Chat GPT & Education
Learning and teaching

Syrielle Montariol & Paola Mejia
On behalf of Antoine Bosselut Tanja Käser
Overview

1. ChatGPT for students
2. ChatGPT for teachers
Motivation

How to fight with ChatGPT, and not against?
How can we use LLM models in education?

A large-scale project at EPFL, started early 2023:

- We collected exercises and solutions from 142 EPFL courses
- 6442 multiple-choice and open-ended questions

We prompted GPT-3.5 and GPT-4 to:

- Generate a solution to a given question: different strategies!
- Grade a given solution
Goal

How good a student are GPT-3.5 and GPT-4?

How good of a grader is GPT-4?
Dataset

6442 questions
4039 MCQs
2403 Open-Answer

4725 in English
1717 in French

1565 from Bachelors
1907 from Masters
2970 from Online Courses
Prompting Strategies

Direct prompting
- Zero-shot
- One-shot
- Expert

Reason-first prompting
- Zero-shot CoT
- Four-shot CoT
- Metacognitive
- Tree of Thought

Reflective prompting
- Self-critique
Prompting Strategies

Direct prompting
- Question-only
- One example
- Ask several times and average

Reason-first prompting
- Ask for reasoning
- Ask for reasoning With 4 examples
- Follow list of reasoning steps
- Simulate a discussion between experts

Reflective prompting
- Validate a given reasoning and answer
Grading Strategies

- GPT-4 Grading
  - MCQ Auto-Scoring
- Human Grading
  - Direct Grading
  - “Gold” Grading

Only a subset of the courses
Can GPT-3.5 and GPT-4 pass EPFL courses?

At best, GPT-4 would pass 85/142 courses

- 40/67 “MCQs courses”
- 45/75 “open-answer courses”

At best, GPT-3.5 would pass 43/142 courses

- 20/67 “MCQs courses”
- 23/75 “open-answer courses”
What impacts their performance?
What impacts their performance?

- Both GPT-4 and GPT-3.5 are better on English questions, both on MCQ and open-answer questions.

- *Computer Software* and *Computer Systems* are the best performing topics

- *Mechanical Engineering, Linguistics* and *General Computer Science* among the topics GPT-4 is weakest at.
Can we use GPT to grade students’ solutions?

GPT-4 vs the courses' teaching staff:
- GPT-4 almost never grades answers as wrong,
- GPT-4 grades answers as correct at a lower rate than humans
- Correlation between human and GPT-4 grading is low.
Summary

**Question:** Integration of different semiconductor materials in heterostructure form is a widely used strategy to tune the functional properties in view of specific applications. Comment on the strain arising in a heterostructure addressing the following question: What are the key parameters that you need to estimate the type and the amount of strain in the system?

**Prompting GPT-4 and GPT-3.5:**
- You are an expert in...
- As you perform this task, follow these steps:
- Let’s think step by step.

**Grading by GPT-4**

**Grading by teaching staff**

**Analysis**
- Course Size
- Topics
- Language
- Year

- LLMs don’t perform better in more general courses
- LLMs often sound like students trying to get points when they don’t know the answer
- GPT-4 gives much more partial grades than humans
Can we use GPT to provide feedback?

**Motivation:** Feedback (formative assessment) is more useful than grades (summative assessment).

How do students perceive AI-generated feedback?

How does it compare to the TA-generated feedback?
Example - Programming Exercise

a) (4 points) Pour une analyse linguistique, on veut écrire une fonction qui estime le nombre de syllabes d’un mot donné. Pour ce faire, on utilise la règle suivante en parcourant les lettres du mot:
- Chaque voyelle (a, e, i, o ou u) et chaque y compte comme une nouvelle syllable...
- ... sauf si la lettre précédente du mot est elle aussi une de ces 6 lettres.
- Un éventuel e à la fin du mot ne compte pas comme nouvelle syllabe.
- Le nombre de syllabes estimé doit toujours être au minimum 1.

Complétez la fonction count_syllables pour qu’elle retourne l’estimation du nombre de syllabes d’un mot passé en paramètre selon ces règles. Vous pouvez partir du principe que la fonction ne recevra que des mots tout en minuscules, sans accents, caractères spéciaux ou ponctuation.
```python
"def count_syllables(x:str) -> int
    s=0
    if x[0:8] == ""a"" or ""i"" or ""y"" or ""e"" or ""o"" or ""u"":
        s=s+1
    i=1
    while i<len(x)-1:
        if x[i:i+1] == ""a"" or ""i"" or ""e"" or ""y"" or ""u"" or ""o"":
            if x[i-1:i] != x[i:i+1]:
                s=s+1
            i=i+1
    if x[len(x):] == ""a"" or ""i"" or ""y"" or ""u"" or ""o":
        s=s+1
    return (s)"
```
Your turn!

Give feedback to the student

5 minutes
a) Votre solution ne respecte pas complètement les règles énoncées dans la consigne. Votre code actuel ne compte que la présence de voyelles différentes dans le mot, et non le nombre de syllabes selon les critères donnés. Vous devez parcourir chaque lettre du mot et suivre les règles mentionnées pour compter les syllabes, y compris la prise en compte de la lettre précédente et de la dernière lettre si c'est un ”e”. Voici une suggestion pour ajuster votre code :

```python
def count_syllables(word: str) -> int:
    syllables = 1
    vowels = "aeiouy"
    for i in range(1, len(word)):
        if word[i] in vowels and word[i-1] not in vowels:
            syllables += 1
    if word.endswith("e"):
        syllables = max(1, syllables - 1)
    return syllables
```
Discussion time

1) First, discuss the following questions in groups.
   a) What do you think of the GPT generated feedback? Is it correct? Is it useful?
   b) Have you already used GPT in your classroom? How?
   c) What are other applications of GPT that you could imagine using in your classroom?

2) Share your ideas!
   a) 2 minutes: present your group’s opinion and idea to everyone
What do you think of the GPT generated feedback? Is it correct? Is it useful?
Have you already used GPT in your classroom? How?
What are other applications of GPT that you could imagine using in your classroom?
Thank you for your participation!

If you are interested in using GPT for feedback, please contact us to participate in our user study.

Contact person: tanya.nazaretsky@epfl.ch