Try to make dataset with llm, it still got hallucinations. So give up.

Using the provided Excel file, which includes API instructions and API names, we need to create a QA dataset for each API. Follow these steps:

Read each line and get the INSTRUCTION and APINAME. For example, if the INSTRUCTION is "Create a memory space for use with the API buffer. The size of the memory space can be specified in one of several units," the correct API will be:

Correct API: CreateApiBuffer (unsigned int channel, unsigned int size, SizeUnit::T sizeUnit)

Then, create three additional answers:

Similar API with wrong functionality: Choose an API with similar keywords in the INSTRUCTION from the list of all APIs. It needs to be diversified as much as possible, choose from the all APINAME in the file. Don’t make up one.

Example: OpenCyclicBuffer (AxisSelection \*pAxisSelection, unsigned int numOfCycles), because it has same key word “Buffer”.

Use other API in the same Class.

Example: FreeApiBuffer (unsigned int channel)

Fake API with incorrect arguments: Create a fake API by modifying the correct API's arguments, or revise as a pseudo API name.

Example: CreateApiBuffer (unsigned int Bufferchannel, unsigned int Buffersize, SizeUnit::T Buffersize), or CreateWMXApiBuffer (unsigned int channel, unsigned int size, SizeUnit::T sizeUnit)

Randomize the order of the answers (A, B, C, D) and specify the correct answer.

For example:

Question: Create a memory space for use with the API buffer. The size of the memory space can be specified in one of several units.

Answers:

A. CreateApiBuffer (unsigned int Bufferchannel, unsigned int Buffersize, SizeUnit::T Buffersize)

B. FreeApiBuffer (unsigned int channel)

C. OpenCyclicBuffer (AxisSelection \*pAxisSelection, unsigned int numOfCycles)

D. CreateApiBuffer (unsigned int channel, unsigned int size, SizeUnit::T sizeUnit)

Correct answer: D

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
Now let’s try the first 5 lines in the file.

Then, make a python file to make the QA dataset.

Prompt for coding:

Help me to make a python code to make a QA dataset. read DataFile - "docs/WMX3API\_CleanedData.xlsx", and write to DatasetFile - "WMX3API\_QA Dataset.xlsx" following the rules below. Read DataFile line by line, if the column 'Instruction' and 'FunctionPython' are not blank, remember the line number as ln, then start to make a QA in the DatasetFile:

1. Column No in DatasetFile adds 1.
2. Copy 'Instruction' of the ln in DataFile to column 'Instruction' in DatasetFile.
3. Make ABCD optional answers.
   1. Copy ‘FunctionPython’ of the ln in DataFile to column A in DatasetFile.
   2. Find all the functions with the same ‘Class’ as the ln in DataFile, and randomly choose one function in ‘FunctionPython’, put it into column B in DatasetFile.

2024-05-20

Evaluation: GPT-3.5-turbe,

1. 7m 39.3s, Total Correct: 57.22%, Total Wrong: 42.78%
2. With concurrent.futures, 1m 32.4s, Total Questions: 769, Correct Answers: 440 (0.57, 57.22%), Wrong Answers: 329 (0.43, 42.78%), temperature=0.5
3. Total Questions: 769,Correct Answers: 460 (0.60, 59.82%), Wrong Answers: 309 (0.40, 40.18%), temperature=0
4. Correct Answers: 435 (**0.57,** 56.57%), Wrong Answers: 334 (0.43, 43.43%), temperature=1
5. Correct Answers: 415 (0.54, 53.97%), Wrong Answers: 354 (0.46, 46.03%), WMX3API\_QA\_Dataset\_Cleaned.json(not reshuffled)
6. model="gpt-4o", Correct Answers: 562 (0.73, 73.08%), Wrong Answers: 207 (0.27, 26.92%), $0.5  
   Correct Answers: 551 (0.72, 71.65%), Wrong Answers: 218 (0.28, 28.35%), $0.45
7. model="gpt-4-turbe", Correct Answers: 539 (0.70, 70.09%), Wrong Answers: 230 (0.30, 29.91%), $0.88;  
   Correct Answers: 541 (0.70, 70.35%), Wrong Answers: 228 (0.30, 29.65%), $0.73