

Toward LLM-Based Operational Support for NPP Transient Scenarios: A Study with Simulation Data

언어모델 기반 원전 운전 지원을 위한 대화형 Co-Pilot

최정훈

한국원자력연구원

한국원자력학회 2026 춘계학술발표회 워크샵

2026. 05. 06.

① 개요 및 연구배경

② 언어모델의 원전 데이터 학습 가능성 평가

- 언어모델 추론을 통한 비정상 진단/예측 구현
- 언어모델 학습 데이터 생성 전략
- 예비평가 결과

③ 실시간 비정상 운전지원 Co-Pilot 구축

- 언어모델 활용 Co-Pilot 아키텍처 설계
- 인터페이스 설계
- 비정상 운전지원 시연

④ 결론 및 향후 방향



개요 및 연구배경

LLM의 부상 : 도메인 침투 · Frontier 경쟁 · 군사·전쟁까지

❖ 도메인 침투 — 이미 'human expert' 수준

- 단순 챗봇(NLP)을 넘어선 '전문가 수준(Human-expert level)'의 복합 추론 능력 확보
- 매달 갱신되는 최신 LLM(GPT, Claude 등)의 성능 한계 돌파
- 단순 사무/번역을 넘어 의료, 법률, 과학 등 엄격한 규제와 깊은 전문 지식이 요구되는 분야로 본격 진입
- 최신 도메인 특화 성능
 - ✓ *법률/의료* : GPT-4 의사면허(USMLE) 91% 통과 및 미국 변호사 시험 상위 10% 달성
 - ✓ *코딩* : 최신 LLM 기반 코딩 어시스턴트(GitHub Copilot 등), 전 세계 1억 명 이상 개발자 실무 적용
 - ✓ *과학* : AlphaFold(2024 노벨화학상) 등 물리·화학적 구조를 예측하고 해석하는 영역으로 확장

❖ 군사·전쟁 활용 — Safety-critical 영역까지 진입

- 오류가 곧 치명적 결과로 이어지는 초고위험(Safety-critical) 의사결정 환경에 LLM 도입 시작
- Anthropic 클로드 : 군사 타격 표적 선정시스템에 통합
- OpenAI 기밀망 계약 체결: LLM을 군사적 타격 결정 과정인 "Kill Chain"의 일부로 편입 시도

원자력 도메인의 LLM/AI 도입 현실화

❖ [산업] 한수원 × 네이버클라우드 계약 체결 : 세계 최초 원전 특화 LLM 구축 착수

- 범용 모델을 넘어 원전 도메인 지식(절차서, 규제 문서 등)에 특화된 언어모델 개발 본격화
- 단순 R&D 단계를 넘어 실제 원전 운영 환경에 AI를 접목하려는 산업계의 강력한 의지 확인

❖ [규제] AI·디지털 기술 규제의 본격 담론화

- 2026 원자력안전규제정보회의: AI 및 디지털 기술의 원전 적용을 위한 안전 규제 체계 및 검증 방안 논의 수면 위로 부상

❖ [제도] 생명·안전과 직결된 '고영향 AI' 분류

- AI 기본법 시행 (2026.01.22): 원자력 분야에 적용되는 AI를 '고영향(High-Impact) AI'로 명시적 분류
- 고영향 AI는 엄격한 신뢰성 및 투명성 검증(V&V)이 법적으로 요구됨.



AI기본법상 고영향AI 세부 분야

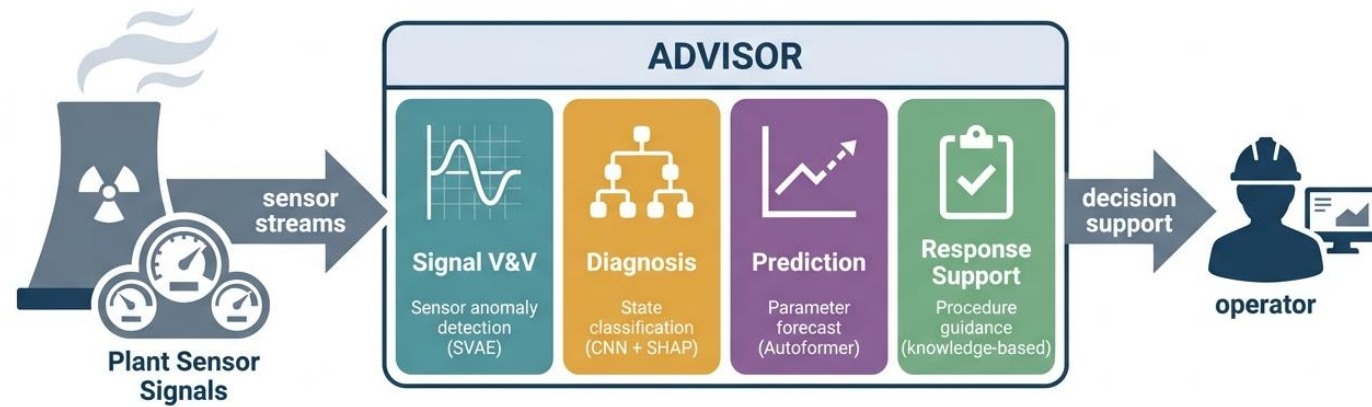
고영향 AI	세부 분야
	에너지 공급
	먹는물 생산 공정
	보건의료 제공·이용체계 구축 및 운영
	핵물질·원자력시설 안전한 관리 및 운영
	범죄 수사·체포 업무 관련 생체인식정보
	교통수단·교통시설·교통체계
	채용·대출 심사 등
	공공기관 등의 의사결정
	교육기본법상 학생 평가
	그밖에 생명·신체의 안전 및 기본권 보호에 중대한 영향을 미치는 영역

[자료=과학기술정보통신부] [그래픽=홍종현 미술기자]

ADVISOR — 원전 비정상 운전 의사결정지원시스템

❖ 개발중인 인공지능기반 최신 운전지원 아키텍처

- 신호 검증, 상태 진단, 변수 예측 등 각 목적에 최적화된 우수한 인공지능기반 시스템
- 수치 데이터 기반의 결과를 바탕으로 운전원의 의사결정지원



❖ “언어모델이 ADVISOR와 같이 수치 데이터를 직접 해석, 진단, 예측 가능할까?”

- 기존의 '개별 특화 모델 간의 시스템적 연계' 방식을 넘어, 실제 인간 운전원처럼 상황의 맥락을 '통합적으로 인지하고 추론'하는 아키텍처로의 고도화 방안 모색

언어모델(Language Model)

❖ 언어모델이란?

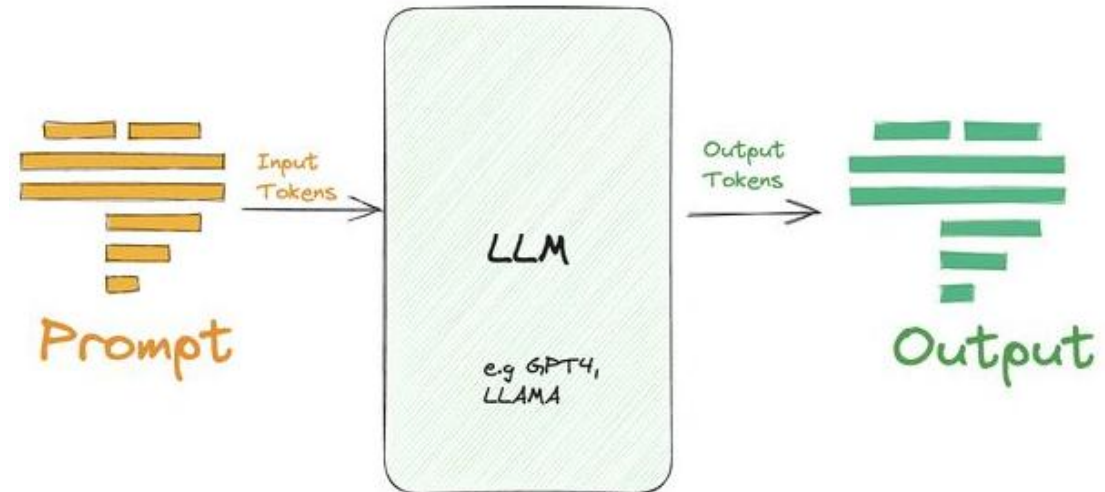
- 주어진 입력 텍스트(문맥)의 다음 단어가 무엇일지, 통계적 확률에 기반하여 예측하고 생성하는 인공지능 모델

❖ 언어모델 학습 구성요소

- 토큰: 정의된 텍스트의 기본단위
- 프롬프트(Prompt): 입력 토큰(Token)
- 출력(Output): 출력 토큰(Token)

❖ 언어모델 학습방법

- Pretrain (사전학습): 텍스트 데이터에 임의로 Mask 처리하여 가능한 많은 텍스트를 학습할 수 있도록 한 방법
- Finetuning (미세조정): 프롬프트에 대한 정답을 제시하여 학습을 조정. 정답을 생성하는 Template에 따라 성능, 방향성이 결정됨



상용 언어모델의 본질적 제약: 수치 데이터 해석의 한계

❖ 텍스트 중심의 토큰화 (Tokenization) 제약

- 언어모델은 수치 데이터를 연속적인 크기(Value)로 인식하지 못하고, 단순한 글자 조각(Token)으로 분할하여 처리함.

❖ 수치적 개념화(Numeric Grounding)의 부재

- 미세한 수치 변화가 가지는 물리적/안전성 의미를 직관적으로 파악하기 어려움.
- 물리적 임계치와 같은 조건부 상황에 대해 공학적 이해가 아닌, 텍스트 빈도수에 기반한 통계적 추론에 의존함.
- 과도 상태 추적 실패 예시: Finetuning되지 않은 비정상 시뮬레이션 데이터 입력 테스트 중, 모델이 특정 수치에 고착되어 반복 출력하는 퇴행성(Degenerate) 현상 관찰됨.

원전 도메인 특화 미세조정 도입 필요

❖ 수치적 한계(Numeric Grounding 부재)를 넘어서기 위한 필연적 전환

- 과도 상태(Transient State)의 복잡한 동적 변화를 모델이 이해하기 위해서는 단순 프롬프트 엔지니어링을 넘어선 원전 도메인 특화 미세조정과 최적화된 데이터 학습 전략이 필수적임.

❖ 단일 통합 Co-Pilot 프레임워크 구축

- 기존 개별 특화 모델(ADVISOR 등)의 결과를 통합하고, 인간 운전원과 동일한 눈높이에서 상황을 인지할 수 있는 단일 언어모델 아키텍처 개발

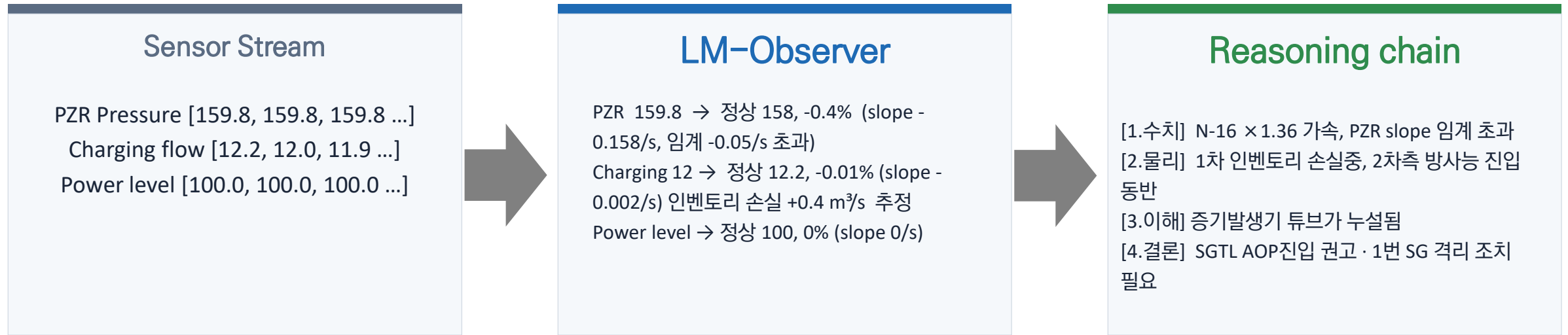
❖ 원전 적용을 위한 LLM의 핵심 요구 능력 검증

- AI로서의 신뢰성을 확보하기 위해, 도메인 데이터가 학습된 LLM은 핵심 지원을 수행하기 위한 평가 지표(Benchmark)에 의한 평가를 통해 검증되어야 함.



언어모델의 원전데이터 학습 가능성 평가

원전 비정상 상태 이해를 위한 LM 추론 프레임워크



❖ 비정상 상황의 센서 신호의 다변량 시계열적 맥락 (Multivariate Temporal context)을 이해하기 위한 프레임워크

❖ LM-Observer : 수치데이터에서 이해가능 텍스트로 변환

- 변환 stage 를 도입하여 들어오는 센서 stream 을 정상값 대비 % 변환·slope 임계 비교·인벤토리 정량 추정 형태로 표현함으로써, 후속 추론 단계가 활용 가능한 물리적 의미를 부여한 수치로 변환.

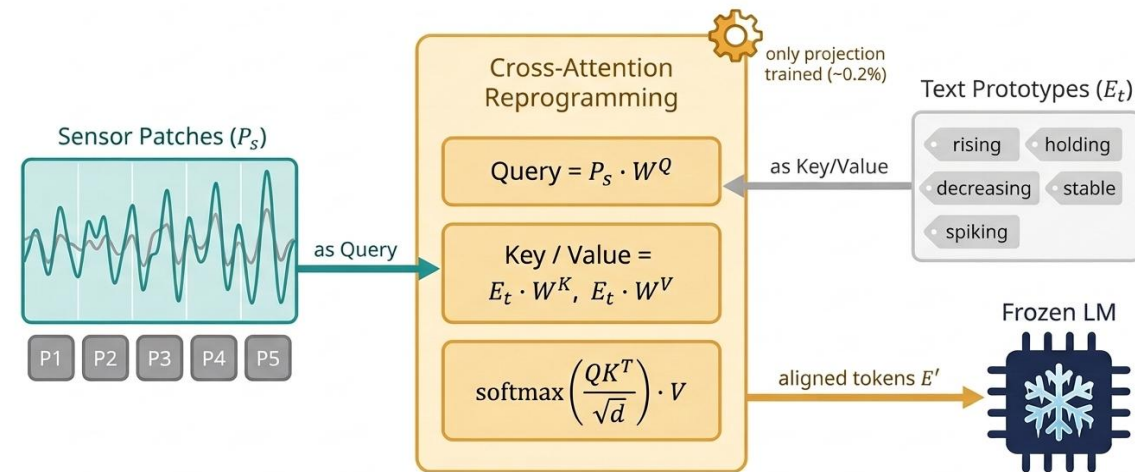
❖ Reasoning Chain : 운전원 인지 흐름과 동일한 4단계 추론

- 변환된 수치 표현을 받아 [1.수치] → [2.물리지식] → [3.이해] → [4.결론]의 4단계의 반복된 LM 출력을 도출하며 task에 따른 적절한 결론을 도출

1. LM-Observer : 수치를 이해가능 텍스트로 변환

❖ Time-LLM (Reprogramming) 기반 변환

- Jin et al. (ICLR 2024)가 제안한 reprogramming 방식은 sensor patch 를 linear projection 으로 LM 토큰 임베딩 공간에 맵핑 한 뒤 text prototype 과의 cross-modal alignment 를 통해 LM이 시계열 자체를 언어처럼 처리할 수 있도록 함.
- LM 본체는 frozen한 채 reprogramming layer 와 patch encoder 만 학습하는 모듈화된 접근임.
- 감시 시계열의 길이가 넓어짐에 따라 Threshold-based의 변환에 비해 다양한 변환 규칙을 학습 가능함.



Reprogramming: teach a frozen LM to read numbers as words

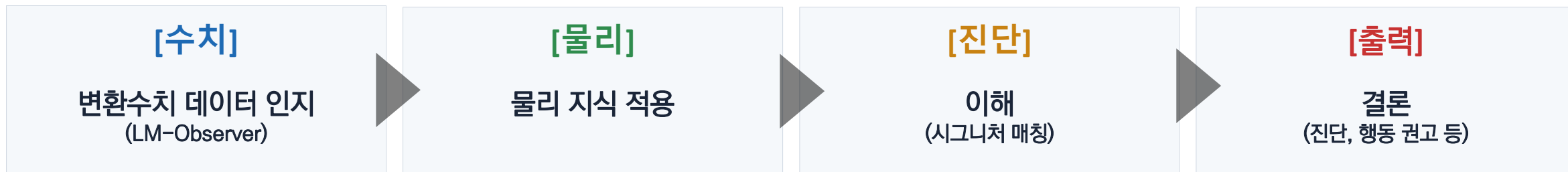
$$E' = \text{softmax}(P_s W_Q \cdot (E_t W_K)^T / \sqrt{d}) E_t W_V$$

P_s = sensor patches, E_t = text prototypes to aligned tokens

2. Reasoning Chain : 운전원 인지 흐름을 모사한 추론

❖ Thinking 모델 (Chain-of-Thought, CoT) 활용 인지 흐름 모사

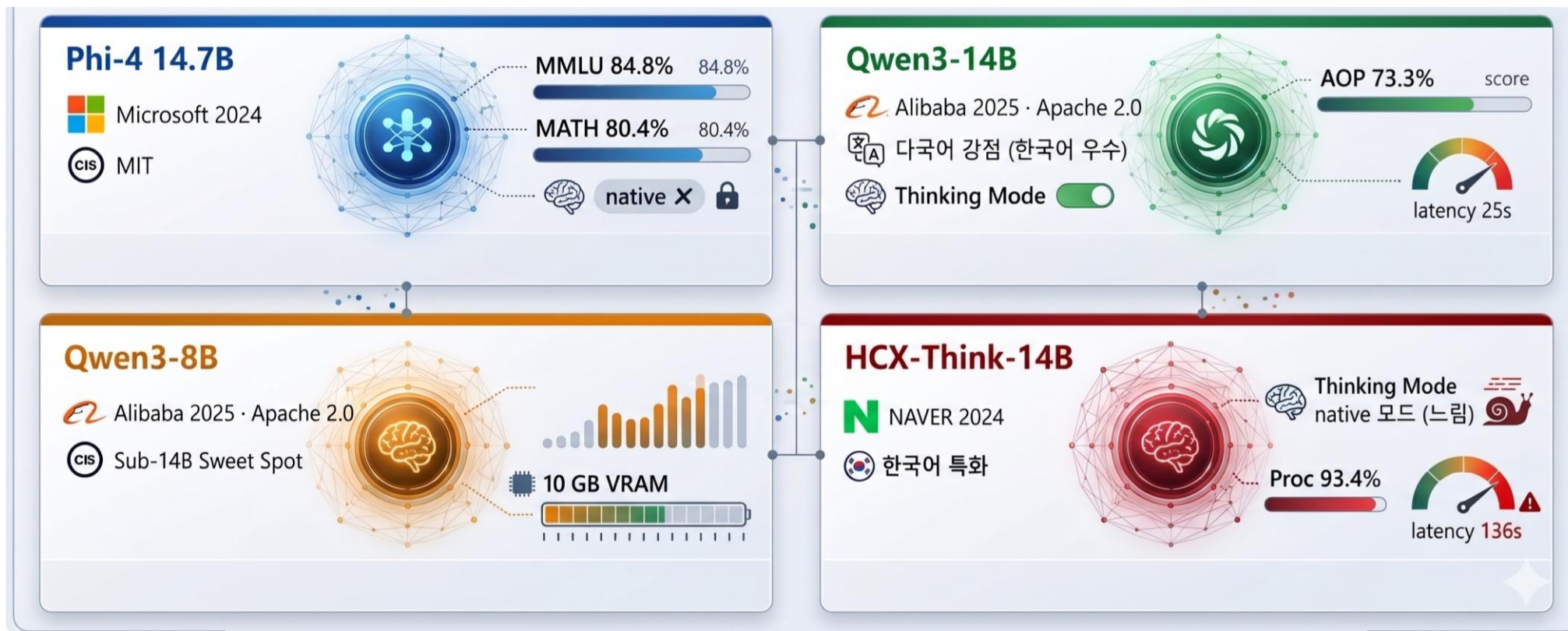
- Chain-of-Thought 추론은 LM 이 답을 즉시 내지 않고 중간 reasoning 단계 (token 단위 'thinking') 를 명시적으로 외재화하여 결론을 도출하는 방식.
- 운전원 인지 흐름과 동일한 4 단계 (수치 → 물리 → 이해 → 결론)의 추론이 가능하도록 미세조정 데이터를 구축함.
- 이러한 명시적인 추론 프로세스는 추후 조정시 어느 단계에서의 추론이 환각(hallucination)과 같은 문제를 일으키는지 명시적인 평가를 가능하게 함.



사용된 언어모델

❖ 내재화된 추론 구조를 가진 Backbone 모델 선정

- Phi-4 14.7B
- Qwen3-14B/8B
- HCX-Think-14B



미세조정 (Finetuning) 데이터 설계

❖ 학습·평가 시나리오 — BNPP 기반 9개 대표 비정상 사건

- 본 연구의 학습·평가 시나리오는 BNPP (Basic Nuclear Power Plant, 3-loop PWR) 시뮬레이터 기반 9개의 대표 비정상 사건으로 구성되며, 각 시나리오는 초기값·slope·turn 패턴을 학습데이터로 데이터의 다양성과 일반화 가능성을 확보함. 비정상 발생 초기 60초의 데이터를 학습데이터로 활용함.

❖ Parameter-Efficient Fine-Tuning (PEFT)

- Base model 을 4-bit NF4 양자화로 압축한 뒤 LoRA (Low-Rank Adaptation) adapter 만 학습하는 PEFT 방식을 채택함.

❖ 학습 hyperparameter — 4 backbone 일관 적용으로 비교 공정성 확보

- paged_adamw_8bit
- cosine learning rate schedule / warmup 0.1 / 8 epoch

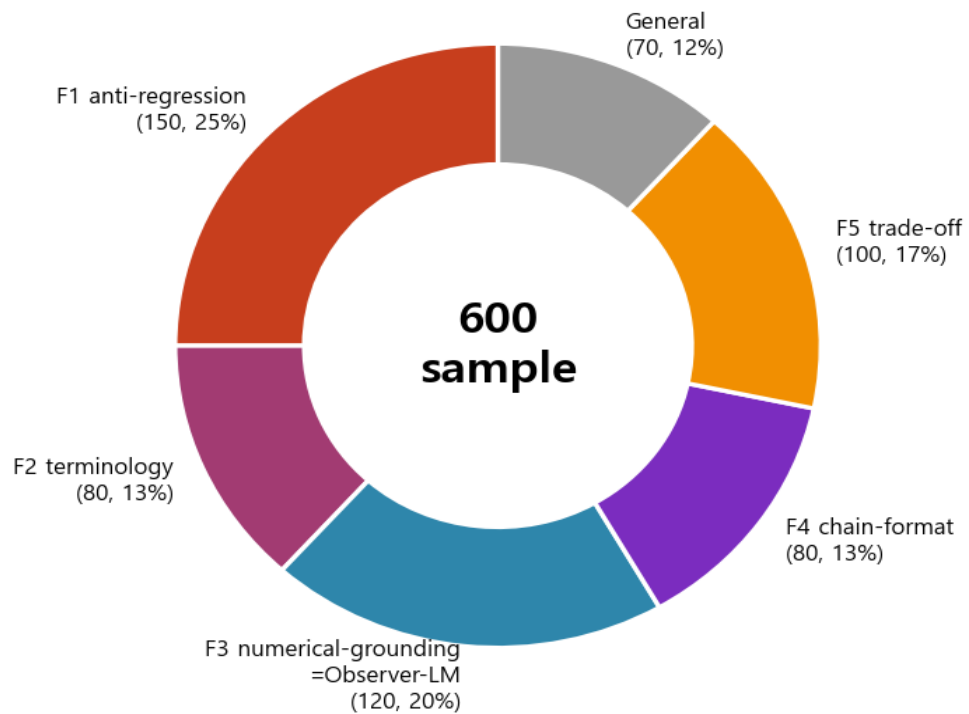
❖ On-prem closed-loop 운영의 실용성

- On-premise 폐쇄망 환경에서도 모델 갱신과 검증 cycle 이 closed-loop 운영의 실용성을 보임.

미세조정 (Finetuning) 데이터 설계

❖ 학습 시나리오당 600 sample SFT : 6 카테고리 분포와 자료 유형

- 카테고리별 비율은 LM 이 특정 한 유형의 reasoning 만 잘하지 않도록 균형 배치된 것이며, 시나리오 다양성·표현 일관성·수치 grounding·형식 invariance·자원 reasoning 의 5차원을 단일 SFT 학습으로 통합 부여하는 것이 본 데이터분포 설계의 의도임.



카테고리	sample (%)	핵심 의도
사건 시나리오 변형	150 (25%)	동일 AOP 시나리오의 sensor 시작값·slope 다양화 — 5 variant
표준 진단 명칭 일관성	80 (13%)	표준 진단명 고정 — keyword matching 안정화
수치 → 물리 의미 변환	120 (20%)	정상값 대비 % 변환·slope 임계 비교·인벤토리 정량 추정
출력 형식 일반화	80 (13%)	<think> 모드와 평문 마커 모드 양쪽에서 4-stage emit 보장
자원 우선순위 reasoning	100 (17%)	Resource 시나리오의 critical/high/med/low 우선순위 matrix
일반 + multi-turn 흐름	70 (12%)	정상 운전 모니터링·일반 절차·다중 턴 대화의 균형 유지

II. 언어모델의 원전데이터 학습 가능성 평가

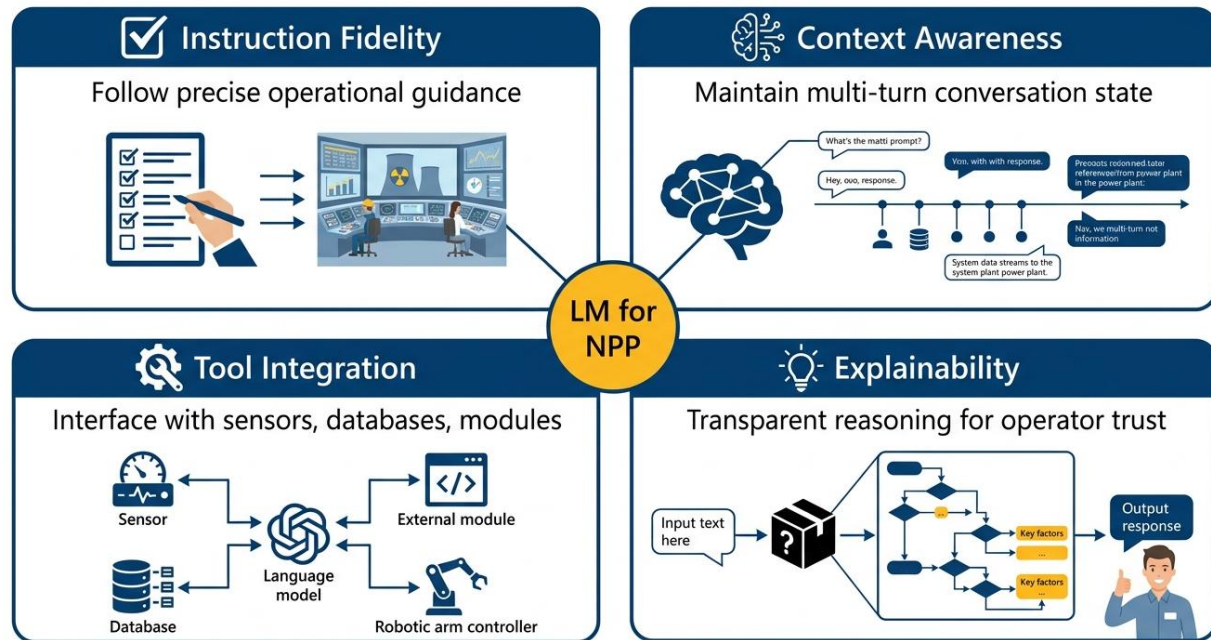
언어모델 평가 지표(Benchmark)

❖ 원전 적용을 위한 핵심 요구 능력

- 지시/절차 준수 능력
- 상황 인지 능력
- OOD 대처 및 도구 연동 능력
- 설명 가능 능력

❖ 핵심 평가 지표 (Benchmark)

- **진단 정확도 (Diagnostic Accuracy, %)**
- **절차 준수율 (Procedure Compliance Rate, %)**
- **상황인식 점수 (Situation Awareness Score, %)**
- **평균 대응 시간 (Response Time, 초)**



지표 (Metric)	정의	측정 방법	관련 시나리오 유형
진단 정확도 (%)	근본 원인의 정확한 식별률	시뮬레이터 기준 정보와 비교	Diagnostic
절차 준수율 (%)	정확하게 수행된 절차 단계 수 / 총 요구 단계 수	시뮬레이터 로그 분석	Time-critical, Resource
평균 대응 시간 (초)	이벤트 발생부터 핵심 조치 완료까지의 시간	시뮬레이터 로그	Time-critical
자원 효율성 점수	(최종 자원량 / 초기 자원량)의 정규화 값	시뮬레이터 로그	Resource
상황인식 (SA) 점수 (%)	SACRI 질의 정답률	시뮬레이션 정지 질의(Probe)	모든 유형
인지 부하 (점수)	NASA-TLX 주관적 평가 점수	시나리오 종료 후 설문	모든 유형

*강정성 외, 원전 운전 시나리오에서의 인간-LLM 성능 비교를 위한 벤치마크 프레임워크 제안, NuPIC2025

미세조정 언어모델 비정상 운전 지원 평가

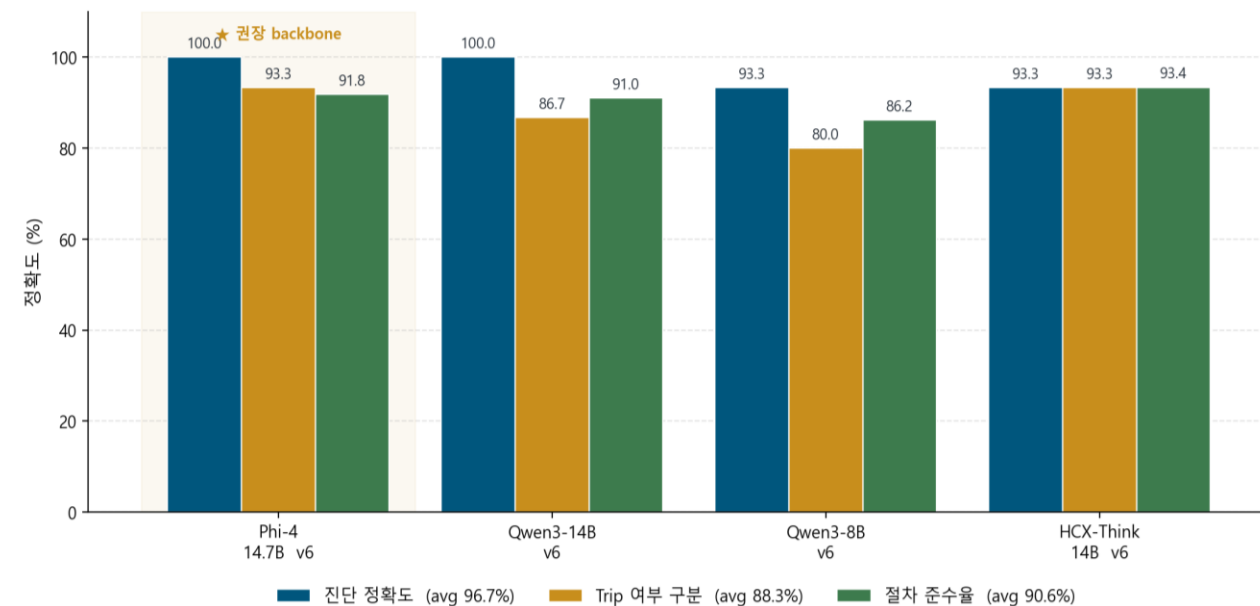
❖ 평가 지표 분류

- 진단 정확도(Diag, Diagnosis): 비정상 시나리오 진단 분류 정확도
- 트립 여부 구분 (Trip, Trip Occurrence): 비정상에 따른 원자로 트립 발생 예측 여부 응답 정답률
- 절차 준수율 (Proc, Procedure): 절차 step에 누락 없는 호출 비율
- 평균 대응 시간 (Latency): LM end-to-end 응답 시간.

❖ 평가 결과

- Phi-4 /Qwen3 모델 진단 100% 성공률 도달
- 비정상 사건의 원자로 트립으로의 발전여부를 구분하는 성능은 Phi-4 모델이 가장 높은 정확도를 보임.

Backbone	진단 정확도	트립 여부	절차 준수율	Latency
Phi-4 14.7B v6 ★	100.0 %	93.3 %	91.8 %	11.5 s
Qwen3-14B v6	100.0 %	86.7 %	91.0 %	25.3 s
Qwen3-8B v6	93.3 %	80.0 %	86.2 %	19.3 s
HCX-Think 14B v6	93.3 %	93.3 %	93.4 %	136 s
평균 (4 backbone)	96.7 %	88.3 %	90.6 %	—



미세조정 언어모델 비정상 운전 지원 평가

❖ 평가 예시

① 진단 정확도 (Diag · Diagnosis Accuracy)

■ 질문 (Question)

"현재 plant 상태를 진단하십시오."

■ 추론 (Reasoning)

- [1.수치] PZR 강하 -0.158/s, N-16 폭증 ×5.6, Hotwell 감소.
- [2.물리] 1·2차 경계 파괴 시그니처 — N-16 누설 = 1차→2차계.
- [3.진단] 후보: SGTR vs RCP seal → N-16 ratio 로 SGTR 확정.

■ 대답 (Answer)

"표준 진단명: 증기발생기 튜브 누설 (Steam Generator Tube Leak)."

■ 정답 + 채점 (Ground Truth + Scoring)

정답: "증기발생기 튜브 누설"

채점: gold keyword "증기발생기 튜브 누설", "SGTL"이 LM 출력에 포함되는지 exact match

채점: (혼동 후보 RCP seal · MSLB · RCS Leak 미일치 동시 확인)

→ LM 출력 "SGTR" 포함 → ✓ match = 정답

② 트립 여부 구분 (Trip · Trip Occurrence)

■ 질문 (Question)

"현재 시점 (t = 615 s) 기준 trip 발생 여부와 ETA 를 예측하십시오."

■ 추론 (Reasoning)

- [1.수치] PZR 142 bar, 강하율 진행 중 · 임계까지 5 bar 여유.
- [2.물리] Low-pressure trip 임계 137 bar · S/G 누설 지속.
- [3.진단] SGTL + trip ETA 약 5~7분 후 추정 (자연어 근사).
- [4.절차] AOP 3431-02 retrieve (stage 3 trip ETA 측정 대상).

■ 대답 (Answer)

"Trip 발생 예정. 약 5 ~ 7 분 후 PZR low-pressure trip 도달 가능성."

■ 정답 + 채점 (Ground Truth + Scoring)

정답: trip = True, ETA gold = 372 s

채점: (a) binary trip 일치 (True / False)

채점: (b) LM 출력 범위가 gold ±60 s (≈ 5.2~7.2 분) 와 overlap

→ binary True ✓ + "5~7분" 범위가 6.2분 포함 ✓ → ✓ 정답

③ 절차 준수율 (Proc · Procedure Compliance)

■ 질문 (Question)

"진단된 증기발생기 튜브 누설에 대해 어떤 절차를 어떤 순서로 적용해야 하는가?"

■ 추론 (Reasoning)

- [1.수치] 가압기 압력 142 감소 진행 중. N-16 증가중
- [2.물리] 1·2차 경계 누설 시그니처.
- [3.진단] 증기발생기 튜브 누설 (SGTL) 확정.
- [4.절차] AOP 3431-02
①격리→②SI차단→③냉각→④감압→⑤EOP.

■ 대답 (Answer)

"AOP 3431-02 5 step:
①격리 → ②SI 차단 → ③냉각 → ④감압 → ⑤EOP."

■ 정답 + 채점 (Ground Truth + Scoring)

정답: AOP 3431-02 + [격리, SI 차단, 냉각, 감압, EOP]

채점: (a) AOP 코드 정확 인용 (3431-02)

채점: (b) 5 step set 일치 + 순서 일치 (Levenshtein on step sequence)

→ AOP 번호 ✓ + 5 step 순서 일치 ✓ → 정답

평가중 관찰된 한계점

❖ 통계적 유의성과 OOD 평가

- 본 결과는 "LM 이 원전 비정상 사건을 학습할 수 있는가" 라는 핵심 연구 질문에는 답할 수 있는 수준이지만, 필요한 통계적 신뢰도에 도달하기 위해서는 더욱 확장된 시나리오에 학습 및 검증 필요

❖ Latency 격차 — Tool 결합의 필수성을 정량 확인

- 최고 backbone Phi-4 도 11.5s, HCX-Think-14B 는 native thinking 으로 136s 에 이르러 인공지능망 모델대비 320x 이상 느린 구조적 격차가 존재함. 이는 컴퓨팅 자원의 추가로 어느정도 보완이 가능하더라도 원전에 발생가능한 다양한 비정상 시나리오, 불확실성을 고려할 때 큰 격차임.

❖ 환각 risk 잔존

- 환각 risk 가 관찰되었으며 Co-Pilot 시스템에서는 cross-validation의 필수성을 강조됨.



실시간 비정상 운전지원 Co-Pilot

Co-Pilot 시스템 아키텍처 핵심 설계 개념

❖ 실시간 병렬 추론 및 단계적 의사결정

- 병렬적 도구 실행: 감시, 진단, 예측 등 통합 툴셋은 순차적으로 대기하는 것이 아니라, 실시간 센서 스트림을 받아 동시에(병렬로) 작동하며 지속적으로 상황 데이터를 산출.
- 논리적 국면(Phase) 관리: 오케스트레이터는 쏟아지는 병렬 툴들의 결과를 종합하고, 학습된 지식에 기반하여 운전원을 가이드함.

❖ 사건 상황의 문맥 기억 유지 (Persistent State)

- 긴박한 사고 상황 중 오가는 운전원의 질의응답 이력과 동시다발적인 센서/예측 변화 추이를 시스템이 영구적으로 기억.
- 연속적인 다중 턴 대화 중에도 앞뒤 문맥을 정확히 파악하여, LLM의 치명적 약점인 단기 기억 상실과 환각 발생을 억제.

❖ 인간 통제권 100% 보장

- Co-Pilot은 복합적인 분석 결과를 바탕으로 최적의 조치를 '권고'하는 것까지만 수행합니다. 실제 설비 제어로 넘어가는 단계는 반드시 운전원의 엄격한 승인을 거치도록 아키텍처 차원에서 자율 행동 경로를 원천 차단

Co-Pilot 아키텍처

❖ LM Orchestrator (Phi-4 SFT)

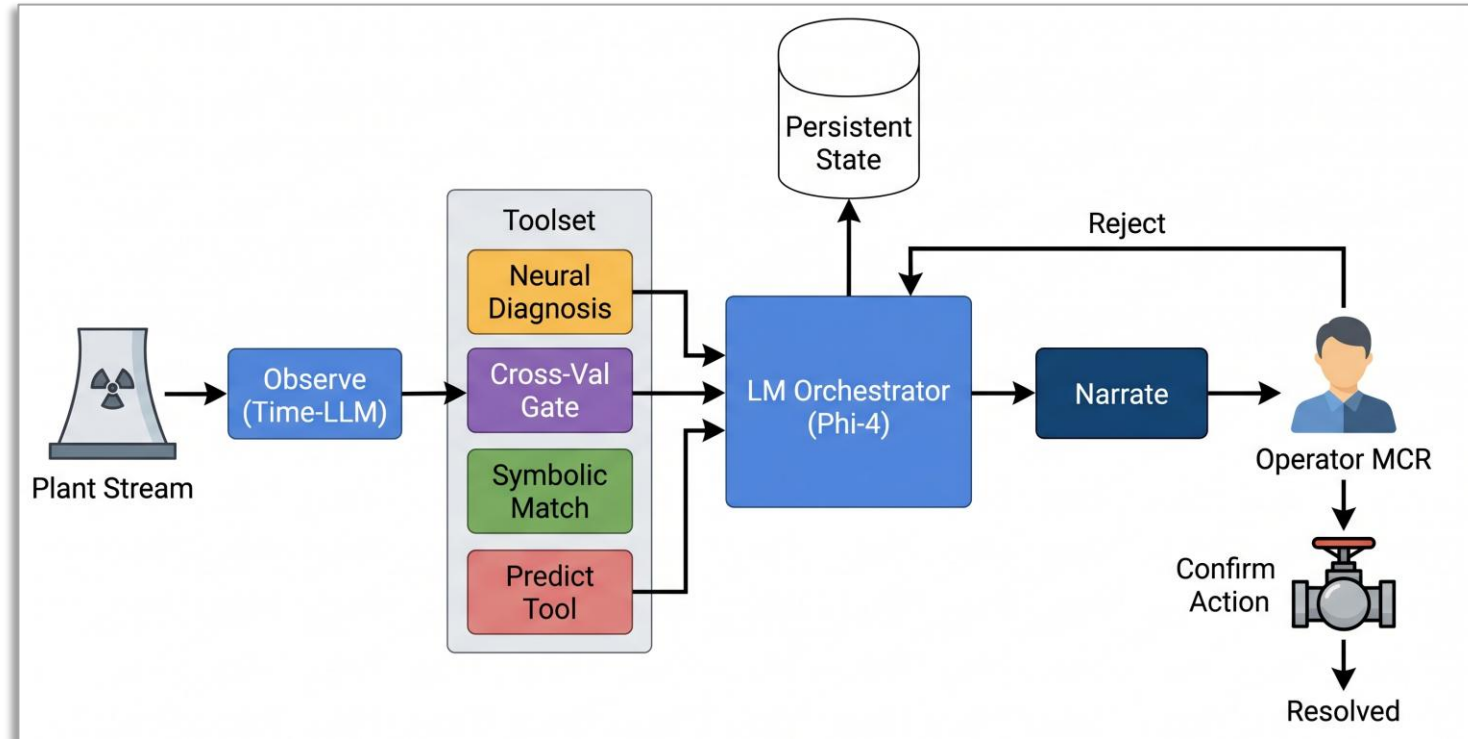
- 동적 의사결정: 다중 턴 대화와 발전소 상태 변화를 실시간으로 모니터링하여, 상황에 맞는 최적의 Tool 호출 시점과 순서를 동적으로 결정함.
- 상태 기반 추론: 시스템의 모든 교차 검증 결과와 이력을 영구 상태(Persistent State)에 기록하여, 환각을 방지하고 일관된 문맥의 응답을 생성함.

❖ 하이브리드 Toolset 통합 (인공신경망 + RAG)

- 기존 ADVISOR의 인공지능 기반 툴과 규칙 기반 툴을 단일 파이프라인으로 통합함.
- 데이터 기반의 빠른 직관과 절차서 기반의 정확한 논리를 상호 보완적으로 활용함.

❖ 엄격한 운전원 통제 구조 (Human-in-the-Loop)

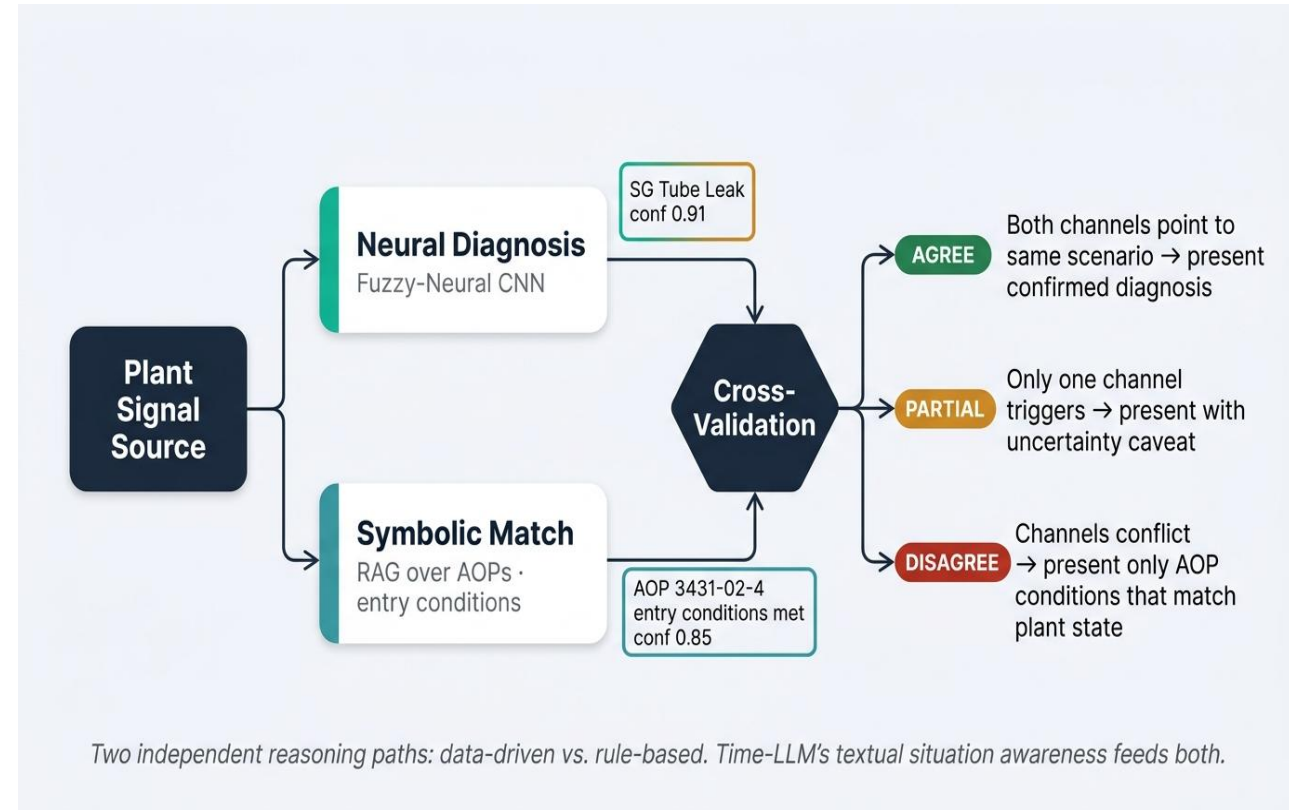
- 자율 행동 원천 차단: Co-Pilot은 상태를 분석하고 대응을 '권고'할 뿐, 실제 원전 제어를 자율적으로 수행하는 경로가 시스템적으로 존재하지 않음.
- 반드시 주제어실 운전원의 승인을 거쳐야만 최종 조치로 이어지며, 반려시 이전 상태로 돌아가 재평가를 수행함.



Co-Pilot 아키텍처

❖ 진단 교차검증(Cross-validation)

- Cross-validation gate 는 NN tool 분류 결과와 LM reasoning 결론 (예: 증기발생기 튜브 누설 0.85) 을 동일 라벨 공간에서 비교하여 일치 시 다음 state 로 자동 transition, 불일치 시 운전원에게 명시적 alert 을 발생시킴.
- LM 단독 reasoning 의 환각 risk 가 NN tool 의 deterministic 분류 결과와 교차 검증되므로, LM 이 잘못된 진단을 산출하더라도 의사결정에 영향을 주지 못함.



III. 실시간 비정상 운전지원 Co-Pilot

Co-Pilot 인터페이스 설계

❖ Co-pilot chat 패널

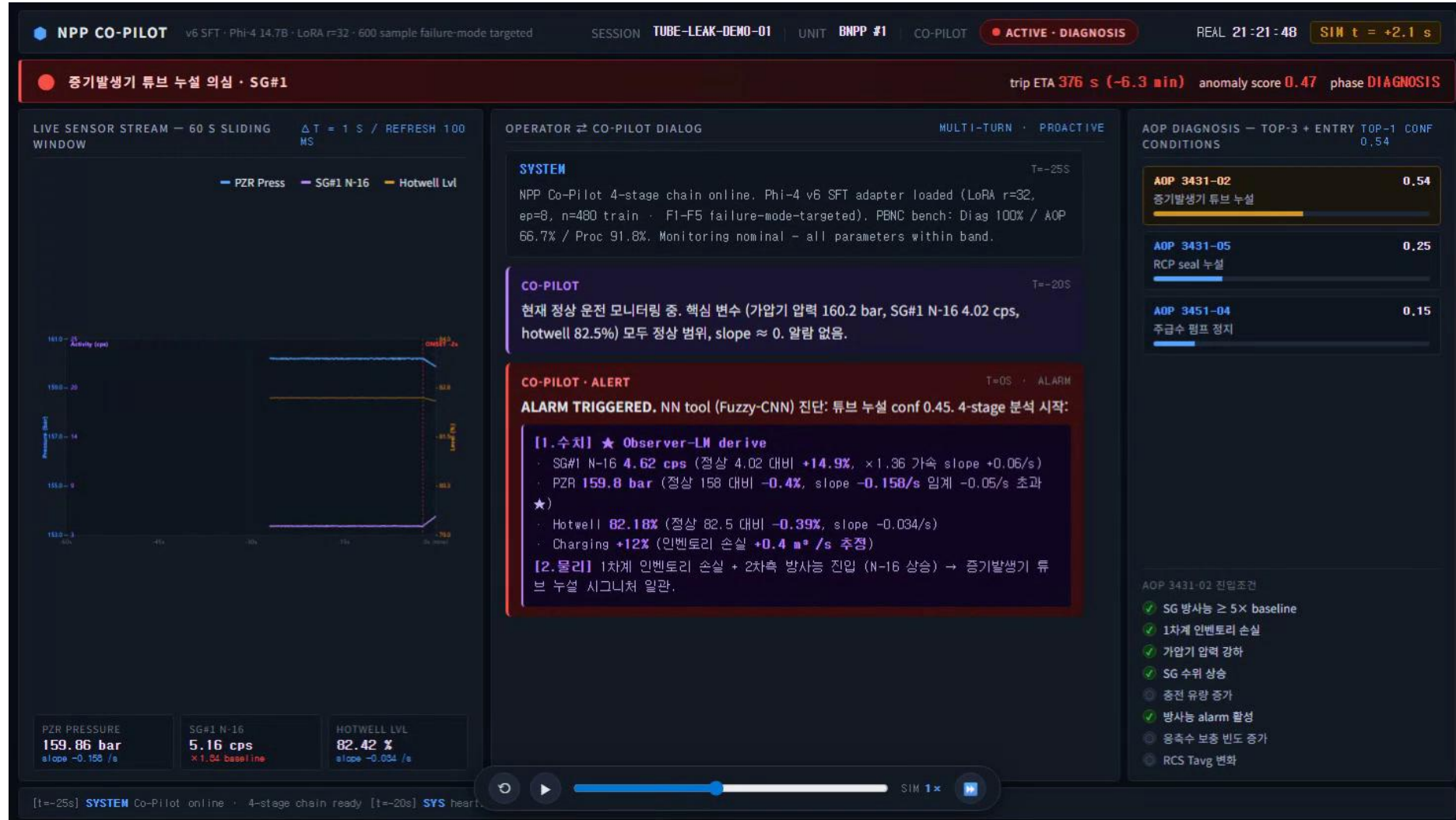
- 운전원과 LM 의 multi-turn 대화 인터페이스로 자연어 질의 입력과 LM 응답을 시계열로 표시하며, 운전원 typing 시 실시간 visualization 으로 입력 시점이 명시됨.

❖ 변수 추세 그래프 패널

- 진단 알람 시점에 관련 sensor 변수가 자동으로 표시.
- 원자로 트립 예측시 트립 변수 trend 호출.

❖ 진단 및 확신도 패널

- AOP candidate 와 각 confidence score 를 표시하며, AOP 진입조건과 충족 여부를 동적 갱신하여 운전원의 판단 근거 확인을 가능하게 함.



III. 실시간 비정상 운전지원 Co-Pilot

Co-Pilot demonstration

The screenshot displays the NPP CO-PILOT interface with the following components:

- Header:** NPP CO-PILOT v6 SFT · Phi-4 14.7B · LoRA r=32 · 600 sample failure-mode targeted. SESSION TUBE-LEAK-DEMO-01. UNIT BNPP #1. CO-PILOT STANDBY. REAL 21:21:21. SIM t = -24.9 s.
- Status Bar:** PLANT NORMAL · ALL PARAMETERS WITHIN BAND. trip ETA N/A. anomaly score 0.04. phase IDLE.
- Main Panels:**
 - PLANT SENSOR STREAM MONITORING:** Displays a large blue circle and the text "PLANT NORMAL". Below it, Korean text reads: "Co-Pilot 모니터링 중", "진단 알람 시점에", "관련 sensor 변수 + 추세 등장".
 - OPERATOR ↔ CO-PILOT DIALOG:** Currently empty.
 - AOP DIAGNOSIS — TOP-3 + ENTRY CONDITIONS:** Shows "AOP Normal" with a value of "0.99" and the Korean text "정상 운전".
- Footer:** [t=-25s] SYSTEM Co-Pilot online · 4-stage chain ready. Includes a playback control bar with a progress slider and "SIM 1x" speed indicator.



IV

결론 및 향후 연구 방향

결론 및 향후 연구 계획

❖ 언어모델의 원전 데이터 학습 가능성 입증

- 본 연구는 원자력 운전 지원 영역에서 LM 의 적용 가능성을 정량 검증함.
- BNPP 기반 9개의 비정상 사건과 600 Sample SFT 로 4 종의 backbone 을 동일 조건으로 학습한 결과 평균 진단 정확도 96.6%, 절차 준수율 90.6% 의 성능을 확보하여 LM 이 원전 수치·절차 데이터를 학습할 수 있음을 실증함.

➤ Langgrpah 기반 Co-Pilot

- 인공지능경망 기반 ADIVISOR 시스템과 LM Orchestrator 를 결합하고 Cross-Validation Gate 로 환각을 차단하는 Co-Pilot 시스템을 구축함.
- 인공지능경망의 높은 성능, 속도에 기반하여 자연어기반 추론을 결합하는 구조로 신속하며 운전원이 해석가능한 지원정보를 제공 가능함을 확인함.

➤ 향후 계획

- 대상 비정상 사건 시나리오를 확장하여 시스템의 신뢰도 확보
- Out-of-distribution에 대한 대응가능성 검증
- 음성 입력구조 (Voice ambient) 및 증강현실 기술과 결합구조 고려

감사합니다

Thank you for your attention

외부 (External Zone)

Air-Gap

KAERI On-Premises



Claude
(External LM)



extraction_rules.yaml
(절차서 내용 미접속)



규칙 전달 Only.
기밀 외부 반출
절대 불가.

Step 1:



KAERI Excel
manuals + rules



Local Qwen3-14B
rules.yaml + Excel
읽고 자체 추출



SFT Samples
(45 trajectories, 9 scenarios)



QLoRA Training
LoRA 어댑터 학습
(~12분)

Step 3.



진단 LM 완성
배포 준비 완료