Brendan Cowan, 1st Year Master's Student
Watanobe Lab., The University of Aizu

Poster Session at Graduate School Information Fair
Enhancing Programming Learning with LLMs: Prompt Engineering and Flipped Interaction
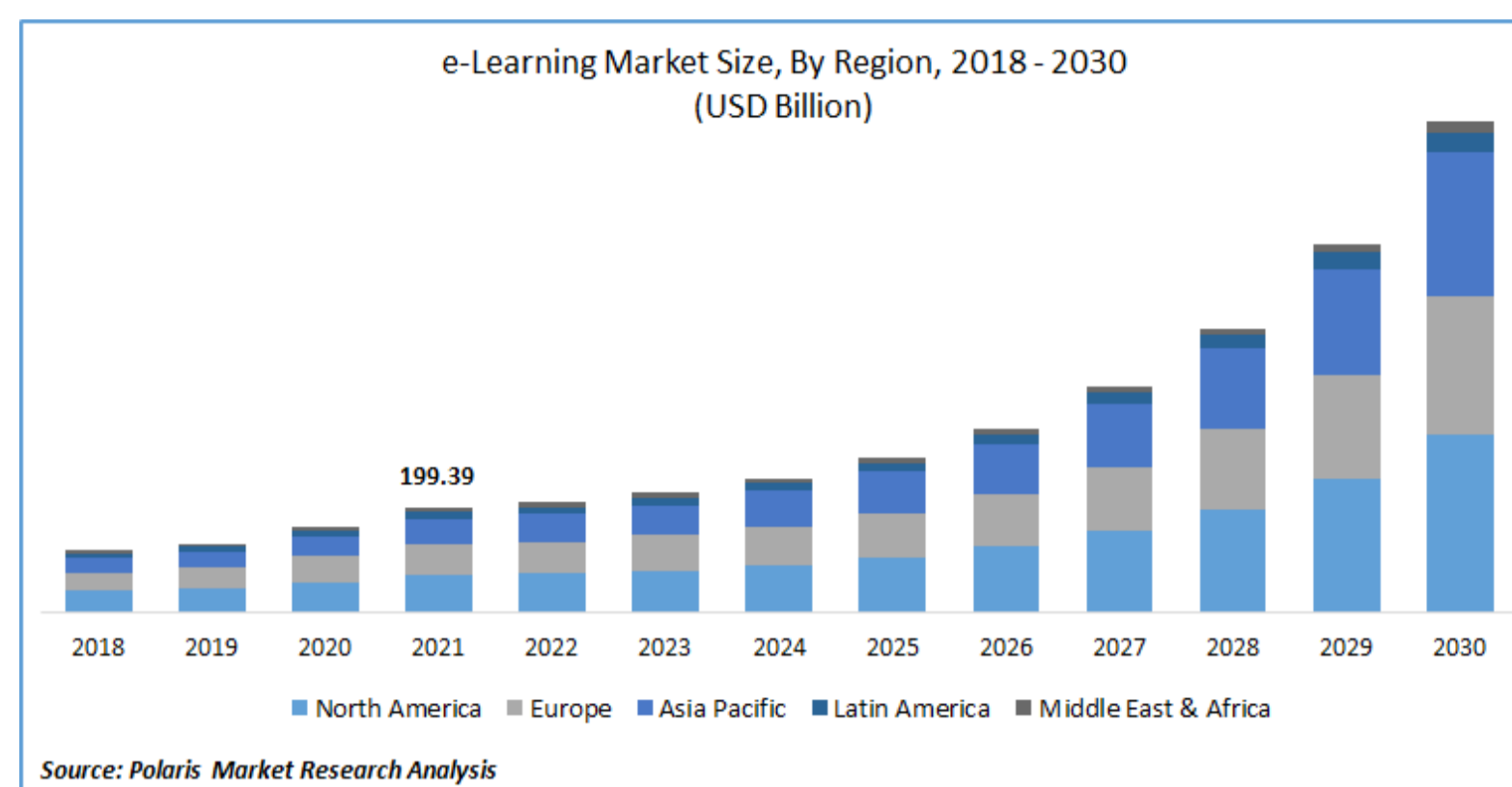
Background

Motivation

Education is becoming more and more virtual, with technologies such as eLearning and remote lectures increasing in popularity. However, virtual study poses the challenge of obtaining good, constructive feedback. Instructors can be busy or slow to answer, and some students may be shy to ask questions.
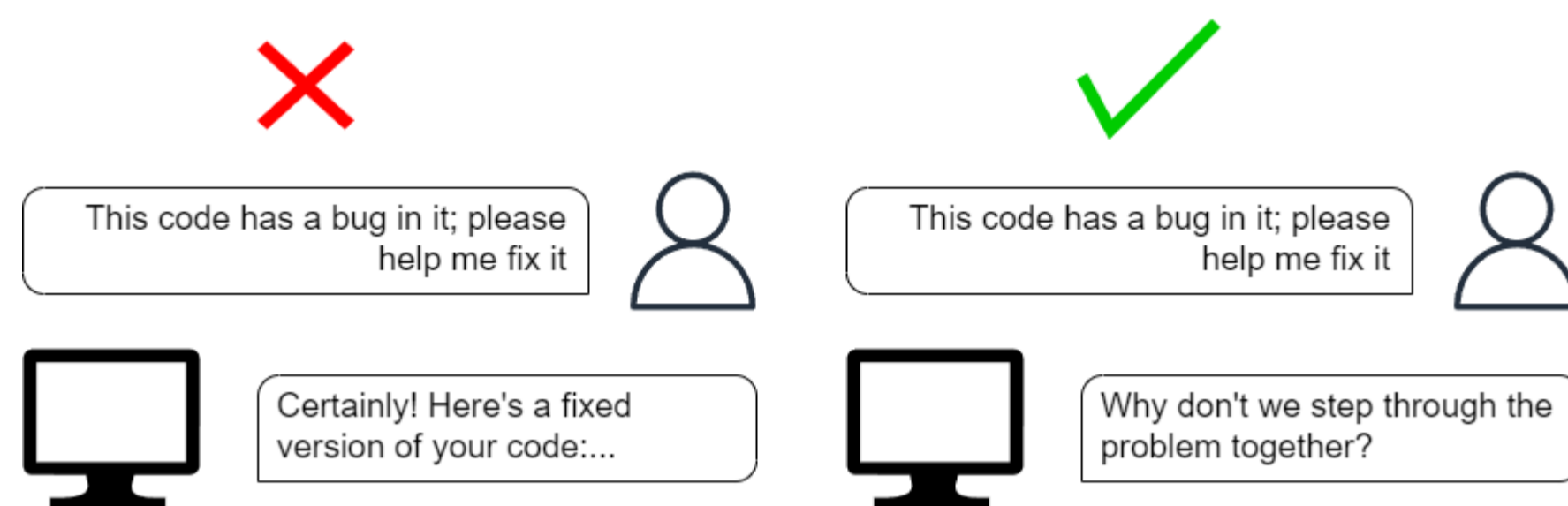
What about ChatGPT?

Recent advances in artificial intelligence and Large Language Models (LLMs) can give students a way to ask for feedback and get immediate results. However, this method may be too easy for students: while an instructor may help a student to learn, an LLM is likely to simply give student raw answers.
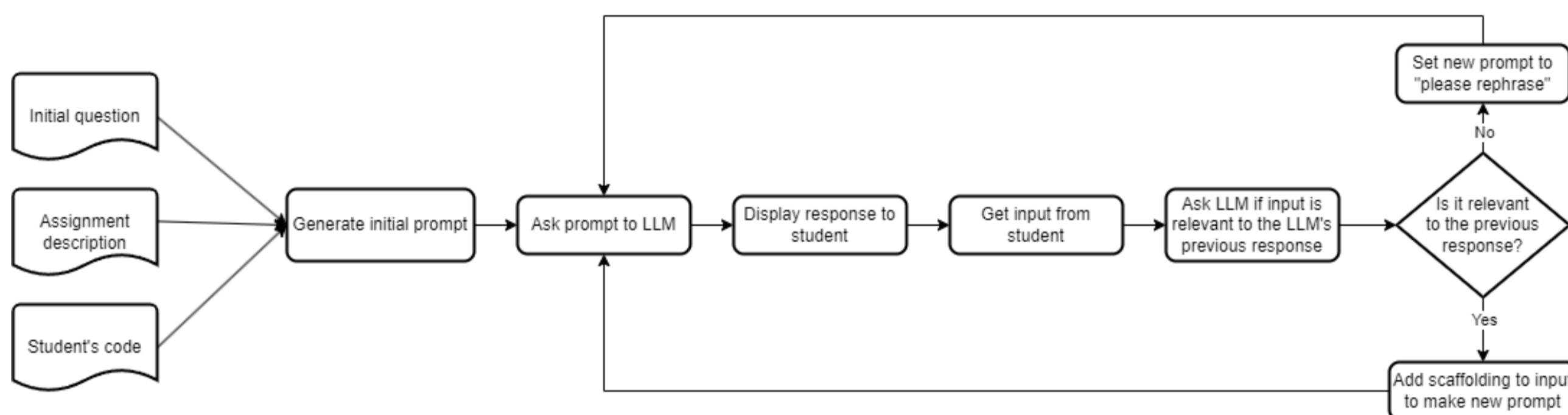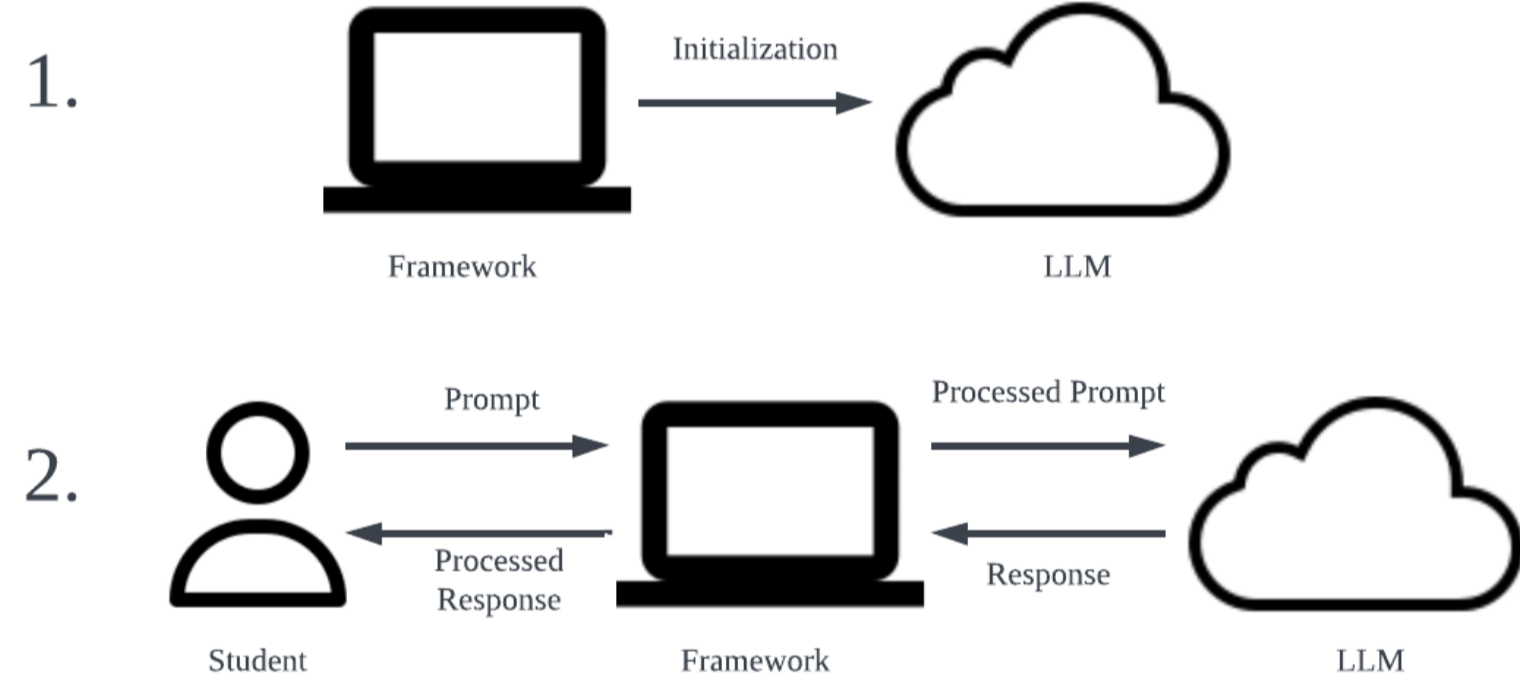
Goal

The goal of this research is to develop and test a way to use LLMs to assist students with their programming assignments in a way that is conducive to their learning.

e-Learning Market Size, By Region, 2018 - 2030 (USD Billion)

199.39

North America | Europe | Asia Pacific | Latin America | Middle East & Africa

Source: Polaris Market Research Analysis

https://www.polarismarketresearch.com/industry-analysis/e-learning-market

❌ This code has a bug in it; please help me fix it
Certainly! Here's a fixed version of your code:...

✅ This code has a bug in it; please help me fix it
Why don't we step through the problem together?

Methodology

1.
Framework → Initialization → LLM

2.
Student → Prompt → Framework → Processed Prompt → LLM
LLM → Response → Framework → Processed Response → Student

Flowchart: Initial question / Assignment description / Student's code → Generate initial prompt → Ask prompt to LLM → Display response to student → Get input from student → Ask LLM if input is relevant to the LLM's previous response → Is it relevant to the previous response? → No: Set new prompt to "please rephrase" / Yes: Add scaffolding to input to make new prompt
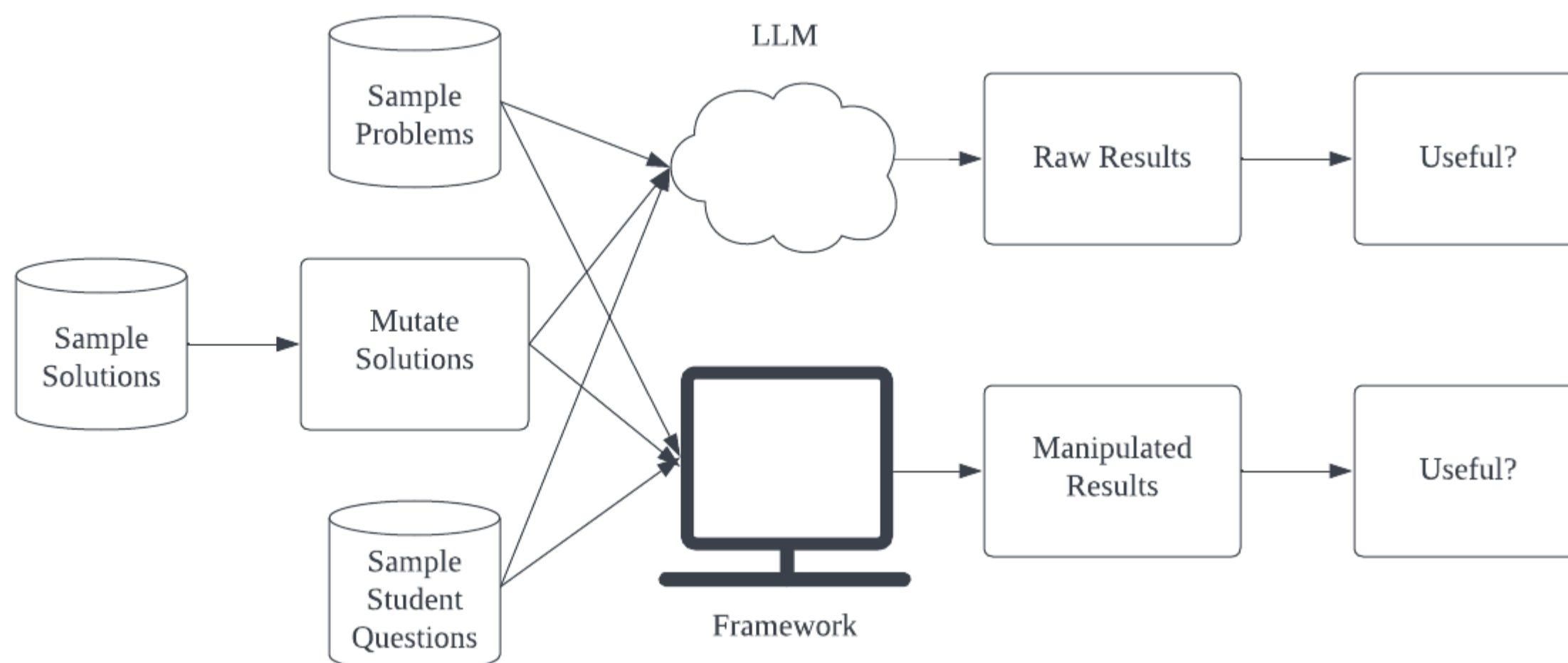
We designed a framework intended to be implemented as an IDE plugin. The student initializes a conversation with the framework from within the IDE, then their code, an assignment description, and an initial prompt from the student are sent to the framework. The framework communicates with an LLM and outputs a refined response to the student.

Results

Experiment Setup

We took sample problems and solutions from geeksforgeeks.com, mutated the solutions so that they mimicked what a student might write, and fed those into both the framework and raw LLM. We then used ChatGPT to generate responses for the student's side of a conversation with the LLM or Framework, and compared the framework's responses to the raw LLM's responses based on two metrics.

Sample Problems → LLM → Raw Results → Useful?
Sample Solutions → Mutate Solutions
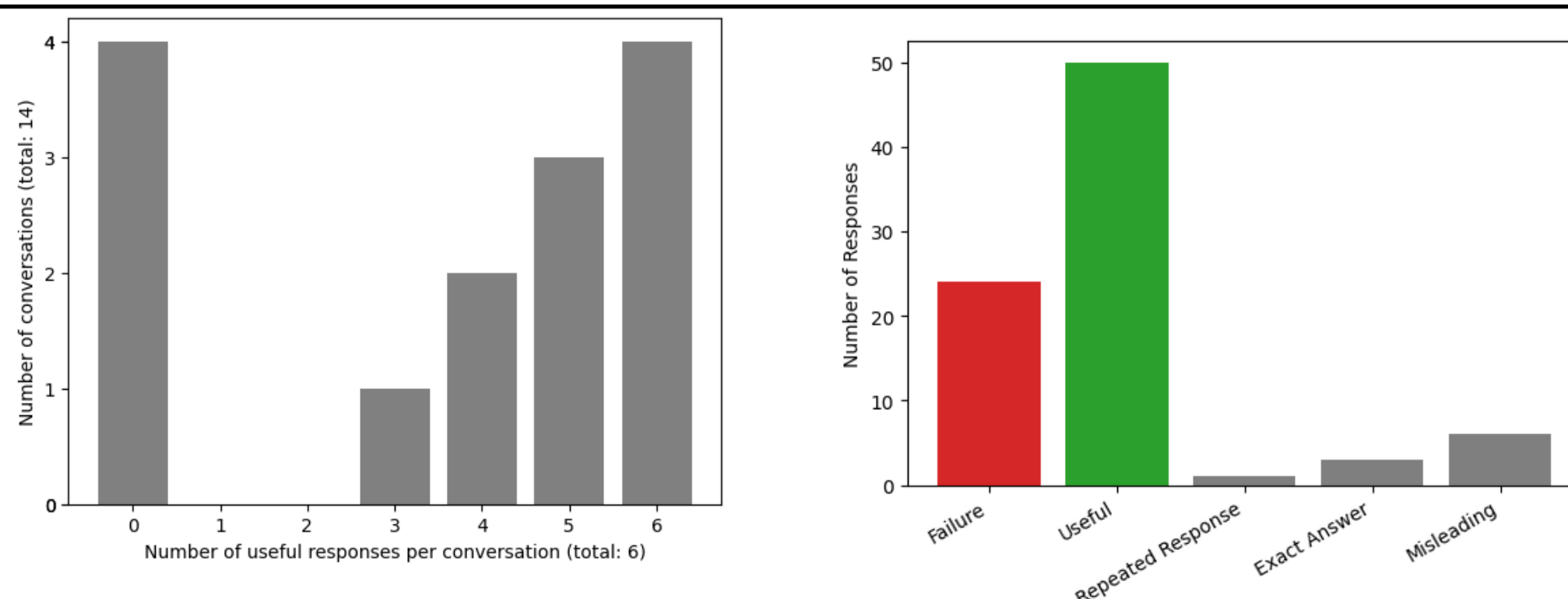Sample Student Questions → Framework → Manipulated Results → Useful?

Overall Conversation

We measured the perceived usefulness of overall conversations. A conversation was considered a pass if it appears to be pushing the student in the right direction without giving any exact answers. In the table below, the Framework Case refers to conversations with our framework, and the Raw Case refers to conversations with LLM alone.

	Passes	Fails
Framework Case	10	4
Raw Case	0	14

Individual responses

We measured the perceived usefulness of individual responses. Here we categorize each response from the framework as Failure (appeared in a failed conversation from the other experiment), Useful (could push the student in the right direction), Repeated Response, Exact Answer, and Misleading.

Brendan Cowan, 1st Year Master's Student
Watanobe Lab., The University of Aizu

Poster ID: 09

Poster Session at Graduate School Information Fair
Enhancing Programming Learning with LLMs:
Prompt Engineering and Flipped Interaction
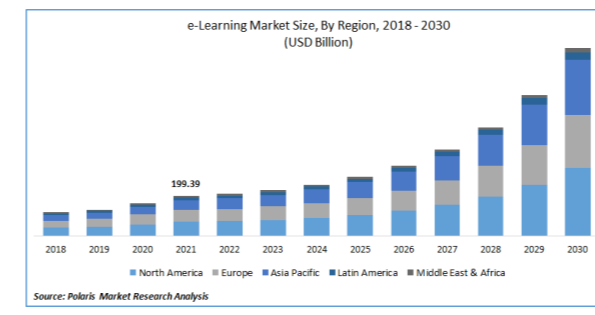
Background

Motivation

Education is becoming more and more virtual, with technologies such as eLearning and remote lectures increasing in popularity. However, virtual study poses the challenge of obtaining good, constructive feedback. Instructors can be busy or slow to answer, and some students may be shy to ask questions.
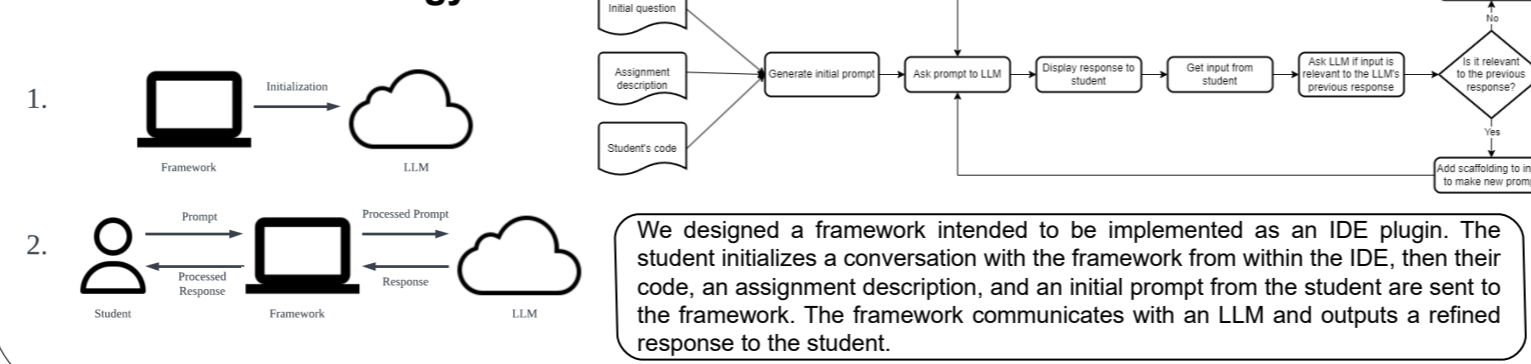
What about ChatGPT?

Recent advances in artificial intelligence and Large Language Models (LLMs) can give students a way to ask for feedback and get immediate results. However, this method may be too easy for students: while an instructor may help a student to learn, an LLM is likely to simply give student raw answers.

Goal

The goal of this research is to develop and test a way to use LLMs to assist students with their programming assignments in a way that is conducive to their learning.
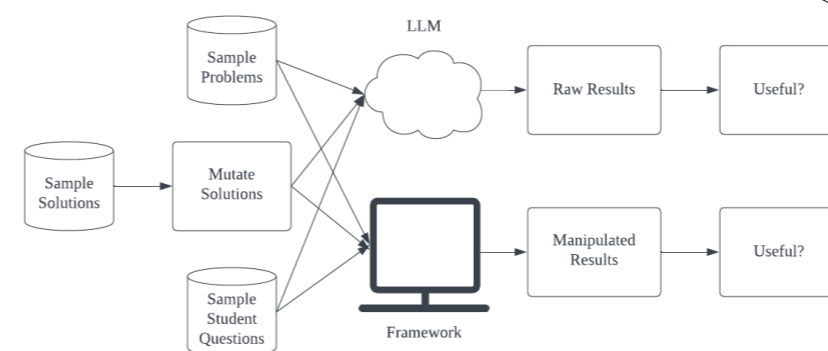

e-Learning Market Size, By Region, 2018 - 2030 (USD Billion)

https://www.polarismarketresearch.com/industry-analysis/e-learning-market

❌ This code has a bug in it, please help me fix it
Certainly! Here's a fixed version of your code:...

✅ This code has a bug in it, please help me fix it
Why don't we step through the problem together?

Methodology

1.

2.

We designed a framework intended to be implemented as an IDE plugin. The student initializes a conversation with the framework from within the IDE, then their code, an assignment description, and an initial prompt from the student are sent to the framework. The framework communicates with an LLM and outputs a refined response to the student.

Results

Experiment Setup

We took sample problems and solutions from geeksforgeeks.com, mutated the solutions so that they mimicked what a student might write, and fed those into both the framework and raw LLM. We then used ChatGPT to generate responses for the student's side of a conversation with the LLM or Framework, and compared the framework's responses to the raw LLM's responses based on two metrics.

Overall Conversation

We measured the perceived usefulness of overall conversations. A conversation was considered a pass if it appears to be pushing the student in the right direction without giving any exact answers. In the table below, the Framework Case refers to conversations with our framework, and the Raw Case refers to conversations with LLM alone.

	Passes	Fails
Framework Case	10	4
Raw Case	0	14

Individual responses

We measured the perceived usefulness of individual responses. Here we categorize each response from the framework as Failure (appeared in a failed conversation from the other experiment), Useful (could push the student in the right direction), Repeated Response, Exact Answer, and Misleading.