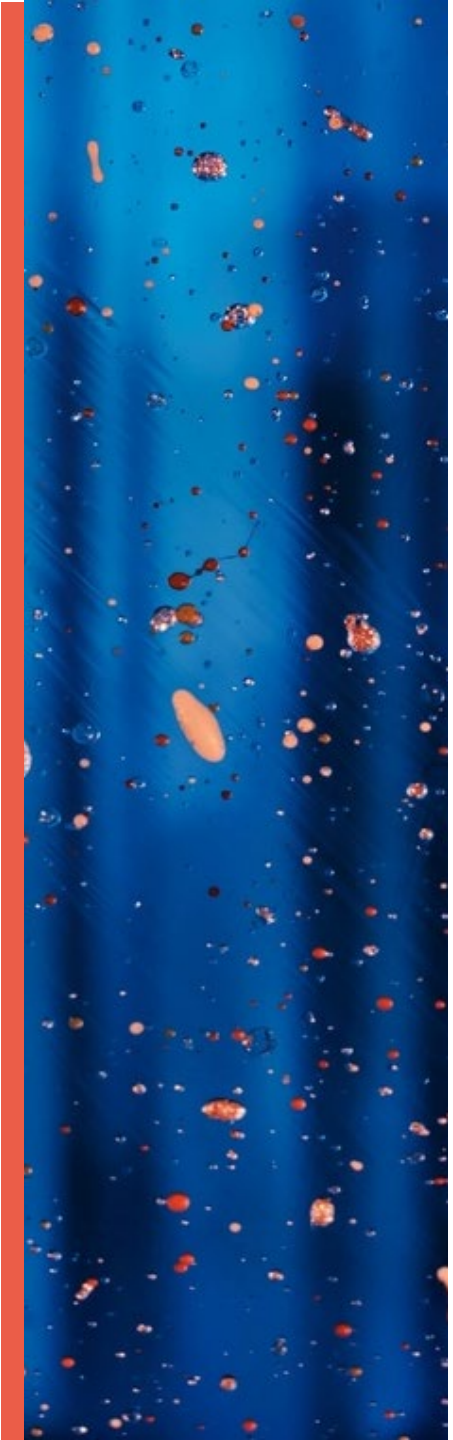


2단계 연구계획-
LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술

카이스트 SWTV 연구실
발표자 이아청

KAIST

2026 겨울 STAAR 워크샵

“AI기반 자동화 테스팅 기술”이 오픈 소스에 기여하고 있는가?

- 구글은 1k 이상의 오픈 소스 프로젝트에
  13k개 이상의 취약점과 50k이상의 버그를 보고

- ZigZagFuzz로 12h내에 15개 오픈 소스 프로젝트에서 85개 버그 보고
  - FFMpeg(멀티미디어 프레임워크)에서 11개 보고
  - Libxml2(XML/HTML 라이브러리)에서 1개 보고
    - “15년”동안 유지되었던 버그

- To Google, “자금을 지원하던가, 버그 제보를 중단하라”

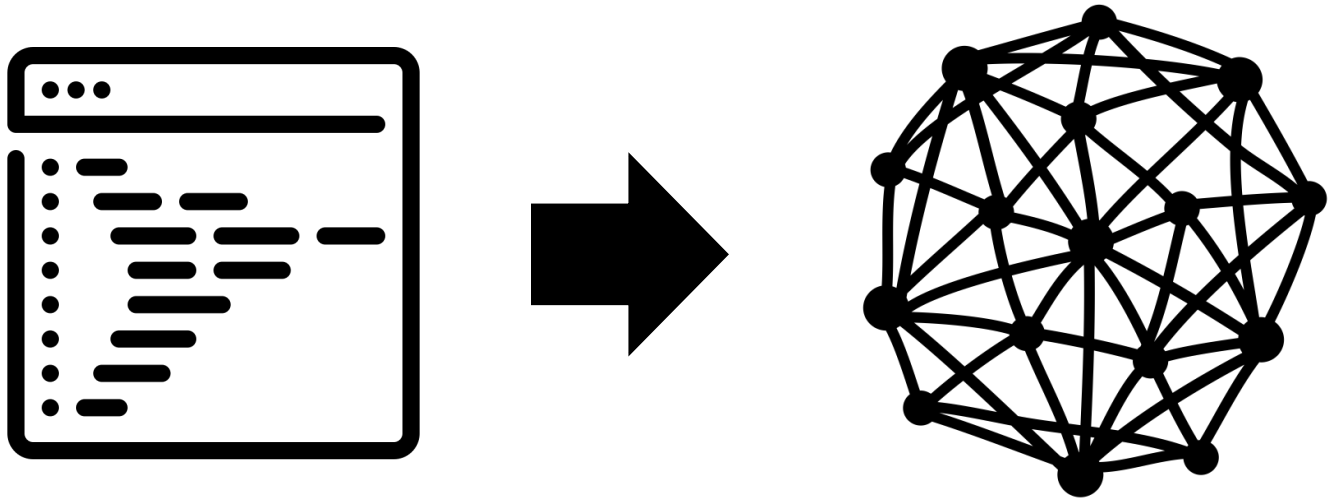
- 너무 많은 AI-생성 리포트로 인해 버그 바운티 프로그램 중단

- 20년 넘게 유지보수를 해오던 Nick Wellnhofer는
너무 많은 보안 리포트 처리가 불가능하다며
유지보수 중단을 선언.

- 구글은 “프로젝트 제로”를 진행중
  - 취약점/버그가 발견되면, 벤더에게 보고 후, 90일 후 공개
  - 90일 안에 고쳐지지 않는다면, 고쳐지지 않은 취약점이 공개

- 유지 보수가 빠르게 되지 않는 오래된 오픈 소스는 어떻게 취약점을 다루어야 하는가?

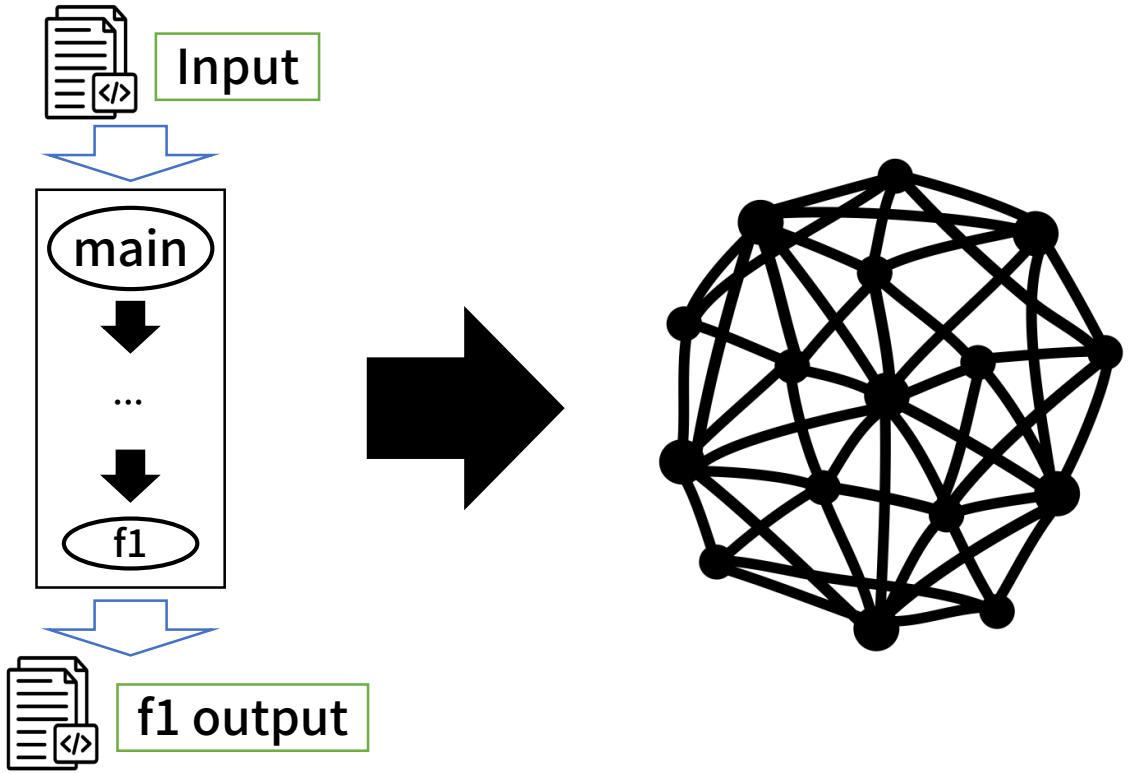
오래된 오픈 소스 코드의 입력/출력을 모조리 배운 네트워크로 대체하면?



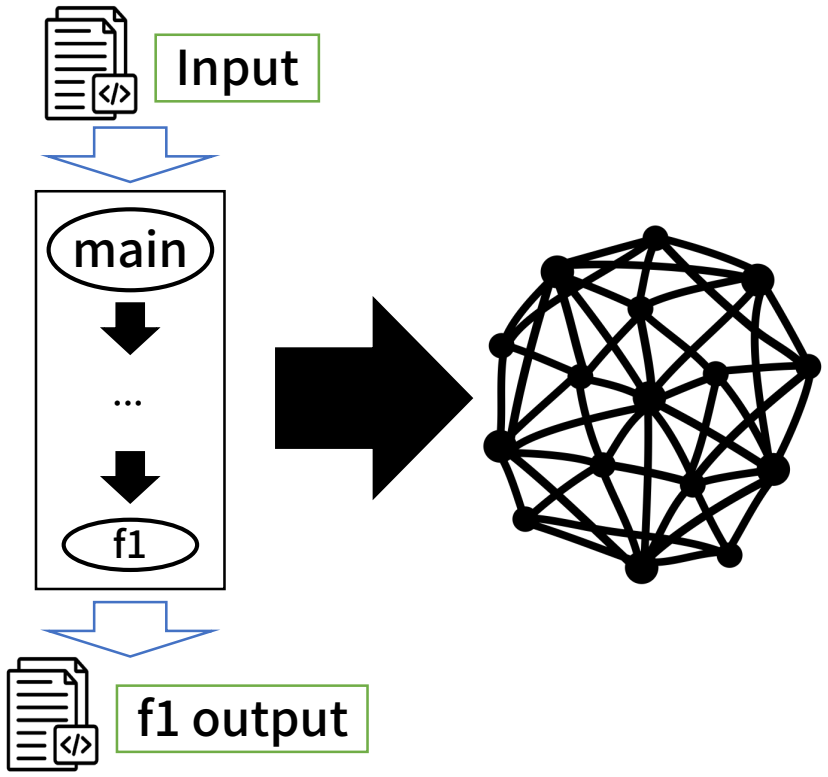
code

network

2단계 연구 계획: LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술

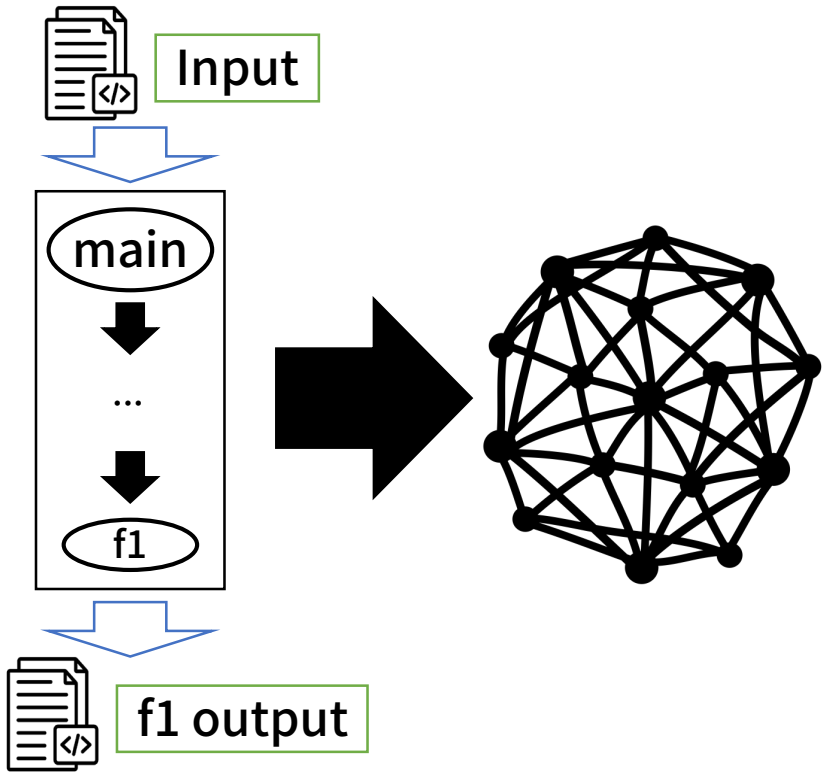


2단계 연구 계획: LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술



62.5% 확률로
같은 BB 커버리지를 가지는
다른 시스템 입력 생성 성공
(평균 BB cov. similarity: 82.7%)

2단계 연구 계획: LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술



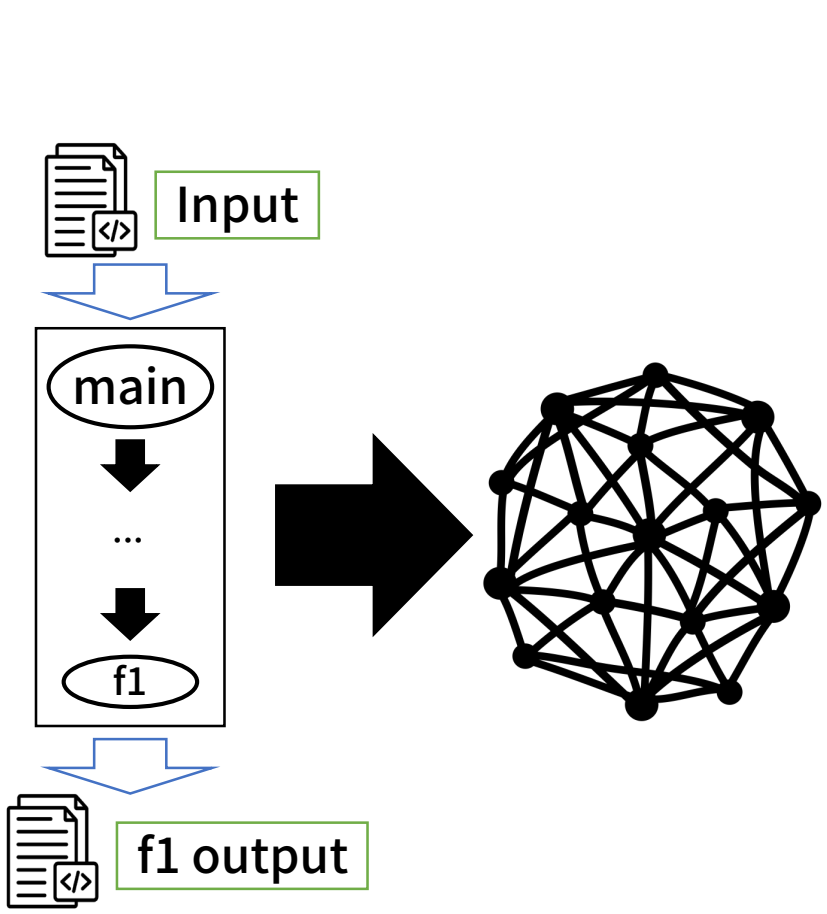
문제 1.

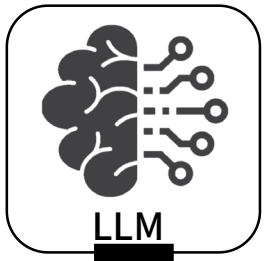
- 확인하고 싶은 지점이 너무 많음.
  - 그러나 모든 함수를 학습하기에는 너무 많은 비용

문제 2.

- 실행해본 적 없는 부분에 대해서는 **아무런 데이터를 얻을 수 없음**

2단계 연구 계획: LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술





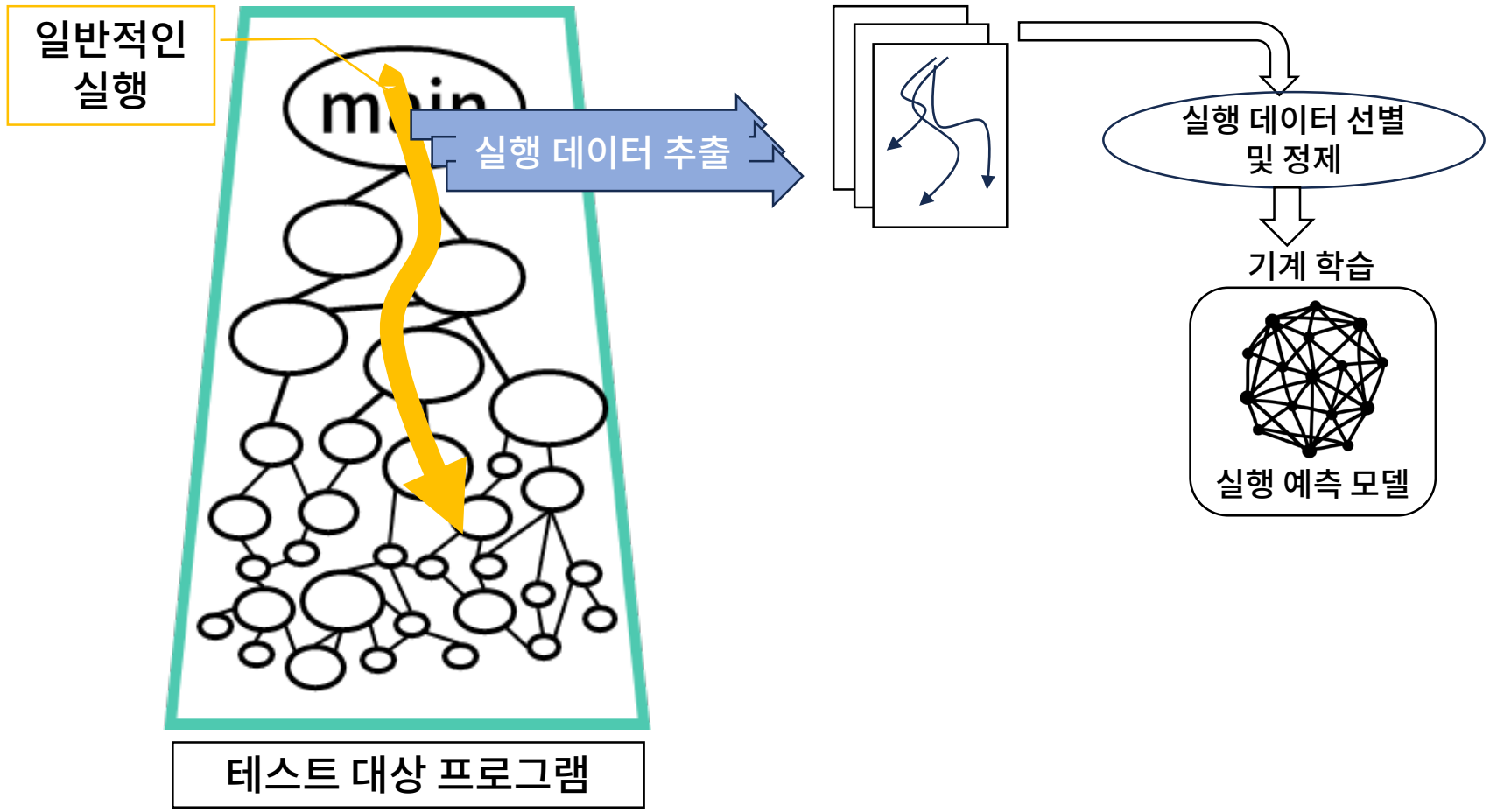
문제 1.

- 확인하고 싶은 지점이 너무 많음.
  - 그러나 모든 함수를 학습하기에는 너무 많은 비용
  - **LLM의 코드 추론 능력으로 분석이 필요한 함수를 취사 선택**

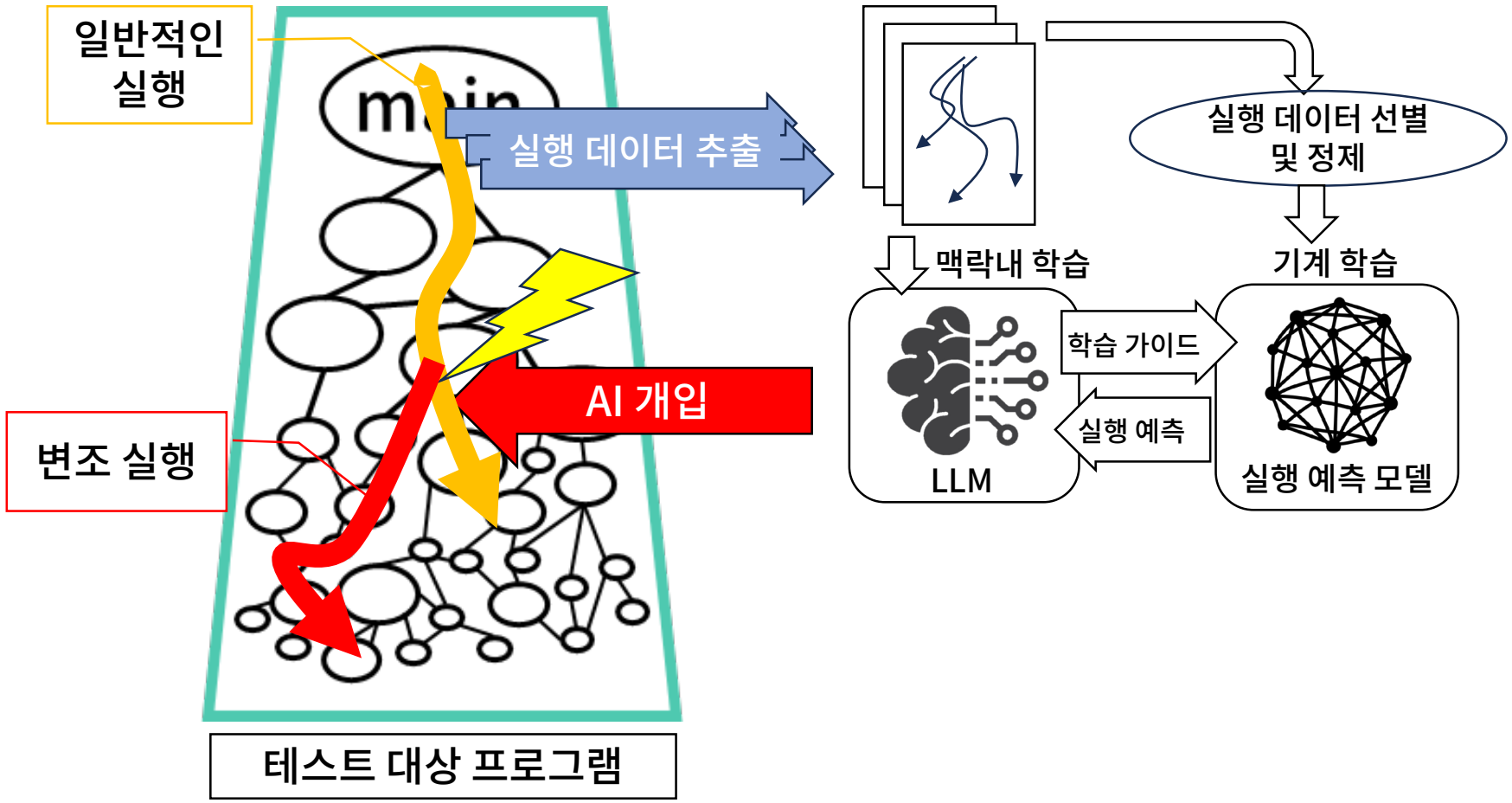
문제 2.

- 실행해본 적 없는 부분에 대해서는 아무런 데이터를 얻을 수 없음
  - **LLM이 지시 하는대로 “변조 실행”을 수행**

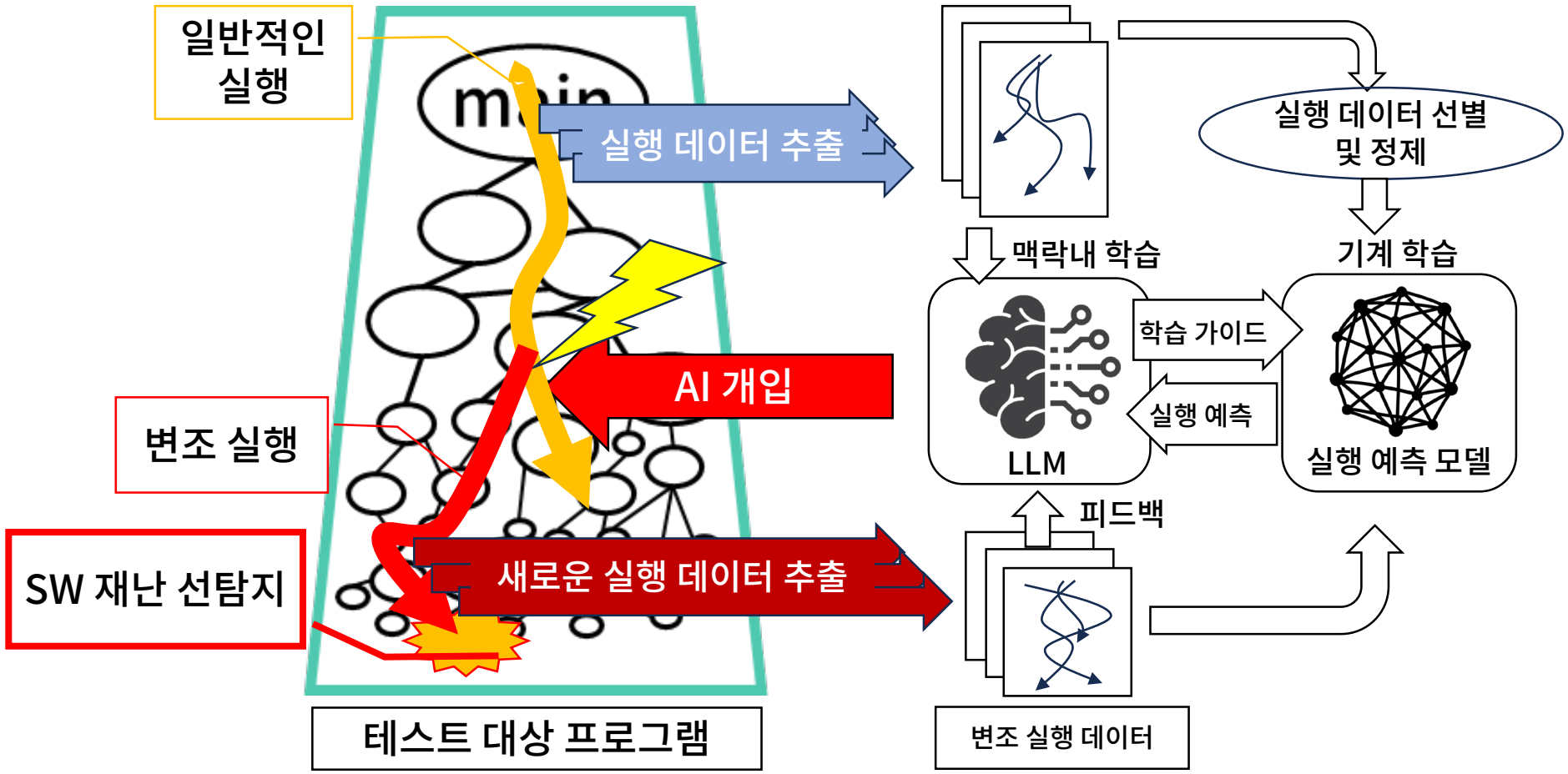
2단계 연구 계획: LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술



2단계 연구 계획: LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술



2단계 연구 계획: LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술



2단계 연구 계획: LLM 기반 학습 및 변조 실행을 통한 SW 재난 선탐지 기술

1. 이 기능을 OSAL, FreeRTOS 등의 실제 사회 infra 시스템에 적용하여 유효성 실증

2. Cross-project 적용
   - 한 프로젝트에서 학습한 실행 예측 모델을 다른 프로젝트에서 재활용하는 방법