Evaluation Instructions

We will have three files to evaluate, each for a different type of prompting strategy:

- **Zero-shot**
- **Zero-shot with an example of how to construct the question**
- **Zero-shot with a description of the construct and how to use it**

Question distribution by level:

- A1 – 3
- A2 – 6
- B1 – 6
- B2 – 6
- C1 – 3

Note! Since we are testing how well the given prompting strategy aligns with the specified parameters (number of questions included), it is possible that some prompts will generate fewer questions than expected. If a topic was skipped during prompting, I will include it in our evaluation, so you do not need to lower your score because of it.

These are all the evaluation questions I used. Please answer all of them (except the 7th one!) with **0 or 1** in the cell for the given question.

1. How unique is the question?
2. Is the answer to the question correct?
3. How well does it relate to the topic?
4. Does the level of the topic align with the content of the question?
5. Does the type of question align with the one stated?
6. Are all the topics represented (number of questions)?
7. Personal score for this level generation (0-10) (please answer after all the questions in the empty row).

These are all the topics that were given to each model. I will also attach the updated JSON file.

A1 - past simple for everyday events and states

A1 - main clause with another main clause

A1 - determiners with nouns

A2 - combining two adjectives with 'but'

A2 - adjectives with 'the most'

A2 - adverbs as modifiers of certainty

A2 - past continuous for description of background events

A2 - present continuous for events in process

A2 - gerunds as nouns

B1 - present continuous for questions about the future

B1 - finite clause after 'than'

B1 - adverbs as modifiers of time

B1 - preposition with no article

B1 - irregular plural noun with 's'

B1 - subordinate clause with conjunctions

B2 - present continuous as a future form

B2 - reported speech using a reporting clause

B2 - verb 'will' in requests

B2 - future perfect continuous

B2 - past perfect continuous results

B2 - past simple questions

C1 - compound adjectives

C1 - verb 'can' with the passive

C1 - passive form; non-finite clauses