PROJECT DOCUMENT

Project Brainstorming Result

Project Name	바이너리 프로그램에서 제어구조를 식별하는 도구 개발

14 조

202002514 안상준
202202602 손예진
202202487 박혜연

지도교수: 조은선 교수님 (서명)

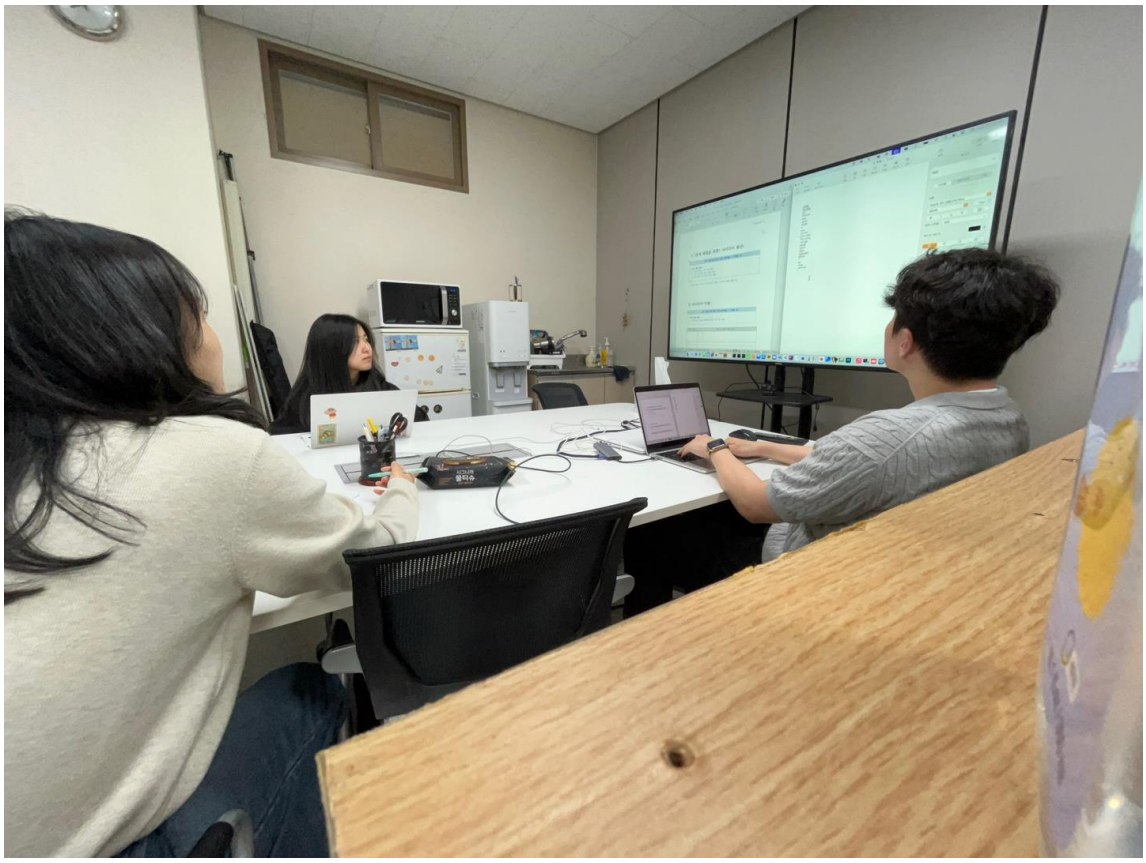
Document Revision History

Rev#	Date	Affected Section	Author
1	2025/3/25	문서 작성	안상준, 손예진, 박혜연

Contents

List of Figure

1.(문제 해결을 위한) 아이디어 발산



연구실에 함께 모여서 회의를 진행하였다. 아이디어 발산을 위하여 각자 연구 주제와 관련된 단어들을 얘기하고 적어 보았다. 주로 난독화에 관한 내용들이 많이 나왔으며, 그 다음 ai에 관련된 단어들이 많이 나왔다. 그 외에 컴파일러와 바이너리 구조에 대한 단어들이 나왔다. 나온 단어과 연관된 단어들을 찾아보며 llm모델의 종류나 버전, 컴파일러와 난독화 도구들의 종류나 옵션들에 대한 정보를 찾아보았다. 이후 찾은 단어들을 Chat Gpt에게 물어 추가적인 내용들을 찾아보았다.

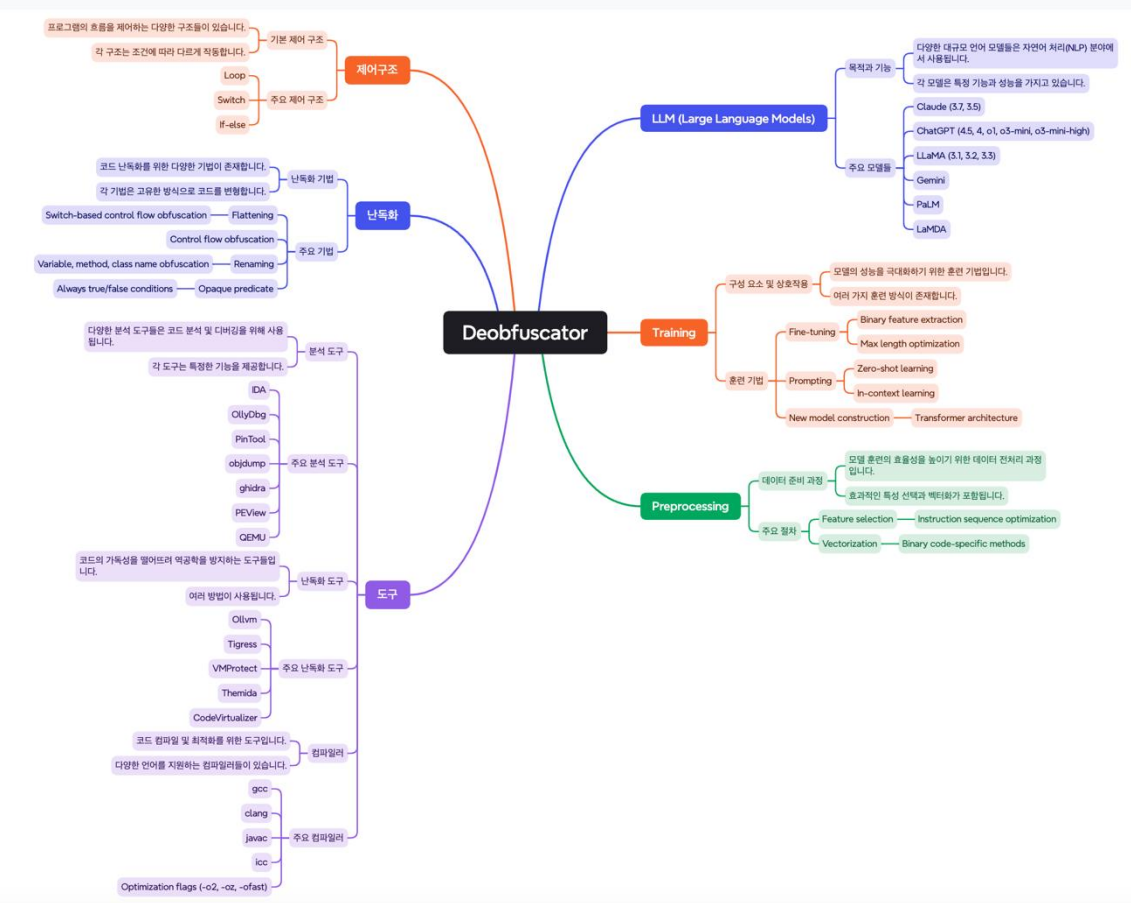
그 결과 약 50개 이상의 단어들이 나왔으며 이를 핵심 개념 4가지로 나누어 정리해 보았다. AI, 도구, 난독화, 제어구조 이렇게 나누어 보았고 각 항목의 하위 개념들을 나누어 정리해 보았다. 개념에 대한 추가적인 설명도 찾아서 정리해 보았다.

다양한 난독화 도구에 대해서 알아보았으며, 이를 통해 적용해볼 난독화 도구와 옵션에 대해서 생각볼 수 있었다. LLM 학습을 위한 방법을 정리해 보았고, 적용할 수 있는 모델에 대해 찾아보며, 어떤 모델 사용이 적합한지 앞으로 고민할 필요가 있을 것 같다.

2. 아이디어 수렴

핵심 개념	정의 및 하위 개념 목록
AI	1. LLM : claude, chat gpt claude version : 3.7, 3.5 chat gpt : 4.5, 4, o1, o3-mini, o3-mini-high LLaMA : 3.1, 3.2, 3.3 Gemini, PaLM, LaMDA 2. Training : fine tuning, prompting, 새로운 모델 구축 fine tuning : binary에서 feature를 추출하여 학습, max length를 조절하여 최적의 length탐색 prompting : ChatGpt, Claude를 이용하여, zero shot이나 in context learning을 사용 feature : Pin Tool을 이용하여 trace를 뽑아 feature로 사용 새로운 모델 구축 : transformer등과 같은 모델 구조를 사용해서 학습 3. Preprocessing : feature selection, vectorize feature selection : instruction sequence 에서 ai 학습에 불필요한 정보 제거로 모델의 효율성을 높임 vectorize : 기존 자연어 처리 방식을 사용하지 않고 바이너리 코드에 맞는 vectorize방식 사용
도구	1. 분석 : IDA, OllyDbg, PinTool, objdump, ghidra, PEView, QEMU 2. 난독화 : Ollvm, Tigress, VMProtect, Themida, CodeVirtualizer 3. 컴파일러 : gcc, clang, optimize, javac, icc, -o2, -oz, -ofast
난독화	1. 기법 : flattening, opaque predicate, control flow, renaming flattening : 함수의 기본 블록을 하나의 큰 switch 문 안에 배치하여 원래의 제어 흐름을 숨김 control flow : 제어 흐름 난독화는 프로그램의 실행 흐름을 복잡하게 만들어 코드 분석을 어렵게 renaming : 변수, 메서드, 클래스의 이름을 의미 없는 문자열로 변경하여 코드의 가독성을 낮춤 opaque predicate : 항상 참 또는 거짓으로 평가되는 조건문을 삽입하여 불필요한 분기를 만듦
제어구조	1. loop, switch, if-else

3. 시각화하기



마인드맵 생성 도구는 Mapify를 사용하였다. Ai를 이용하여 마인드맵을 생성해주는 도구로 아이디어 발산을 통해 도출해낸 단어들을 prompting을 통하여 마인드맵을 생성하였다. 앞서 직접 만들었던 상위개념을 5가지로 변경하였다. Ai를 좀 더 세분화 하여 LLM, Training, Preprocessing 이렇게 세 개의 상위 개념을 정의한 것을 볼 수 있었다.