In-context-learning prompt for LLM's Self-feedback at translation:
You are an annotator for the quality of machine translation. Your task is to identify errors and assess the quality of the translation.
Based on the source segment and machine translation surrounded with triple backticks, identify error types in the translation and classify them. The categories of errors are: accuracy (addition, mistranslation, omission, untranslated text), fluency (character encoding, grammar, inconsistency, punctuation, register, spelling), locale convention (currency, date, name, telephone, or time format) style (awkward), terminology (inappropriate for context, inconsistent use), non-translation, other, or no-error.
Each error is classified as one of three categories: critical, major, and minor. Critical errors inhibit comprehension of the text. Major errors disrupt the flow, but what the text is trying to say is still understandable. Minor errors are technically errors, but do not disrupt the flow or hinder comprehension.

Source: "'大众点评乌鲁木齐家居商场频道为您提供高铁居然之家地址，电话，营业时间等最新商户信息，找装修公司，就上大众点评"' Translation: "'Urumqi Home Furnishing Store Channel provides you with the latest bussiness information such as the address, telephone number, bussiness hours, etc., of high-speed rail, and find a decoration company, and go to the reviews."' Annotate errors in the translation. MQM annotations:

"of high-speed rail" is a critical accuracy/addition error
"go to the reviews" is a major accuracy/mistranslation error
"etc.," is a minor style/awkwards error

Source: "'I do apologise about this, we must gain permission from the account holder to discuss an order with another person, I apologise if this was done previously, however, I would not be able to discuss this with yourself without the account holders permission."' Translation: "'Ich entschuldige mich dafür, wir müssen die Erlaubnis einholen, um eine Bestellung mit einer anderen Person zu besprechen. Ich entschuldige mich, falls dies zuvor geschehen wäre, aber ohne die Erlaubnis des Kontoinhabers wäre ich nicht in der Lage, dies mit dir involvement."' Annotate errors in the translation. MQM annotations:

'involvement' is a major accuracy/mistranslation error
'the account holder' is a major accuracy/omission error
'wäre' is a minor fluency/grammar error
'dir' is a minor fluency/register error

Source: "'Talks have resumed in Vienna to try to revive the nuclear pact, with both sides trying to gauge the prospects of success after the latest exchanges in the stop-start negotiations."' Translation: "'Ve Vídni se ve Vídni obnovily rozhovory o oživení jaderného paktu, přičemže obě partaje se snaží posoudit vyhlídky na úspěch po posledních výměnách v jednáních."' Annotate errors in the translation. MQM annotations:

've Vídni' is a major accuracy/addition error
'the stop-start' is a major accuracy/omission error
'partaje' is a minor terminology/inappropriate for context error

Source: "'Talks have resumed in Vienna to try to revive the nuclear pact, with both sides trying to gauge the prospects of success after the latest exchanges in the stop-start negotiations."' Translation: "'Ve Vídni se ve Vídni obnovily rozhovory o oživení jaderného paktu, přičemže obě partaje se snaží posoudit vyhlídky na úspěch po posledních výměnách v jednáních."' Annotate errors in the translation. MQM annotations:

Source: "'O ko ago ilekun WiFi, O wi."' Translation: "'He locked the WiFi door, he said."' Annotate errors in the translation. MQM annotations:

Output for Feedback:
'He locked the WiFi door' is a critical accuracy/mistranslation error

Table 9: Those are the feedback in context learning examples (three examples) that we used to prompt all LLMs across four language directions at Flores200. In this example, the source translation is Yourba text "O ko ago ilekun WiFi, O wi.". The English candidate text is "He locked the WiFi door" from GPT-4. GPT-4's self-feedback outputs "'He locked the WiFi door' is a critical accuracy/mistranslation error".

In-context-learning prompt for LLM's refinement at translation:
Please fix all errors. You can rewrite translation if translation is bad.

Source: "'大众点评乌鲁木齐家居商场频道为您提供高铁居然之家地址，电话，营业时间等最新商户信息，找装修公司，就上大众点评"'
Translation: "'Urumqi Home Furnishing Store Channel provides you with the latest bussiness information such as the address, telephone number, bussiness hours, etc., of high-speed rail, and find a decoration company, and go to the reviews."'

Feedback: 'of high-speed rail' is a critical accuracy/addition error
'go to the reviews' is a major accuracy/mistranslation error
'etc.,' is a minor style/awkward error

Improved Chinese-to-English translation:
Dianping Urumqi Renovation and Design Channel will provide you with the address, phone number, operation time and other information of HSR Easyhome, and please come to Dianping if you are looking for a renovation company.

Source: "'I do apologise about this, we must gain permission from the account holder to discuss an order with another person, I apologise if this was done previously, however, I would not be able to discuss this with yourself without the account holders permission."'
Translation: "'Ich entschuldige mich dafür, wir müssen die Erlaubnis einholen, um eine Bestellung mit einer anderen Person zu besprechen. Ich entschuldige mich, falls dies zuvor geschehen wäre, aber ohne die Erlaubnis des Kontoinhabers wäre ich nicht in der Lage, dies mit dir involvement."'

Feedback:
'involvement' is a major accuracy/mistranslation error
'the account holder' is a major accuracy/omission error
'wäre' is a minor fluency/grammar error
'dir' is a minor fluency/register error

Improved English-to-German translation:
Ich bitte um Entschuldigung, aber wir benötigen das Einverständnis des Kontoinhabers, um eine Bestellung mit einer anderen Person zu besprechen, falls es schon eingeholt wurde, entschuldige ich mich, aber ich kann dies ohne das Einverständnis des Kontoinhabers nicht mit Ihnen besprechen.
Source: "'Talks have resumed in Vienna to try to revive the nuclear pact, with both sides trying to gauge the prospects of success after the latest exchanges in the stop-start negotiations."'
Translation: "'Ve Vídni se ve Vídni obnovily rozhovory o oživení jaderného paktu, přičemže obě partaje se snaží posoudit vyhlídky na úspěch po posledních výměnách v jednáních."'

Feedback:
've Vídni' is a major accuracy/addition error
'the stop-start' is a major accuracy/omission error
'partaje' is a minor terminology/inappropriate for context error

Improved English-to-Cezch translation:
Ve Vídni byly obnoveny rozhovory o oživení jaderného paktu a obě strany se snaží odhadnout, jaké jsou vyhlídky na úspěch po posledních výměnách názorů v rámci přerušených jednání.

Source: "'O ko ago ilekun WiFi, O wi."'
Translation: 'He locked the WiFi door, he said."'

Feedback:
'He locked the WiFi door' is a critical accuracy/mistranslation error

Improved English-to-Yoruba translation:
Output for refinement:
"He secured the WiFi connection, he said."

Table 10: Those are the refinement in context learning examples (three examples) that we used to prompt all LLMs across four language directions at Flores200. In this example, the source translation is Yourba text "O ko ago ilekun WiFi, O wi.". The English text is GPT-4's output "He locked the WiFi door.". We gathered GPT-4's self-feedbak: "'He locked the WiFi door' is a critical accuracy/mistranslation error". GPT-4 outputs a refined output: "He secured the WiFi connection, he said.", according to the feedback and prior step generation.

Prompt for LLM's initial generation for reasoning:
You are a competitive math problem solver. Please generate a step-by-step solution. Your final answer must be enclosed in LaTeX's boxed tag.

How many vertical asymptotes does the graph of $y = \frac{2}{x^2+x-6}$ have?

Output for math answer:
A vertical asymptote is a vertical line that the graph of a function approaches but never crosses. A function has a vertical asymptote at $x = k$ if it has infinite limits as x approaches k from the left or right.For the rational function $y = \frac{2}{x^2+x-6}$, we know that it could have vertical asymptotes wherever the denominator is zero, because division by zero is undefined and yields infinity.Therefore, to find the vertical asymptotes, we solve the equation $x^2 + x - 6 = 0$. We apply factorization and get $(x - 2)(x + 3) = 0$. This gives us $x = 2$ and $x = -3$.In any case, we need to check if any of these could be holes instead of vertical asymptotes. This would be the case if $x = 2$ or $x = -3$ also made the numerator zero. However, the numerator is a constant and won't be zero. Therefore this function has two vertical asymptotes at $x = 2$ and $x = -3$. The answer is $\boxed{2}$.

Table 11: This is the prompt we used to generate mathematical reasoning. The output in this example comes from GPT-4. We first used this prompt to generate one reasoning path as initial generation. Then, we sample 10 different reasoning paths and use majority vote to find the most confident answer. To generate self-feedback, we compare majority voted answer and initial answer. If they are matched, feedback indicates answer is correct and no change should be made. If they are not matched, the feedback indicates answer is wrong and inital answer will be replaced by the majority voted answer. We alternate above two steps and iteratively refine the reasoning paths.

Concept-to-text prompt for LLM's initial generation:
You are generating text based on specified words. Do not provide any explanations or text apart from the text output.

Please generate a sentence that contains the exact string matches for the following concepts:

['vest', 'snow', 'remove', 'silhouette', 'dunk', 'step', 'item', 'toddler', 'plaster', 'wall', 'piece', 'tool', 'circle', 'pull', 'wrap', 'headphone', 'deck', 'suit', 'press', 'mother', 'pit', 'cupcake', 'priest', 'rink', 'counter', 'rest', 'piano', 'wine', 'serve', 'winter']

Output for Concept-to-Text generation:

In the winter, the toddler wearing a vest, pressed a piece of wet plaster against the wall in a circle, while his mother baking a cupcake was removing the snow from the deck with a tool; switching steps to the rhythm of a piano piece echoing from the headphones wrapped around her neck, she caught a glimpse of the priest serving wine, his silhouette dunking a cupcake into a pit at the rink counter, under the restful gaze of a suit-clad man.

Table 12: This is the concept-to-text prompt we used to test all LLMs at Commongen Hard. This example contains the output from GPT-4.