

Input:

I want you to act as a paraphrase identification expert for Germany .
Sentence 1: Der Ponoru ist eine Nebenfluss des Horezu in Rumänien .
Sentence 2: Der Ponoru ist ein Nebenfluss des Horezu in Rumänien .
Question: Does Sentence 1 paraphrase Sentence 2? Yes or No?
You should retell the sentence 1 and sentence 2 in English.
You should provide a yes or no answer to the question: Does Sentence 1 paraphrase Sentence 2?
You should step-by-step answer the request.
You should tell me the answer choosing either yes or no in this format 'Answer :'.

Retelled Sentence 1: The Ponoru River is a tributary of the Horezu River in Romania .
Retelled Sentence 2: The Ponoru River is a tributary of the Horezu in Romania .
Answer: Yes

... (the remaining 4 examples)

I want you to act as a paraphrase identification expert for Germany .
Sentence 1: Lito spielte für den Zingone Club Football
Sentence 2: Zingone spielte für den Club
Question: Does Sentence 1 paraphrase Sentence 2? Yes or No?
You should retell the sentence 1 and sentence 2 in English.
You should provide a yes or no answer to the question: Does Sentence 1 paraphrase Sentence 2?
You should step-by-step answer the request.
You should tell me the answer choosing either yes or no in this format 'Answer :'.

text-davinci-003 output:

Retelled Sentence 1: Lito played football for the Zingone Club.
Retelled Sentence 2: Zingone played football for the Lito Club.
Answer: Yes

gpt-3.5-turbo output:

Retelled Sentence 1: Lito played football for the Zingone Club.
Retelled Sentence 2: Zingone played football for the Lito Club.
Answer: No

Answer:

No

Figure 17: A Germany example of the PAWS-X benchmark using XLT and the corresponding outputs under the few-shot setting.

Input:

Answer the question in one or a few words in Japanese : 世界で一番背の高い人類は誰ですか？

text-davinci-003 output:

最高身長者。

gpt-3.5-turbo output:

誰？

Answer:

ロバート・ワドロー

Figure 18: A Japanese example of the MKQA benchmark using **basic prompt** and the corresponding outputs under the **zero-shot** setting.

Input:

I want you to act as a question answering expert for Japanese .
Question: 世界で一番背の高い人類は誰ですか？
You should retell the question in English.
You should answer the question in English in one or a few words.
You should step-by-step answer the request.
You should tell me the answer in one or a few words in Japanese in this format 'Answer : '.

text-davinci-003 output:

Question in English: Who is the tallest human being in the world?
Answer: Robert Wadlow (2.72 m).
Answer in Japanese: 答え：ロバート・ワドロウ (2.72 m) 。

gpt-3.5-turbo output:

Retelling: Who is the tallest person in the world?
Answer: Robert Wadlow
Step-by-step response:
1. The question is asking for the tallest person in the world.
2. The answer is Robert Wadlow, who was 8 feet 11 inches (2.72 meters) tall.
Answer in Japanese: 答え：ロバート・ワドロー

Answer:

ロバート・ワドロー

Figure 19: A Japanese example of the MKQA benchmark using **XLT** and the corresponding outputs under the **zero-shot** setting.

Input:

I want you to act as a question answering expert for Japanese .
Question: フルハウスでスティーブが出てきたのはどのシーズンですか？
You should retell the question in English.
You should answer the question in English in one or a few words.
You should step-by-step answer the request.
You should tell me the answer in one or a few words in Japanese in this format 'Answer:'.

Question: what season does steve come in full house?
English Answer: 5.0 seasons.
Japanese Answer: 5.0 時節

... (the remaining 4 examples)

I want you to act as a question answering expert for Japanese .
Question: 世界で一番背の高い人類は誰ですか？
You should retell the question in English.
You should answer the question in English in one or a few words.
You should step-by-step answer the request.
You should tell me the answer in one or a few words in Japanese in this format 'Answer:'.

text-davinci-003 output:

question: who is the tallest human being in the world
english answer: sultan kösen
japanese answer: スルタン・ケーセン

gpt-3.5-turbo output:

question: who is the tallest person in the world
nenglish answer: robert wadlow
japanese answer: ロバート・ワドロー

Answer:

ロバート・ワドロー

Figure 20: A Japanese example of the **MKQA** benchmark using **XLT** and the corresponding outputs under the **few-shot** setting.