DEOBFUSCATOR

문제 정의서

2025.04.05

컴퓨터융합학부 202002514 안상준
인공지능학과 202202487 박혜연
컴퓨터융합학부 202202602 손예진

Contents

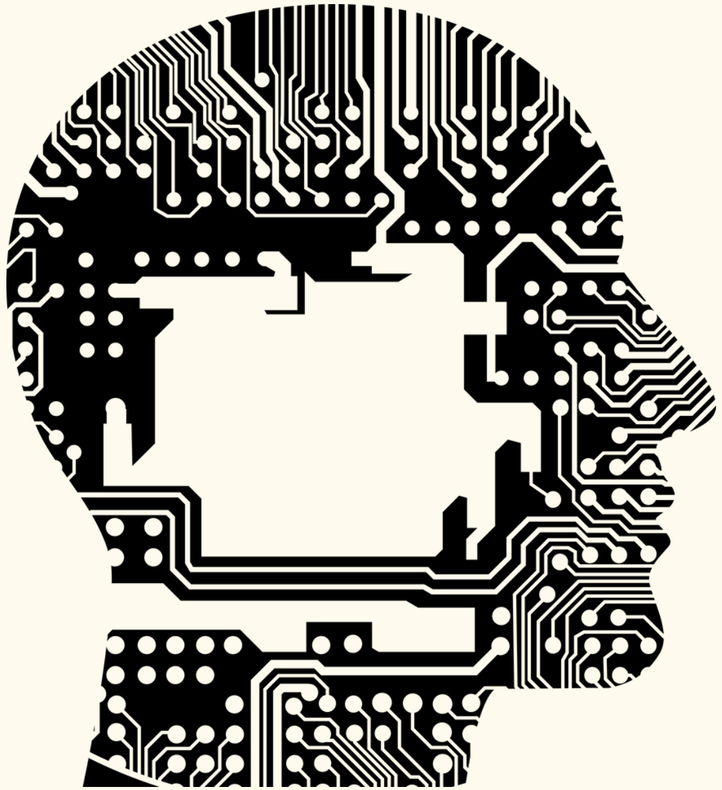
01 연구 개발의 필요성

기존 연구의 한계

- 특정한 난독화 도구나 기법에 최적화 되어있는 기존 역난독화 도구
- LLM을 역난독화에 사용하지 않은 이유 : 신뢰성 문제 및 데이터셋 문제

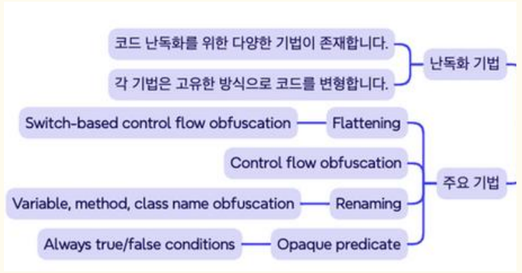
연구 개발의 개선 방향

- 데이터셋을 직접 구축하고 이를 LLM 학습에 사용
- 다양한 데이터셋 확보 → 플랫폼에 독립적으로 사용 가능한 도구



02 연구 개발의 목표 및 내용

LLM을 활용하여 가상화 난독화 된 코드를 역난독화 하는 것

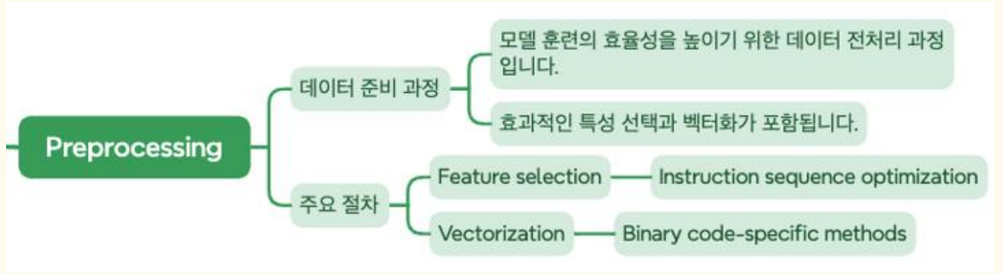


난독화 기법

flattening,
opaque predicate,
control flow, renaming 등

LLM Training

fine-tuning을 하거나
few-shot을 사용해
prompting을 하는 방식



Preproccessing

최대한의 성능을 얻기 위해
데이터셋에 feature selection과
max length 조절이 필요

Vectorize

기존 자연어 처리 방식이
아닌 바이너리 코드에 맞는
벡터화 방식을 사용

03 이해당사자 인터뷰 / 설문 인사이트

Q1 기존 역난독화 방식에 대한 불편한 점이 있습니까?

A1 기존 방식들은 특정 난독화 도구에 맞춰진 경우가 많아 약간의 변형이 있다면, 범용성이 낮다는 문제점이 있다.

A2 일반적으로 난독화 된 데이터를 바로 사용할 수 없어 전처리 및 단순화 작업을 직접 수행해야한다는 문제점이 있다.
또한, 이 작업이 올바르게 수행되었는지 확인하는 것도 어렵다.

Q2

LLM을 사용하여 역난독화하는 방식에 대해 어떻게 생각하십니까? 어떤 효과나 부작용을 예상하십니까?

A1

기존 방식보다 더 포괄적인 역난독화 가능하지만,
수학이나 알고리즘으로 검증하지 못해 성능에 대한 신뢰도가 낮을 수 있다.

A2

기존 방식보다 더 포괄적인 역난독화 가능하지만,
수학이나 알고리즘으로 검증하지 못해 성능에 대한 신뢰도가 낮을 수 있다.
LLM의 입력에는 제약이 존재하고, *Hallucination 현상을 우려한다

*Hallucination 현상 (환각 현상) : 인공지능 모델이 실제 데이터나 학습한 정보와 일치하지 않는 잘못된 결과를 생성하는 현상

Q3 어느 정도의 정확도가 보장되어야 사용할 의향이 있으십니까?

A1 50% 이상만 되어도 참고용으로 사용할 의향이 있다.

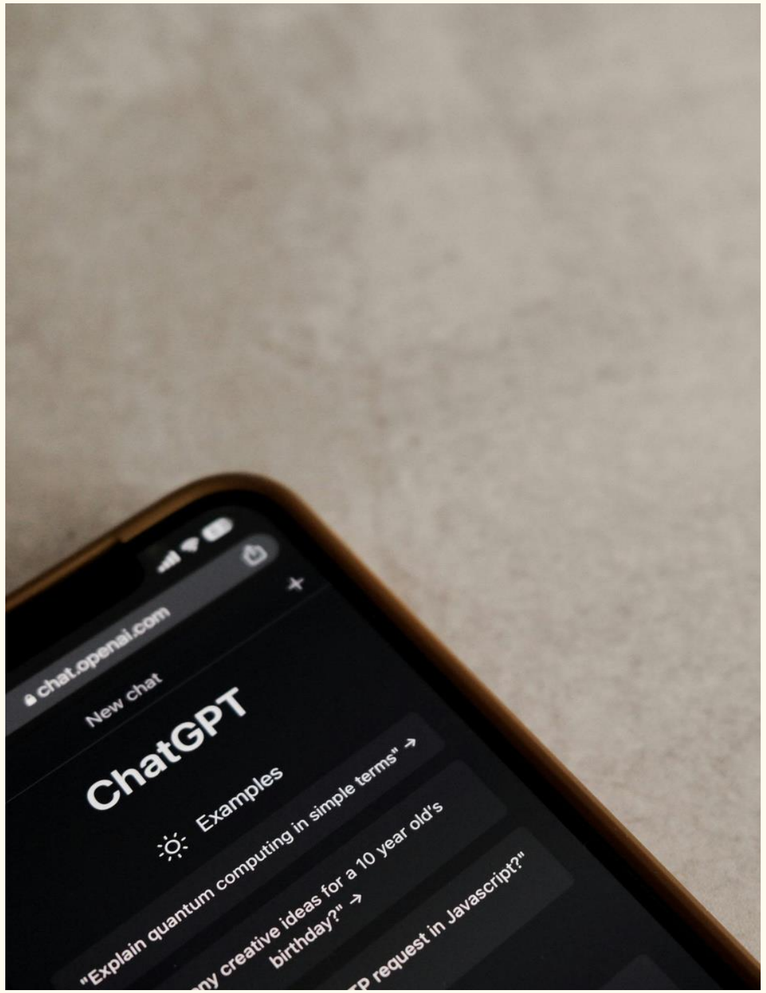
A2 평가 기준에 따라 다르겠지만, 분석가들이 잘못된 정보에 시간을 낭비하지 않도록 FP(False Positive)와 FN(False Negative) 값이 낮아지게 해야할 것이다.

Q4 추가적으로 제안하고 싶으신 방안이 있으십니까?

A1 특정 구조의 유무보다는 어느 부분에 나타나는지 시각화 해주는 도구

A2 코드의 크기가 커져도 쉽게 사용할 수 있으면 유용할 것 같다.

03 이해당사자 인터뷰 / 설문 인사이트



인터뷰 INSIGHT

난독화 방식의 여러 변형에 대응할 수 있는 도구가 필요

명확한 평가 기준 필요
FP, FN 을 고려하기 위한 precision, recall 을 포함

결과를 시각화 하기 위한
LLM 학습에 사용할 데이터 전처리 방안 고려

04 기대 효과 및 향후 확장 가능성

01

데이터셋을 직접 구축하고 LLM 학습에 사용
LLM의 신뢰성 및 데이터의 사용 가능성 입증

02

기존 LLM 의 바이너리 코드에 특화되어 있지 않아 낮았던 정확성에 대하여
바이너리 코드로 fine-tuning 하여 더 높은 정확도 획득 및 시간 비용 감소

05 연구 개발의 추진전략 및 방법

학습

바이너리 코드에서의 switch, if-else, while 구조의 패턴을 확인
난독화, 컴파일러, 프로그램 사용법 등 프로젝트에 필요한 기본 지식을 학습

분석

난독화를 적용하고 LLM에 입력하여 코드의 구조를 식별하는지 확인하는
실험을 진행, GitHub API를 사용한 데이터셋 확보

적용

확보한 데이터셋에 대한 처리 및 데이터셋 전처리 후에 LLM에 직접 학습을
진행하여 실험해보고 성능을 개선해 나갈 예정

06 참고문헌 (Reference)

MOHSENI, SEYEDREZA, ET AL. "CAN LLMS OBFUSCATE CODE? A SYSTEMATIC ANALYSIS OF LARGE LANGUAGE MODELS INTO ASSEMBLY CODE OBFUSCATION."ARXIV PREPRINT ARXIV:2412.16135(2024).

LI, XUEZIXIANG, YU QU, AND HENG YIN. "PALMTREE: LEARNING AN ASSEMBLY LANGUAGE MODEL FOR INSTRUCTION EMBEDDING."PROCEEDINGS OF THE 2021 ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY. 2021.

박성우, 박용수 "VMPROTECT의 역공학 방해 기능 분석 및 PIN을 이용한 우회 방안" 정보처리학회논문지. 컴퓨터 및 통신시스템 10.11 PP.297-304 (2021) : 297.