

CodeTeacher: 생성형 AI 기반 맞춤형
Python 코드 학습 보조 시스템 개발

제출자 : 임 소 정

2215969/ 소프트웨어융합전공

1. 프로젝트 개요

1.1 프로젝트 개요

CodeTeacher는 Python을 배우는 초보 개발자, 컴퓨터공학 전공 학생, 그리고 코딩 튜터링을 받는 학습자들이 **복잡한 코드 구조를 쉽게 이해할 수 있도록 돕는 생성형 AI 기반 학습 보조 시스템**이다.

사용자는 웹 기반 앱에 Python 코드를 입력하면, 시스템은 자동으로 각 줄에 대한 **간단한 영어 주석, GPT 기반 한국어 설명**, 그리고 정확한 흐름을 시각적으로 표현한 **코드 흐름도(CFG 기반)**를 생성한다.



이 시스템은 AI가 코드의 구조와 의미를 학습자의 눈높이에 맞춰 설명해주는 개인화된 학습 경험을 제공하는 것을 목표로 한다. 사용자는 자신의 이해 수준에 맞춰 설명 스타일을 선택할 수 있으며, 설명 결과를 PDF로 저장하여 복습에도 활용할 수 있다.

1.2 문제 정의

본인을 비롯한 많은 학생들이 단순한 **코드 실행은 가능하지만, 코드의 구조와 흐름을 논리적으로 이해하지 못하는 문제**를 겪고 있다. 구체적으로, 다음과 같은 문제들을 겪고 있다.

- **코드의 각 줄의 역할을 해석하는 데 시간이 오래 걸린다.**
: 코드 한 줄 한 줄이 어떤 기능을 수행하는지 스스로 파악하기 어렵고 그에 사용된 문법 또한 확인해야 하므로 학습 속도가 느려진다.
- **기존의 학습 도구는 설명과 시각화의 깊이가 부족하다.**
: 대부분의 플랫폼은 코드 실행 결과만 보여주거나 어떠한 목적이 있을 때 코드를 자동으로 완성하도록 돕는 것이 대부분이라 동적인 흐름 이해 자체를 돕는 기능이 부족하다.
- **생성형 AI 도구를 활용하는 데에도 진입장벽이 존재한다.**
: ChatGPT나 유사한 도구들을 활용하더라도 효과적인 학습을 위해서는 프롬프트 설계나 반복적 활용법에 대한 이해가 필요하며 코드를 온전히 입력하기 위해 무수히 많은 질문을 입력해야 한다.

1.3 접근 방법

선택한 GenAI 모델	ChatGPT-3.5 와 CodeT5+
모델 선택 이유	<div data-bbox="518 425 922 461">① GPT - 3.5 (ChatGPT API)</div> <div data-bbox="1117 421 1366 506">   </div> <ul style="list-style-type: none"> ● 자연어 처리 기반의 대형 언어 모델로, 코드 요약 및 설명 생성에 적합하다 ● 단기 프로젝트 기준으로 비용이 비교적 저렴하며, 접근성이 높다 ● 다만, context window 제한(약 4K tokens) 으로 인해 긴 코드의 해석 정확도에 한계가 있다. <div data-bbox="518 880 1099 916">② CodeT5+ (HuggingFace 오픈소스 모델)</div> <ul style="list-style-type: none"> ● 코드 전용으로 사전학습된 모델로, 라인 단위 주석 생성에 특화되었다 ● GPT의 입력 한계를 보완하고, 코드 구조를 요약해주는 역할로 활용 가능하다 ● 오픈소스로 제공되며, 추가 비용 없이 로컬에서 자유롭게 사용 가능하다
적용 전략	<p>총 2번의 분석을 거치는 모델 조합 구조를 만들고자 한다.</p> <ul style="list-style-type: none"> ● 1차 분석 (CodeT5+) <ul style="list-style-type: none"> 사용자가 입력한 코드를 라인별로 요약 및 주석화하여 코드 흐름의 구조를 미리 파악한다 ● 2차 분석 (GPT-3.5) <ul style="list-style-type: none"> CodeT5+가 생성한 주석과 원본 코드를 함께 GPT에 입력하여, 더 정확하고 자연스러운 한국어 코드 설명 생성한다 <p>이러한 모델 구조를 통해 GPT 단독 사용 시 발생할 수 있는 맥락 손실과 비용 문제를 보완하고, 동시에 사용자가 이해하기 쉬운 설명 + 해석 정확도를 실현하고자 한다.</p>

2. 프로젝트 내용

2.1 기술개발 목표

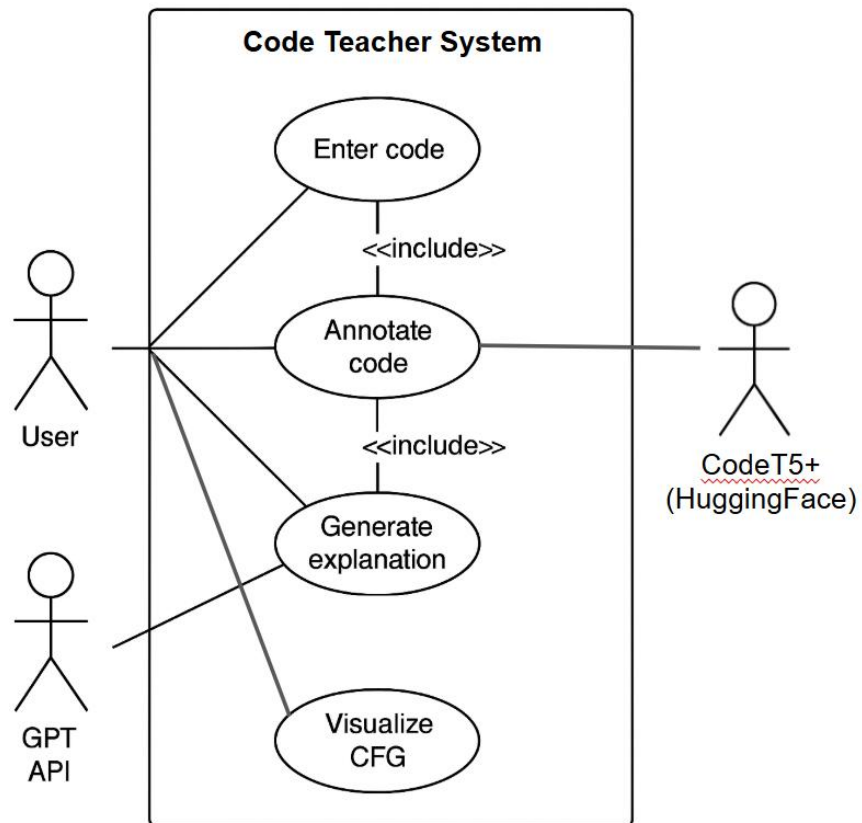
최종 목표	Generative AI 및 Model을 활용하여 사용자가 Python 학습을 용이하도록 하는 시스템 개발하기
세부 목표	<ul style="list-style-type: none">● 코드의 각 줄이 수행하는 역할을 AI가 자동으로 설명해줄 수 있어야 한다.● 복잡한 코드의 흐름과 구조를 정확하고 시각적으로 표현하여 학생들이 직관적으로 이해할 수 있게 해야한다.● 학습자의 이해 수준에 따라 설명 스타일을 선택할 수 있어야 한다.● 설명 결과는 언제든지 복습할 수 있도록 PDF로 제공할 수 있어야 한다.● 별도의 소프트웨어 설치 없이 웹에서 쉽게 사용할 수 있어야 한다.

2.2 개발 내용

개발 환경 및 스택	<p>① 개발 환경</p> <ul style="list-style-type: none">● 구현 플랫폼 : 로컬 PC (NVIDIA GPU 미탑재 환경)● 개발 관리: Github 저장소 (https://github.com/Sophia680102/CodeTeacher) <p>본 프로젝트에서는 클라우드 기반 모델인 CodeT5+ 와 GPT-3.5 API를 활용하되, 로컬 환경에서도 충분히 실행 가능한 구조를 목표로 한다.</p> <p>위 AI 모델들은 외부 API 호출을 통해 동작하며, 모델 자체의 로딩이나 실행은 로컬에서 직접 수행하지 않기 때문에 고성능 GPU 없이도 시스템 구동이 가능하다.</p> <p>또한, CFG 시각화 및 Streamlit 기반 웹 앱 실행 역시 GPU 의존도가 낮아, 전체 시스템은 CPU 기반의 일반적인 로컬 환경에서도 안정적으로 작동할 것으로 예상된다.</p>
------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	② 사용 기술 스택 (대표적인 것만) :																				
	<table><tr><th>항목</th><th colspan="2">도구/ 라이브러리</th></tr><tr><td>언어</td><td colspan="2">Python</td></tr><tr><td>프레임워크</td><td colspan="2">Streamlit (웹 인터페이스)</td></tr><tr><td>AI 모델</td><td colspan="2">CodeT5+ , OpenAI GPT-3.5</td></tr><tr><td>시각화</td><td colspan="2">pycfg (Python Control Flow Graph 생성기)</td></tr><tr><td>PDF 생성</td><td colspan="2">reportlab 또는 fpdf 라이브러리</td></tr></table>			항목	도구/ 라이브러리		언어	Python		프레임워크	Streamlit (웹 인터페이스)		AI 모델	CodeT5+ , OpenAI GPT-3.5		시각화	pycfg (Python Control Flow Graph 생성기)		PDF 생성	reportlab 또는 fpdf 라이브러리	
	항목	도구/ 라이브러리																			
	언어	Python																			
	프레임워크	Streamlit (웹 인터페이스)																			
	AI 모델	CodeT5+ , OpenAI GPT-3.5																			
	시각화	pycfg (Python Control Flow Graph 생성기)																			
PDF 생성	reportlab 또는 fpdf 라이브러리																				
주요 기능	시스템의 주요 기능은 다음과 같이 설정했다.																				
	순서	주요 기능	설명																		
	1	사용자 코드 입력 및 설명 스타일 선택	사용자가 Python 코드를 입력 및 GPT 설명 스타일 선택 가능. - {학생용/개발자용/튜터스타일}																		
	2	라인별 주석 생성	CodeT5+ 모델이 각 줄마다 간단한 영어 주석 생성																		
	3	GPT 설명 생성	GPT-3.5가 전체 코드와 CodeT5+ 주석을 참고하여 자연스러운 한국어 설명 제공																		
	4	코드 흐름도 시각화	Control Flow Graph (CFG)를 활용한 코드 흐름도를 생성하여 PNG 로 저장																		
	5	PDF 내보내기 기능	GPT가 생성한 한국어 설명을 PDF로 저장 가능																		
	6	Streamlit 기반 웹 UI	전체 기능을 Streamlit으로 구성하여 직관적이고 반응 빠른 인터페이스 제공																		

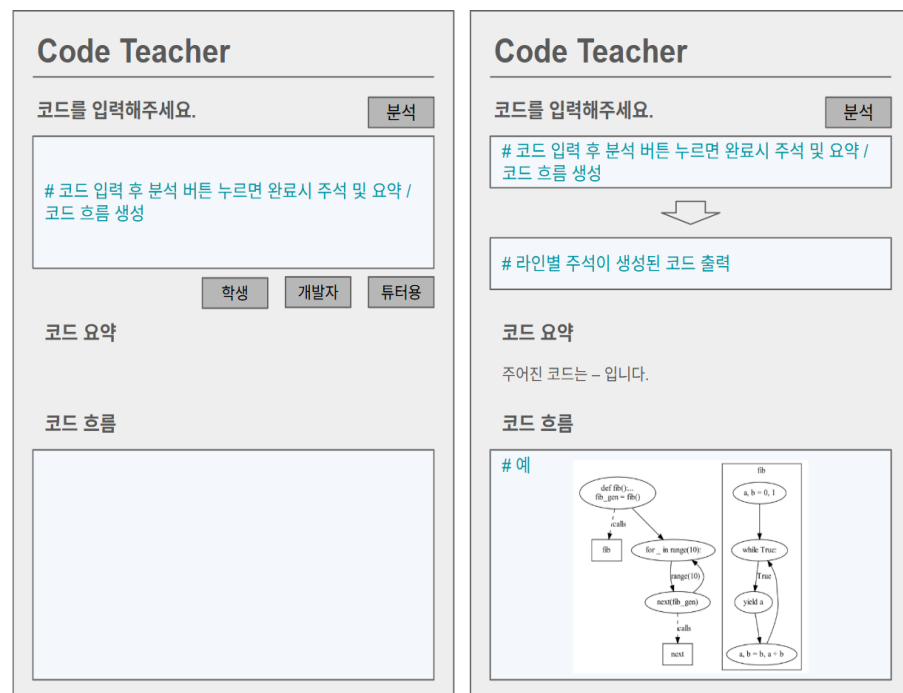
다음은 본 시스템의 핵심 기능을 요약한 UML 다이어그램이다.



[그림 1. CodeTeacher UML]

예상 결과

최종 Streamlit 화면 구성은 다음과 같이 하고자 한다.



[그림 2. CodeTeacher UI]

2.3 추진 일정¹

Timeline	5 - week 1	5 - week 2	5 - week 3	5 - week 4
기능 구현(1)	기능별 모듈 최소 실행			
기능 구현(2)		세부 기능 보완		
UI 강화			UI 다듬고 사용성 높이기	

본 프로젝트는 ChatGPT와 같은 생성형 AI를 적극 활용함으로써, 훨씬 높은 개발 생산성과 자동화 수준을 기대할 수 있다. 따라서, 핵심 기능 구현에 소요되는 시간은 비교적 짧은 것으로 예상된다.

그러나, 본 프로젝트는 로컬 PC 환경에서 실행되며, API 연동, 모델 호출, 파일 처리 등 다양한 컴포넌트를 포함하기 때문에 예기치 못한 오류나 환경 의존적인 이슈 발생 가능성도 존재한다.

또한, 개발자(본인)의 현재 프로그래밍 역량과 코드 해석 능력이 완전하지 않음을 고려하여 각 단계별 충분한 완충 시간과 점진적 개발 흐름을 반영하여 위와 같이 계획을 수립하였다.

¹ 일정에 여유가 있다면 chatbot 기능을 추가 구현하고자 함.

3. 기대효과 및 확장성

3.1 기대효과

- 줄 단위 주석 + 자연어 설명 + 흐름도 시각화를 결합하여, 초보자도 코드의 실행 흐름과 핵심 로직을 보다 쉽게 이해할 수 있어 학습 효율이 높아질 것이다.
- GPT 설명 스타일 선택 기능을 통해 학습자의 수준이나 학습 목적에 맞는 설명 방식 제공이 가능하다.
- 스스로 코드를 작성하고 시스템의 설명과 비교하면서 학습자의 메타인지 향상을 기대할 수 있다.
- GPU가 없는 일반 로컬 PC에서도 충분히 작동 가능하도록 시스템을 구축하여 누구나 본 프로젝트를 오픈소스로서 활용할 수 있게 된다.

3.2 확장성

- 대화형 챗봇 기능 추가 : GPT를 기반으로 코드 관련 질문에 실시간 답변하는 챗봇 기능을 추가할 수 있다.
- Python 외에 Java, JavaScript 등 타 언어로도 분석 기능을 확장할 수 있다.
- 사용자의 코드 실력/ 이해도 분석 후, 학습 수준에 따른 개인화된 피드백 제공 시스템을 구축하여 발전시킬 수 있다.
- 교수자용 모드를 추가하여, 학생 코드 자동 분석 및 과제 피드백 시스템으로 발전시킬 수 있다.
- 코딩 교육 웹사이트 등과 연계하여 실습 중심 교육 환경 구현이 가능해질 수 있다.

4. 참고 문헌

논문	CodeT5+: Open Code Large Language Models for Code Understanding and Generation (https://arxiv.org/abs/2305.07922)
----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------