Technical Interview Project: Exemplar Answer Generation with OpenAI API

CURA

Background

Cura Education hosts an extensive library of online courses, each containing multiple tasks designed for student engagement and learning. Within these tasks, students are required to answer various questions and submit their responses. Providing exemplar answers is crucial for teachers, as it sets a clear benchmark for evaluating student submissions. To assist teachers in assessing student work more effectively, Cura aims to generate exemplar answers for each question, offering a reliable reference point for grading and feedback.

Project Overview

This project aims to integrate the OpenAI API to generate exemplar answers for questions. Each input includes context of student task content, question, and assessment rubrics. The goal is to automatically generate high-quality answers based on provided input and ensure they align with the rubrics used for evaluation.

Example student task content(partial):

How does our digestive system work?

Each part of the digestive system helps break down and move food through the **gastrointestinal tract** (or digestive tract) which leads from the mouth to the anus.

The mouth

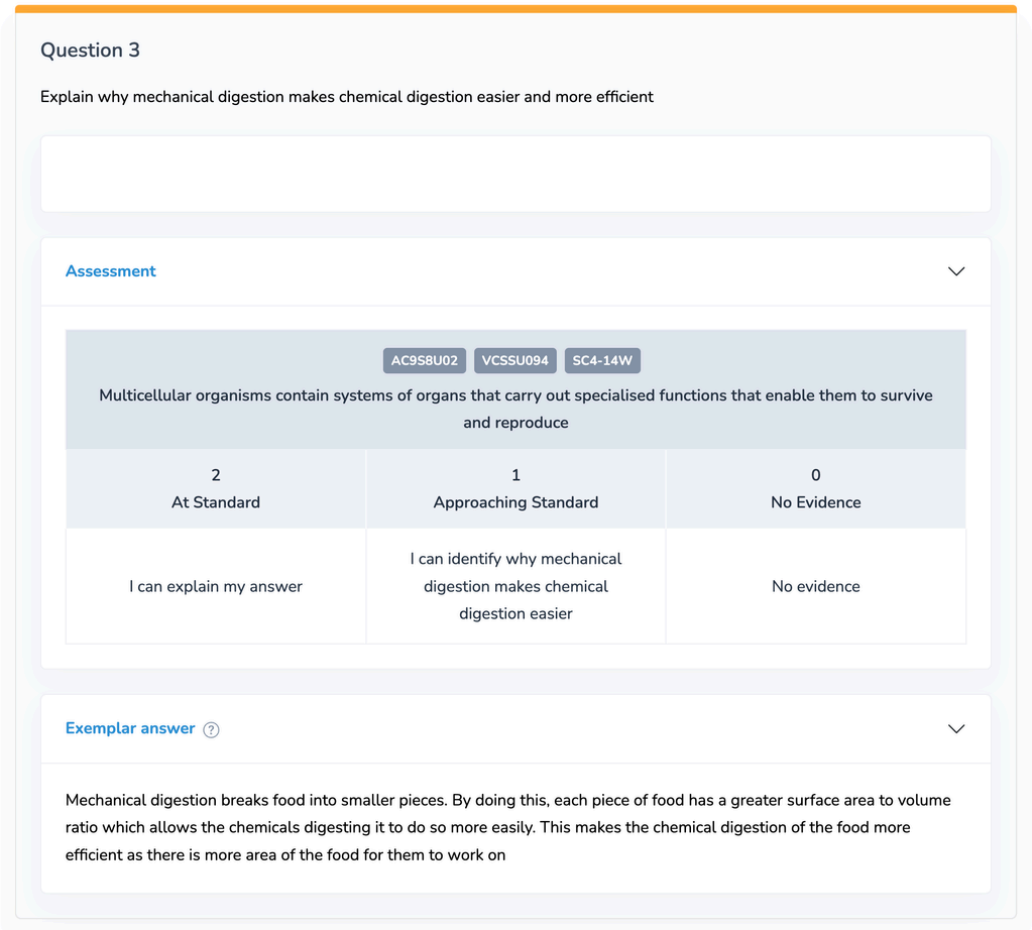
Our mouth is responsible for physically and chemically breaking down food. Physically, our teeth, jaw, and tongue break down food into smaller pieces through biting, grinding, and chewing. This is called **mechanical digestion**.

Salivary glands make saliva to help make the food slippery. Saliva ensures food can be easily moved around our mouth and swallowed. Saliva also contains **amylase** which is an **enzyme** that chemically breaks down carbohydrates (like bread) into simple sugars (like glucose). By breaking down our food into smaller pieces through mechanical digestion, the enzymes in chemical digestion can act on this food more easily.

The oesophagus

The oesophagus is a muscular tube that transports food and drink from the mouth to the stomach. When we swallow food, these muscles squeeze the food down through a process called **peristalsis**. This massaging of food down your oesophagus is so powerful that you can even swallow if you are hanging upside down.

Example question, assessment rubrics, and desired exemplar answer:



Technical Requirements

1. **Data Source**: You are provided with a set of 100+ rows of **training data**, including task content, questions, rubrics and exemplar answers.
2. **Core Objective**: Use the provided training data to develop a Python solution that integrates with the **OpenAI API** (GPT-4o mini). You goal is to train the LLM to generate exemplar answers for questions.
    a. Input:
        i. Task content as context
        ii. Question
        iii. Assessment Rubrics
    b. Output:
        i. Exemplar answer
3. **Result Evaluation**:
    a. Provide some form of evaluation for the generated answers against the sample training data. (e.g. k fold cross validation or other methods)

Bonus:

1. **Evaluation**:
    a. Analyse the evaluation results, e.g. how good is the exemplar answer quality of the LLM model you have trained. We would like to see your thought process in this step.
    b. Some creative ways to evaluate out-of-sample or unseen data. This will help generating new exemplar answers in the future.
2. **Python Testing**: Include automated tests for your code (e.g., using pytest) to demonstrate that it runs correctly and can handle potential edge cases.

Submission Guidelines

1. Submit the project as a **public GitHub repository**.
2. Include a file to briefly explain:

- The overall structure of your solution
- How to run the code
- Any testing or evaluation logic included in the project

3. Ensure that the project can be run without external configurations (e.g., include requirements.txt for easy setup).

Please email your submission to jun@curaeducation.com

Time Expectation

You should not spend more than **1-2 days** working on this project.

Don't hesitate to email us if you have any questions.

Good luck, and we look forward to reviewing your submission!

Appendix

Training Data Structure

Example:


```
[{
        "question_id" : "8b1d6474-a38a-4456-8ac2-74ff054b9d67",
        "task_id" : "41f89fd8-0d25-4aa7-848c-6840f01023da",
        "question" : "Which fin design worked best? Why?",
        "rubric" : "{\"items\": [\"I can explain my response\", \"I can identify which fin design worked best\", \"No evidence\"], \"criteria\": \"Decide which variables should be changed, measured and controlled in fair tests and accurately observe, measure and record data\", \"total_score\": \"2\", \"curriculum_codes\": {\"au\": [\"AC9S5I03\", \"VCSISU084\", \"ST3-1WS-S\"]}}",
        "answer" : "\"Fin designs that have three or four triangular fins that are at the base end of the rocket work best. They stabilise the rocket and make it more aerodynamic or streamlined\"",
        "task_title" : "Designing your rocket",
        "task_content" : "Designing your rocket    Building phase     The shape, weight, and size of a rocket, and it's design of nose cone and fins, all affect how aerodynamic or efficient it will be.     Being efficient allows a rocket to use less fuel while travelling long distances or overcoming gravity to take off and escape our atmosphere.     Rockets need to go straight up when launching and not veer to one side or roll when travelling through space. An effective  nose cone  and  fins  will help to stabilise your rocket.     So, to create a rocket that can be launched into space, you must design:     Which fin design worked best? Why? "
    },
    {
        "question_id" : "8add417a-4108-43e6-b033-4fa72ac4f709",
        "task_id" : "35d95305-bc3a-461a-ba96-83063257f32d",
        "question" : "Which planet had the fastest and slowest orbits? How long are they and why do you think this occurs?",
        "rubric" : "{\"items\": [\"I can explain why these planets have the fastest and slowest orbits\", \"I can identify which planet has both the fastest or the slowest orbit, and how long that orbit is\", \"I can identify which planet has either the fastest or the slowest orbit, and how long that orbit is\", \"No
```

```
evidence\"], \"criteria\": \"Earth is part of a system of planets orbiting around a star (the Sun)\",
\"total_score\": \"3\", \"curriculum_codes\": {\"au\": [\"AC9S6U02\", \"VCSSU078\", \"ST3-10ES-S\"]}}",
        "answer" : "\"Mercury has the fastest orbit of 88 days. This happens as it is the closest planet to
the Sun. Neptune has the slowest orbit of 60,190 days. This happens as it is the furthest planet from the
Sun\"",
        "task_title" : "The orbits of other planets",
        "task_content" : "The orbits of other planets    Share a video, photos, or explanation of the model
your team made    Which planet had the fastest and slowest orbits? How long are they and why do you think this
occurs?   Which planet has no tilted axis? Which planet has the most tilted axis? How does this affect these
planets? &amp;nbsp; "
    },
...
]
```


question_id (String): A unique identifier for the question.

- Example: "8b1d6474-a38a-4456-8ac2-74ff054b9d67"

task_id (String): A unique identifier for the task that the question belongs to.

- Example: "41f89fd8-0d25-4aa7-848c-6840f01023da"

question (String): The actual question posed to the students.

- Example: "Which fin design worked best? Why?"

answer (String):

- The exemplar answer for the question, which serves as a reference for the teacher.

rubric (JSON formatted String): A detailed rubric that includes:

- **items** (List of Strings): Key performance indicators for assessing the student's response. items[0] is for the highest score, and items[items.length - 1] is for the lowest score(0)
- **criteria** (String): Describes what students need to do to meet the assessment standards.
- total_score (String): The maximum score a student can achieve for the question.
- curriculum_codes (Object): Curriculum reference codes based on the educational standards, which might differ by region (e.g., Australian curriculum codes).


```
{
  "items": [
    "I can explain my response",
    "I can identify which fin design worked best",
    "No evidence"
  ],
  "criteria": "Decide which variables should be changed, measured and controlled in fair tests and accurately
observe, measure and record data",
  "total_score": "2",
  "curriculum_codes": {
    "au": ["AC9S5I03", "VCSISU084", "ST3-1WS-S"]
  }
}
```