b>1,323,635</b>
<b>총 데이터 수</b>							<b>68,829,020</b>

# 3. 연구개발 성과

- ❖ CIPS 조기 진단 (미래 ASI값 예측)모델 개발
- ❖ 학습모델 : 비전 트랜스포머 모델
  - ❖ 데이터의 시계열 패치(patch)를 Self-Attention 메커니즘으로 상관관계를 학습
  - ❖ 크러드 침착이 포함된 과거 50일간의 축적 데이터를 바탕으로 미래 ASI 예측 모델
- ❖ 예측 성능 평가: ASI 예측치 절대오차 **0.0074** (RMS)



CIPS 조기진단 모델 구조



CIPS 발생 시 ASI 예측 결과 (HFP 운전 시)

# 3. 연구개발 성과

## ❖ CIPS 조기 진단 (미래 ASI값 예측)모델 개발

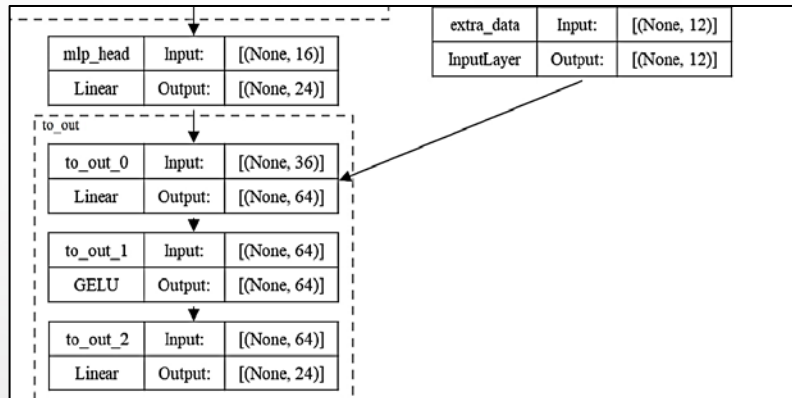
### ❖ 학습모델 : 비전 트랜스포머 모델

❖ 장기 저출력 운전 경험 노심에서의 CIPS 예측

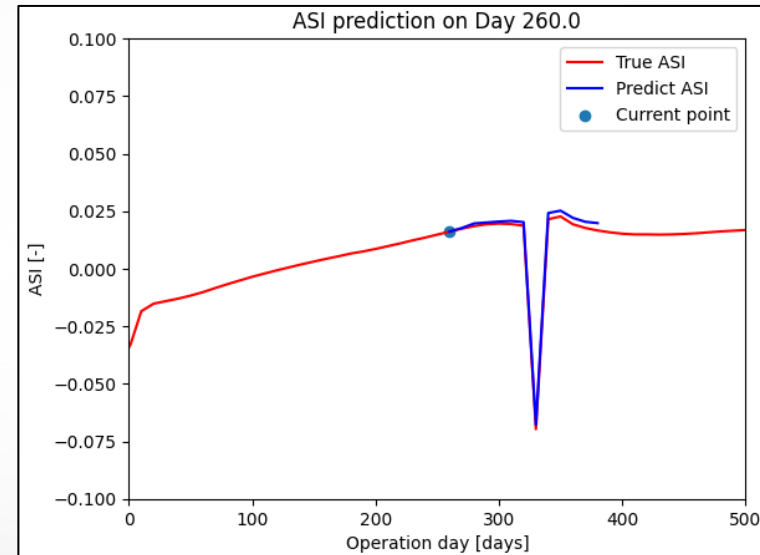
❖ 제어봉 삽입 및 보론 주입으로 출력 감발 시, 그로인한 인도된 ASI 변화 발생

(\*CIPS로 인한 ASI 변화와 구분 필요)

❖ 모델 출력중에 장기 저출력 운전 계획 데이터를 추가 입력하여 출력 감발로 인하여 ASI 값이 변화한 것임을 모델이 학습하도록 함.



출력중에 미래 장기저출력 계획 데이터를 추가 (extra\_data)



CIPS 발생 시 ASI 예측 결과  
(EFPD 320~330에서 장기저출력 운전 계획)

# 3. 연구개발 성과

❖ 냉각수 입구온도 비대칭 진단 모델 개발

❖ 학습모델 : 자동회귀 디코더 모델

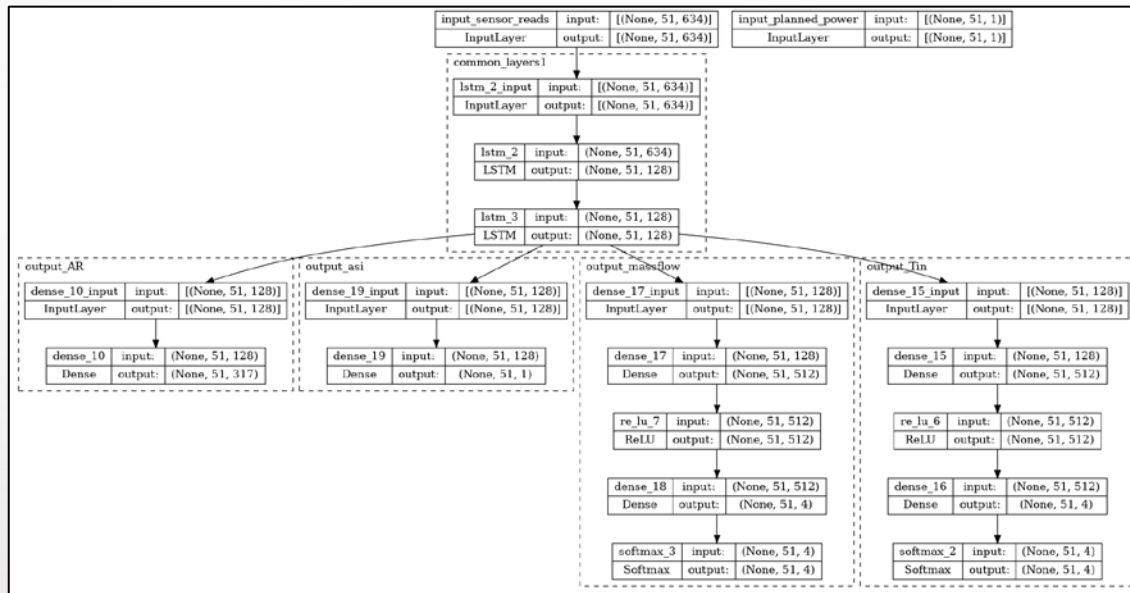
❖ 시퀀스의 이전 입력에서의 측정값으로 다음 성분을 예측하는 디코더 모델

❖ 자동회귀(DR) 정의 : 노내계측기 신호 이전스텝에서의 변화량

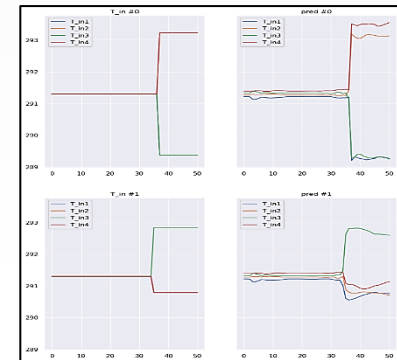
$$DR = x_{SPND}(t) - x_{SPND}(t - 1) = \Delta x$$

❖ AR 값을 예측하는 training으로 1단계 학습, 이후 2단계 학습에서 출력값 학습

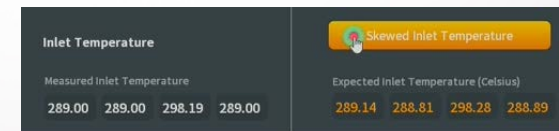
❖ 예측 성능 : 온도 예측값 상대오차 **0.42% (RMS)**



냉각수 입구온도 비대칭 진단 모델 구조 (2단계 학습)



냉각수 입구온도 비대칭 진단 결과  
(좌: 실제값, 우: 모델 예측값)



냉각수 입구온도 비대칭 진단 결과 GUI

### 3. 연구개발 성과

#### ❖ PI 신호 비정상 진단 모델 (노내계측기 신호, 제어봉 위치지시기 신호)

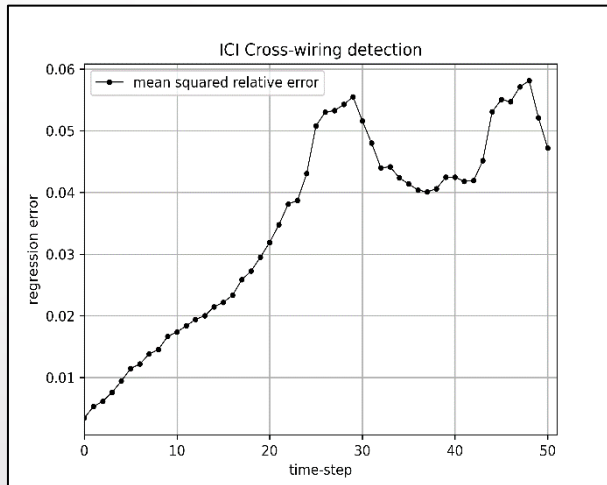
→실제 제어봉 정상 유무	제어봉 위치 정상 (True)	제어봉 위치 비정상 (False)	
↓PI 데이터 정상 유무	제어봉위치 신호 정상 (Positive)	정상운전 (True Positive)	제어봉 오정렬 상황 (False Positive)
제어봉 위치 신호 비정상 (Negative)	제어봉 위치지시기 오지시 상황 (True Negative)	PI 데이터를 통한 제어봉 위치 진단 (False Negative)	

제어봉 위치지시기 오지시 (True-Negative) 진단과  
제어봉 오정렬 (False-Positive) 진단 구분

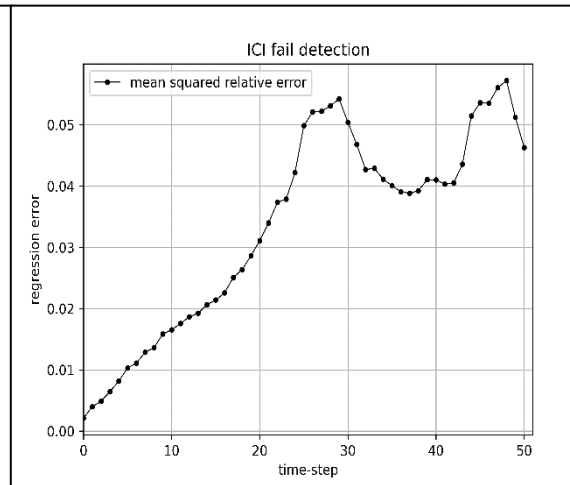
- ❖ 실제 노심은 정상운전상태이나, 신호체계의 고장으로 비정상신호가 발생한 경우
- ❖ 학습모델 : 자가회귀 오토-인코더 모델
  - ❖ 비정상 및 정상 데이터를 입력 받아 순수한 정상 데이터를 복구하는 모델
  - ❖ 인코딩 과정에서 신호 오류의 주요 특성이 잠재 공간에 희석되어 디코딩 과정에서 정상 데이터로 복원
  - ❖ 이상신호 진단 : 입력 데이터와 정상으로 복원된 출력 데이터 간 오차 발생 시, 입력 데이터에 비정상 데이터가 포함되었음을 진단

# 3. 연구개발 성과

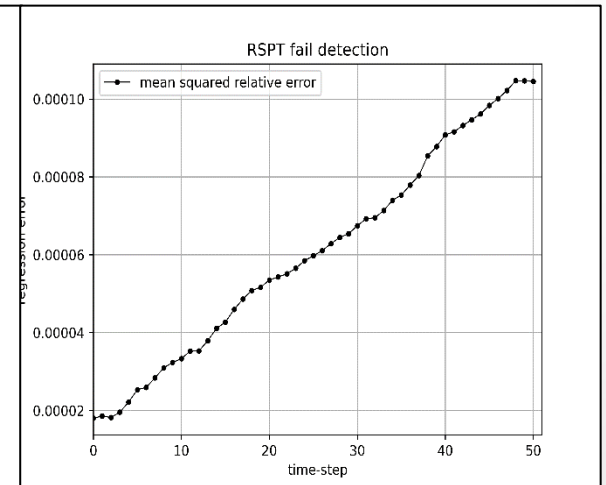
- ❖ PI 신호 비정상 진단 모델 (노내계측기 신호, 제어봉 위치지시기 신호)
- ❖ 학습모델 : 자기회귀 오토-인코더 모델
  - ❖ 자기차분(DR) 정의 : 입력 신호 이전스텝에서의 2차 변환량
 
$$DR = \Delta x_{SPND}(t) - \Delta x_{SPND}(t - 1)$$
  - ❖ 3단계 Staged learning 적용
    - ❖ 1단계 : DR값 학습
    - ❖ 2단계 : 정상 노내계측기 신호값 복원 학습
    - ❖ 3단계 : 비정상신호의 정상복원 학습
- ❖ 복원 모델 성능 평가 : MSE (최대 0.06%)



노내계측기 신호교차 복원 모델  
MSE vs time-step



노내계측기 신호고장 복원 모델  
MSE vs time-step



제어봉 위치지시기 오지시 복원 모델  
MSE vs time-step

### 3. 연구개발 성과

- ❖ PI 신호 비정상 진단 모델 (노내계측기 신호, 제어봉 위치지시기 신호)
- ❖ 진단 모델 성능 평가 : Confusion Matrix (**최소 Recall 93%**)
  - ❖ Accuracy =  $(TP + TN) / (TP + TN + FP + FN)$
  - ❖ Precision =  $(TP) / (TP + FP)$
  - ❖ Recall =  $(TP) / (TP + FN)$
  - ❖ F1 score =  $2 \times \frac{Precision \times Recall}{Precision + Recall}$
- ❖ ICI 고장의 경우, FAIL 상황 및 SUSPECT 상황을 구분하여 진단
- ❖ 신호 비정상 진단의 경우, Recall 및 F1 score가 중요한 지표

모델	ICI 교차			ICI 고장				CEAC 오지시		
	BOC	MOC	EOC	BOC (FAIL)	BOC (SUSPECT)	MOC (FAIL)	EOC (FAIL)	BOC	MOC	EOC
시점										
Accuracy	100%	99%	98%	100%	100%	100%	100%	100%	100%	100%
Precision	99%	84%	84%	100%	96%	100%	100%	100%	100%	100%
Recall	99%	98%	93%	100%	99%	100%	100%	96%	95%	95%
F1 score	99%	90%	88%	100%	98%	100%	100%	98%	97%	97%

# 3. 연구개발 성과

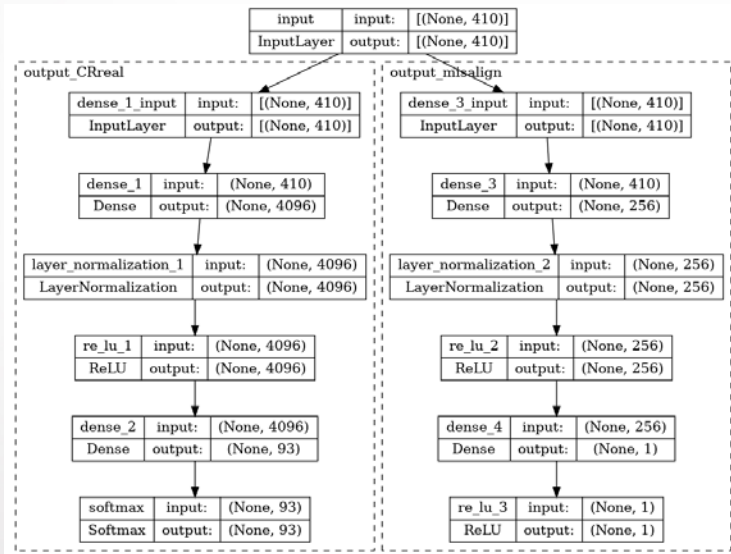
## ❖ 제어봉 오정렬 진단 모델 개발

## ❖ 학습모델 : 소프트맥스 다층 퍼셉트론 모델

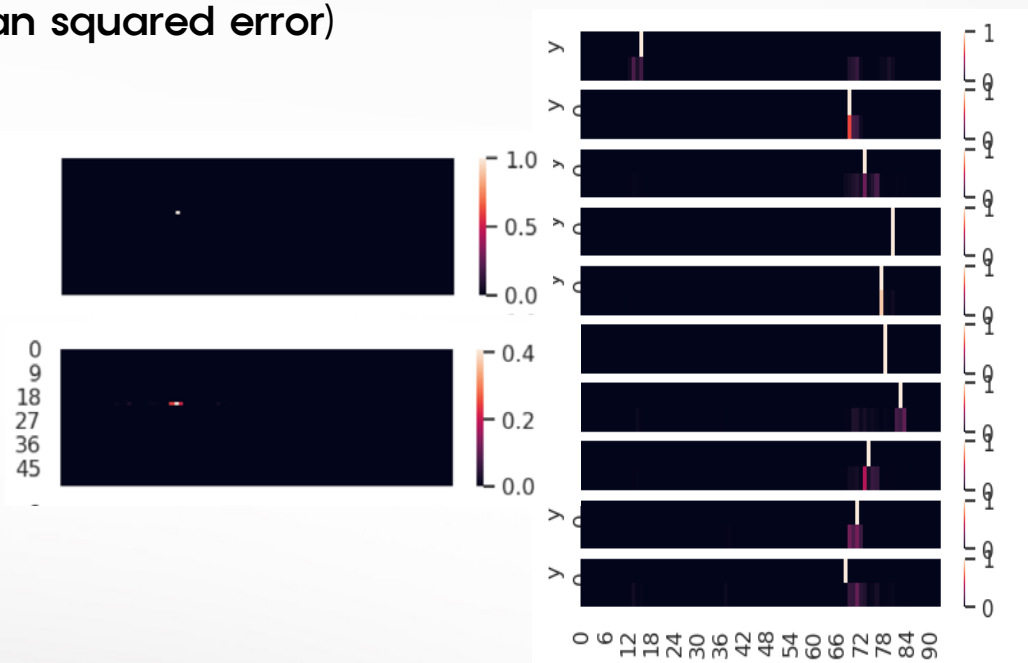
- ❖ 출력층의 활성화 함수를 소프트맥스로 정의한 다층 퍼셉트론 (MLP) 모델.
- ❖ 제어봉 위치 예측 모델로서 PDIL 위반 진단 동시 수행

## ❖ 예측 성능 :

- ❖ 발생시점 및 인덱스 예측 모델 : **0.054%** (Binary Cross-entropy Error)
- ❖ 오정렬 편차 예측 모델 : **0.15%** (Mean squared error)



제어봉 위치 예측 모델 구조



제어봉 오정렬 발생 시점 및 인덱스 예측

제어봉 오정렬 편차 예측

# 3. 연구개발 성과

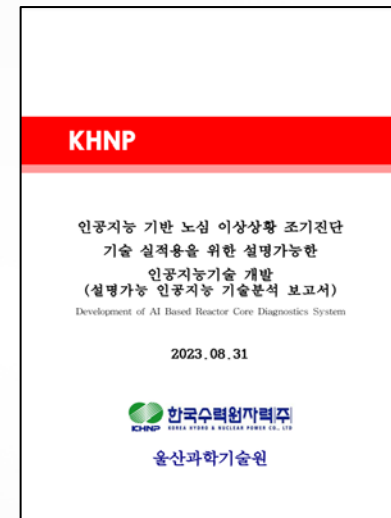
## ❖ 성과목표 ② : 설명가능 인공지능 기술 검토 보고서 작성 및 설명가능 AI 기술 선정의 적정성 검증

- MDI를 통해 학습 과정에서 트리 기반 기계학습 모델이 중점적으로 학습한 특성을 파악할 수 있다.
- LIME은 샘플의 특성을 임의로 변형하여 모델 성능의 변화를 관측하고, 이를 이용해 단순한 대리 모델을 학습하여 복잡한 원형 모델을 설명한다.
- SHAP는 특성이 예측에 기여한 정도를 측정하기 위해 모든 가능한 특성 집합(Subset)에 대한 예측값의 변화를 평가하므로, 특성 간 상호작용 포착이 가능하다.

표 8 노심 이상 상황에서의 계측기 신호 편차 상위 15개와 XAI가 평가한 중요 특성 15개의 비교

비정상 상황	비교 샘플	MDI	PI	LIME	SHAP
제어봉 비정상	6스텝 삽입	14	14	6	9
	34스텝 삽입			5	12
입구온도비대칭	온도 편차 1.0°C	10	10	2	6
	온도 편차 2.7°C			9	10
크리드발생	ASI편차 1.5%	10	10	8	10
	ASI편차 2.9%			9	10
ICI 교차	ICI#3-ICI#8	15	15	6	10
	ICI#3-ICI#7			6	10
일치하는 특성 수		49	49	51	77
비교 기준 특성 수		60	60	110	110

XAI 기술 선정의 적정성 검증  
MDI, PI 평가 결과 타당성 일치특성 비율  
**82% (49/60) 이상 달성**

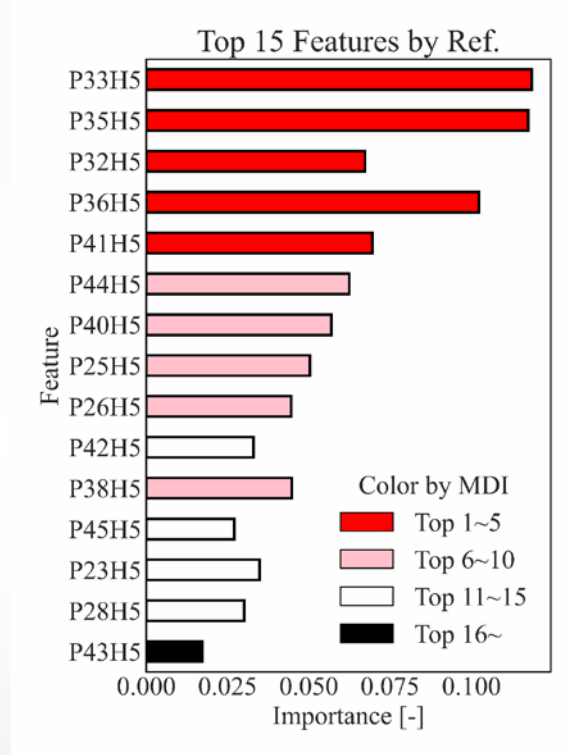
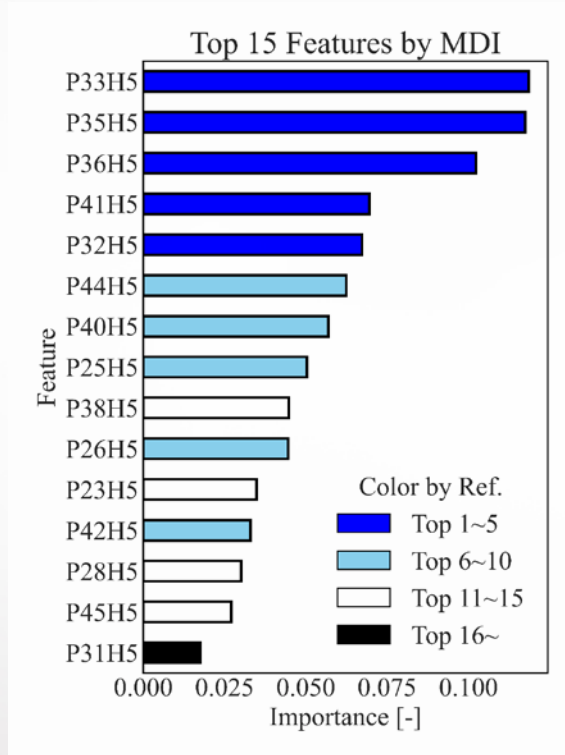


성과물 : 설명가능 인공지능  
분석 기술보고서

# 3. 연구개발 성과

## ❖ 노심 이상진단 모델에 대한 설명가능 인공지능 검토 요약

- MDI: 트리 기반 기계학습 모델이 중점적으로 학습한 특성 및 구조 파악
  - 분할 과정의 무작위 선정 빈도 MDI에 민감한 영향을 보임
  - 트리 기반 모델에만 사용할 수 있는 한계



### 3. 연구개발 성과

#### ❖ 노심 이상진단 모델에 대한 설명가능 인공지능 검토 요약

- **SHAP** : 머신 러닝 모델의 예측 결과를 해석하기 위한 방법 중 하나로, 개별 특성이 예측 결과에 얼마나 기여하는지를 측정.
  - 이론적으로 모든 AI 모델에 적용 가능하지만, 아래 SHAP 값 계산의 복잡성 때문에 사실상 **트리 기반 모델에만 사용할 수 있음.**

$$\text{Shapely value} : \phi_i = \sum_{S \in F \setminus i} \frac{|F| - |S| - 1}{|F|!} [f_{S \cup i}(x_{S \cup i}) - f_S(x_S)]$$

- **LIME** : 복잡한 AI 모델을 분석하여 간단한 대리 모델을 생성하여 모델 설명
  - 높은 상관관계를 가지는 데이터에서의 적용에 한계가 있다.
  - LIME은 값 **급격하게 변하는 상황**(ex, 노내계측기 교차) 한계

$$\text{LIME explanation}(x) = \operatorname{argmin}_{g \in G} (\mathcal{L}(f, g, \pi_x) + \Omega(g))$$

### 3. 연구개발 성과

#### ❖ 노심 이상진단 모델에 대한 설명가능 인공지능 검토 요약

- **PI** : 모델의 입력데이터 중 **무작위 인덱스의 데이터를 “생략”** 하여 입력 시 모델 정확도의 영향 평가
  - 단일 특성 값의 변화 예측에 영향 성능 변화 작음 (상관관계의 다른 특성)
  - 문제의 특성에 따른 적합한 성능 지표를 선택 필요

*PI* 지표 :  $f_{PI,i} = Perform.(X_{base}) - Perform.(X_{shuffled})$

- **LRP** : 각 레이어에서 발생한 **출력의 기여**를 역전파(decomposition) 하여, **입력 데이터의 중요도를 추적**. 입력 레이어까지의 기여도를 계산하면, 입력 데이터에 대한 중요도를 계산할 수 있다.
  - 이미지 기반 데이터와 딥러닝 모델에 적합하며, 인공지능 판별 결과의 입력 기여도와 은닉 레이어의 기여도를 이미지 형태로 시각화할 수 있다.
  - 시계열 의존성 데이터 사용에는 한계가 존재하여, 스냅샷 데이터를 추출하여 적용이 필요하다.

$$J_{th} \text{ layer relevance score} : R_j = \sum_k \frac{a_j \rho(w_{jk})}{\sum_{0,j} a_j \rho(w_{jk})}$$

### 3. 연구개발 성과

#### ❖ 노심 이상진단 모델에 대한 설명가능 인공지능 비교

- XAI 기술 비교군 : MDI, PI, LIME, SHAP
- 노심 이상상황: XAI 기술 비교용 단순화 학습데이터

이상상황	샘플링 인자	비정상 레이블링 기준
제어봉비정상	CEA#6 편차발생	CEA#6 편차발생 5스텝 이상
냉각재 불균형	사분출력 편차	온도 편차 0.6°C 이상
크러드 발생	크러드 생성속도	60일 이내 ASI 차이 3%
ICI 교차	ICI3과 교차 ICI 선택	ICI 교차 여부

이상상황	편차 발생 위치	신호 최대 편차	상위 특성 간 편차	데이터 변화
제어봉비정상	제어봉 주변, 노심 상부	~16.87 %	3.88%	점진적
냉각재 불균형	편차 발생 사분면 (전역)	~1.36	0.207%	점진적
크러드 발생	노심 상부, 외곽	~8.90%	0.300%	점진적
ICI 교차	교차발생 계측기	~32.91%	8.181%	급격

### 3. 연구개발 성과

#### ❖ 노심 이상진단 모델에 대한 설명가능 인공지능 성능 비교

- 비교기준: 비정상상태-정상상태 특성차이( $X_{비정상} - X_{정상}$ ) 상위 특성 선정
- 지역해석: 무작위 선정 100개 Positive 샘플에 대해 각 상위 15개 특성 비교

이상상황	MDI	PI	LIME	SHAP
제어봉 비정상	14	14	13.31	13.26
입구 온도 불균형	10	10	2.64	8.30
크러드 발생	10	10	8.72	10.29
ICI 교차	15	15	5.54	9.99
일치하는 특성 수	49 (81%)	49 (81%)	30.21 (60%)	41.84 (83%)
비교 기준 특성 수	60		50	

- MDI, PI, SHAP 방법의 “설명가능성” 80% 이상

# 3. 연구개발 성과

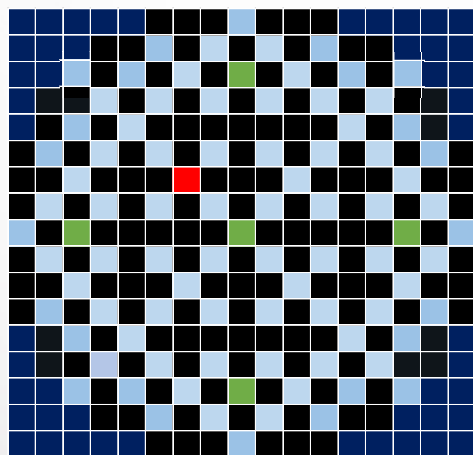
## ❖ 성과목표 ③ : XAI 이상진단 시 판단 근거 타당성

### ❖ 제어봉 오정렬 모델의 계층별 타당성 전파(LRP)

- 입력 데이터와 출력 데이터 간 관련성을 제시하여 모델 예측 결과에 대한 eXplain
  - LRP 적용 예시 시나리오
  - 출력 제어를 위한 R5 bank 100cm 삽입 도중, 운전원의 실수로 R3 bank인 79번 제어봉이 함께 삽입됨
  - 제어봉 오정렬 진단 모델의 입력 타당성 분석 및 가시화

$$\begin{aligned} \forall_k : z_k &= \epsilon + \sum_{0,j} a_j \cdot \rho(w_{jk}) && \text{(forward pass)} \\ \forall_k : s_k &= R_k / z_k && \text{(element-wise division)} \\ \forall_j : c_j &= \sum_k \rho(w_{jk}) \cdot s_k && \text{(backward pass)} \\ \forall_j : R_j &= a_j c_j && \text{(element-wise product)} \end{aligned}$$

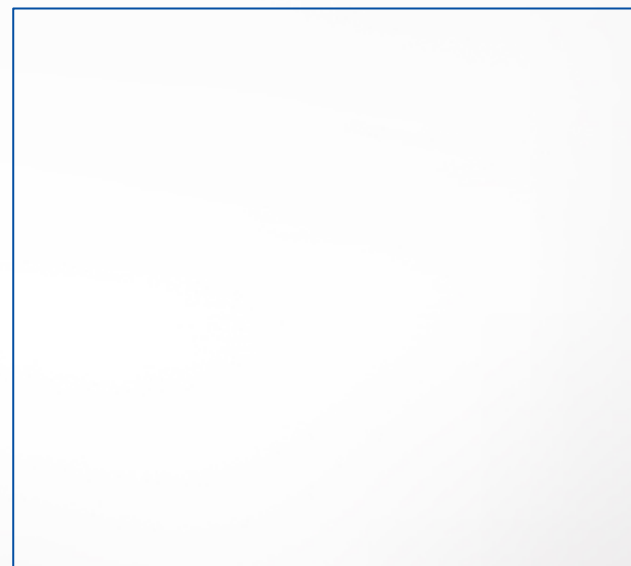
LRP process



제어봉 삽입 위치  
 녹색 : 출력제어 / 적색 : 오정렬



제어봉 오정렬 진단 모델



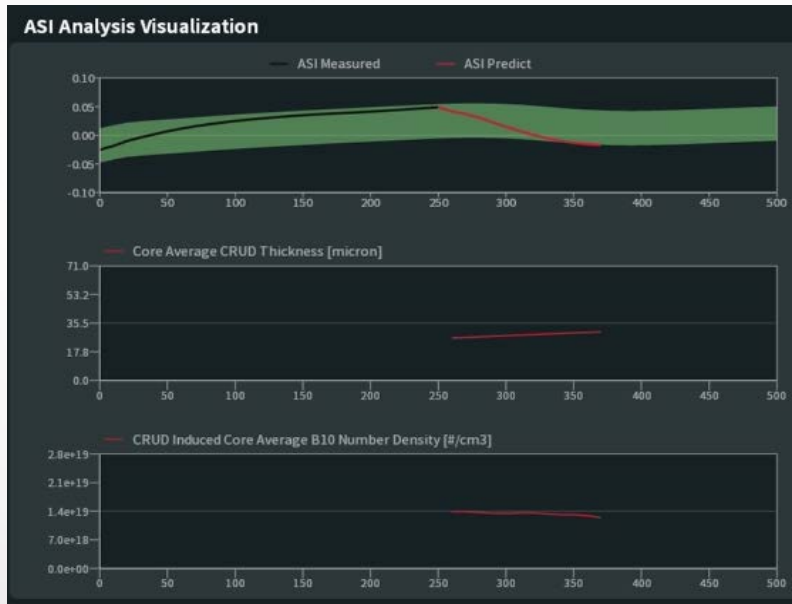
제어봉 오정렬 진단 모델  
 입력층 타당성 (Relavance)

# 3. 연구개발 성과

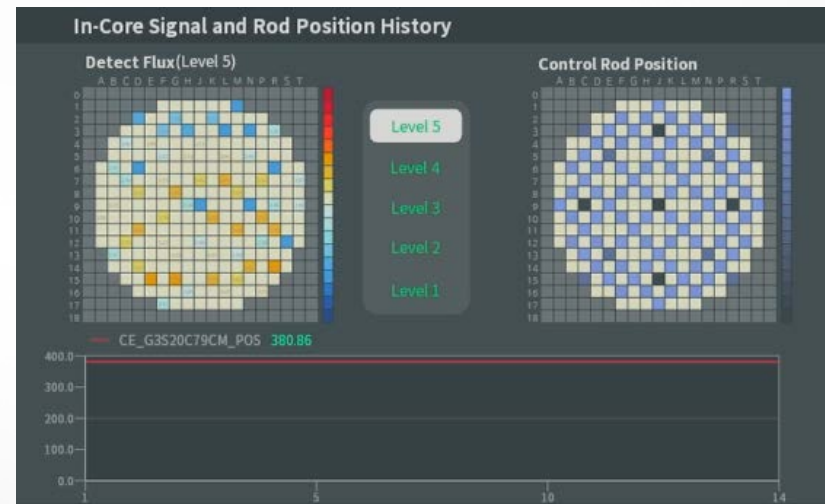
## ❖ 성과목표 ③ : XAI 이상진단 시 판단 근거 타당성

### ❖ 노심 내 크러드 침적량 예측 모델

- ASI 예측 결과에 대한 설명으로, 노심 내 평균 크러드 두께와 크러드로 인한 붕소 수밀도 예측결과를 동시에 가시화하여 운전원에게 현 상황을 설명.
- 미래의 ASI값 예측을 하며 AOA 발생 시 추가정보로 크러드 예측정보 가시화
- 크러드 발생으로 인한 노내계측 신호 PI 데이터 실측값을 추가 정보로 제시



XAI for ASI Prediction



PI 데이터 가시화 UI

# 3. 연구개발 성과

## ❖ 설명가능 인공지능 GUI 개발



온라인 모드 시스템 구성

- PI 중개서버: 발전소 데이터를 전달하는 서버
- Snapshot Generator: PI 중개서버의 데이터를 사용하여 Incore Signal 데이터를 생산하는 모듈
- DAQ: 다른 서버로부터 데이터 적재 및 데이터를 화면에 전송하는 서버
- 노심 비정상 인공지능 서버: 인공지능 모델 실행과 입출력 파일을 관리하는 서버

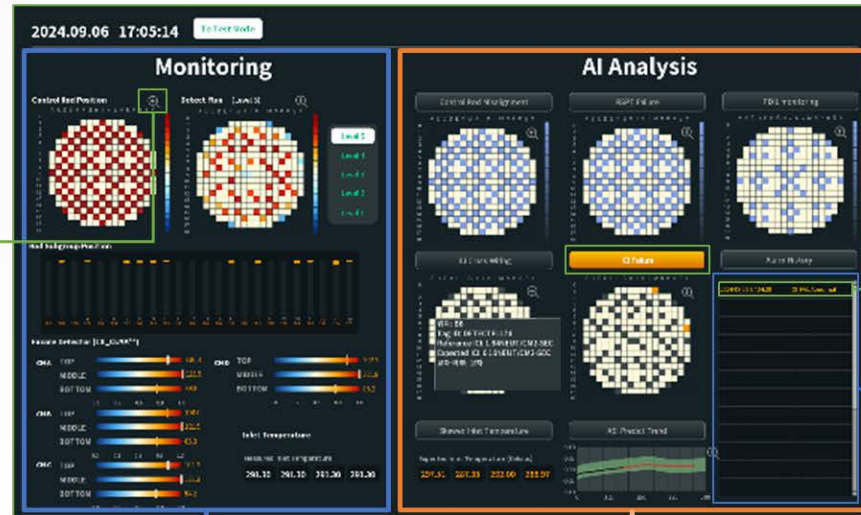
온라인 모드 시스템 프로세스

1. DAQ는 PI 중개서버로부터 실시간 데이터 적재
2. Snapshot Generator는 DAQ를 통해 SnapShot DB에 결과 적재
3. 노심 비정상 인공지능 서버는 DAQ에 모델 입력 데이터 요청
4. 입력 데이터를 사용하여 모델 구동
5. 모델 결과 DB 적재 및 로그 기록
6. 화면은 인공지능 결과를 전송받아 가시화



# 3. 연구개발 성과

## ❖ 설명가능 인공지능 GUI 개발



클릭시 컴포넌트 확대/축소

클릭시 상세 페이지 진입

비정상 발생 시간  
비정상 발생 종류 제공

### 모니터링 페이지

- 온라인 모드일 경우 PI 중개서버로부터 전송받은 데이터 실시간 가시화
- 테스트 모드일 경우 가장 최근 모델이 사용한 입력데이터 가시화
- Control Rod Position, ICI Signal, Rod Subgroup Position, EXI Signal, Inlet Temperature 가시화 컴포넌트 제공

### AI 분석 페이지

- AI의 출력 결과와 비정상 판단 결과 가시화
- 비정상 판단 히스토리 로그 기능 제공
- 비정상 라벨 또는 로그를 클릭하여 상세 페이지 이동
- 컨투어 객체의 Signal에 대응하는 노드 클릭시 상세 메시지 제공

## 4. 연구 추진 노력

### ❖ 1차년도 중간평가 의견에 대한 조치사항

- 본 과제 결과물이 한수원 내부에 활용될 수 있도록 유도
  - 한수원 활용성 증대를 위해 한수원에서 개발 중인 Web-based System과 연계를 위한 대상 호기/주기 통합 (1차년도 OPR1000 → 2차년도 APR1400)
- 한수원의 원자로 실측 데이터 확보를 위한 세부적인 방안 수립 필요.  
원자로 실측 데이터 미확보 시 대안 필요
  - 실측데이터 미확보 대안으로, 한수원으로부터 PI 데이터 운전이력 자료를 받아 데이터 생성에 적극 반영 (출력 감발 시, 붕소/제어봉 동시사용 여부 등)
  - 해당 시점만 진단하는 스냅샷 데이터에서 시계열 데이터로 데이터 차원 확장
- 장기 저출력 운전 경험 노심 대응 데이터 추가 생산 및 노심 이상상황 데이터와 통합

# 5.기술의 완성도 및 활용 계획

## ❖ 인공지능 진단 시스템 적용 범위 확장

- 다양한 노심 상태(출력, 연소도, 제어군 위치)에 대한 진단모델 확장
- 이상진단 범위 및 상세진단 범위 확장
  - 예) 제어군 별 삽입한계 도달 등 운전제한범위 도달 알람모델

## ❖ 설명가능 인공지능(XAI) 결과를 통한 특성 최적화/특성 엔지니어링

- 각 이상상황별 중요한 특성에 집중해 모델의 성능 향상
- 비선형 관계, 상호작용 고려해 해당 정보 반영한 새로운 피쳐 설계

## ❖ 한수원 Web-based 시스템 연계

- 해당 호기/주기 데이터 학습데이터 생산
- 노심 이상 진단모델 및 XAI 모델 결과 가시화
- XAI 기반 노심 이상상황 조기진단 인공지능 모델 연계모듈 개발
- 플랜트 공정 시스템으로부터 데이터 취득, 이상 진단 모델 연계, AI 연계에 필요한 GUI 인터페이스 및 모델 연동
- ✓ 플랜트 공정 시스템과 연동을 수행하기 위한 API 또는 라이브러리를 한수원에서 제공 시 온라인 인터페이스로 설계 및 개발, 플랜트 공정 시스템과 온라인 연결이 불가능한 경우 오프라인 데이터를 활용하여 연계

# 5.기술의 완성도 및 활용 계획

## ❖ 기술 실적용 시 완성도

실적용 시 활용 계획	방법	연구내용	실적(%)
인공지능 의사결정 신뢰성 제고	설명가능 인공지능 기술 적용	XAI 기술 자료수집 및 분석	100
		XAI 기술 적합성 비교	100
		XAI 모델 구현 및 학습	100
		판단 근거의 타당성 검증	100
운전원 실용성 향상	운전관련 정보 제공	기술지침서 제한사항 연계	100
		상세진단 인공지능 진단체계 개발	100
		상세진단 인공지능 학습	100
한수원 내부 활용성 향상	한수원 Web-based 시스템 및 GUI 연계	대상 호기 학습데이터 생산	100
		대상 호기 인공지능 모델 학습	100
		GUI 시각화 상태변수 생산	100

## 6. 결론

- ❖ 한수원의 실제운전 자문을 받아, 7개 노심 이상상황에 대한 RAST-K 기반 학습데이터 생산 및 다양한 이상진단 모델 개발.
- ❖ 각 7개 이상진단 AI 모델의 블랙박스적 특성으로 인한 인허가 문제 해결을 위한 “설명 가능한 인공지능 기술” 검토 및 적용.
- ❖ 개발기술의 실적용을 위해, 공참기관 ENU 주식회사 협업으로 디지털 트윈 서버 페이지로 구동 가능한 웹기반 GUI 연계.
- ❖ SCI급 논문 3건, 관련 특허 출원 2건 (중 1건 등록), 한수원 프로그램 등록 6건, 기술보고서 1건 등 당초 계획 이상의 성과를 제출.



**Q & A**

A decorative horizontal bar consisting of three segments: a blue segment on the left, a green segment in the middle, and a blue segment on the right. It is positioned above and below the 'Q & A' text.

# 추가 보충 슬라이드

## ❖ 인공지능 진단 모델 입출력

- 7개 노심 이상상황 학습데이터 및 진단 모델 입출력 요약

노심 이상상황 진단	회귀모델	모델 입력	모델 출력
제어봉 오정렬 PDIL 위반	소프트맥스 다중 퍼셉트론	1. 노내계측기 신호 2. 노외계측기 신호 3. 노심 출력 [%]	제어봉 위치 [cm]
CIPS 조기 진단 (120일 뒤 ASI 예측)	비전 트랜스포머	1. 노내계측기 신호 2. 노외계측기 신호 3. 노심 출력 계획 [%] 4. 제어봉 위치 [cm]	(미래 예측값) 1. ASI 값 [-] 2. 크러드 두께 [micron] 3. 크러드 침적된 봉소 수밀도 [#/cc]
냉각수 입구온도 비대칭	자동회귀 디코더	1. 노내계측기 신호 2. 노외계측기 신호 3. 노심 출력 [%]	냉각수 입구온도 [°C]
노내계측기 신호 Fail	자가회귀 오토인코더	1. <b>비정상</b> 노내계측기 신호 2. <b>정상</b> 노외계측기 신호 3. 노심 출력 [%] 4. <b>정상</b> 제어봉 위치 [cm]	<b>정상</b> 노내계측기 신호
노내계측기 교차 설치	자가회귀 오토인코더	1. <b>교차된</b> 노내계측기 신호 2. <b>정상</b> 노외계측기 신호 3. 노심 출력 [%] 4. <b>정상</b> 제어봉 위치 [cm]	<b>정상</b> 노내계측기 신호
제어봉 위치지시기 오지시	자가회귀 오토인코더	1. <b>정상</b> 노내계측기 신호 2. <b>정상</b> 노외계측기 신호 3. 노심 출력 [%] 4. <b>비정상</b> 제어봉 위치 [cm]	<b>정상</b> 제어봉 위치 [cm]