# Bootstrapping Grounded Chain-of-Thought in Multimodal LLMs for Data-Efficient Model Adaptation

Jiaer Xia[1] Bingkui Tong[2] Yuhang Zang[3] Rui Shao[4] Kaiyang Zhou[1]✉
[1]Hong Kong Baptist University
[2]Sichuan University [3]Shanghai AI Lab [4]Harbin Institute of Technology (Shenzhen)

## A. Appendix

### A.1. Evaluation of Generalization Ability

Theoretically, our approach can stem the generalizability from clearer and better alignment training data. Compared to the original data, which only includes simple answer annotations, CoT data provides the detailed underlying structures, relationships, and logical patterns within the data. This enables clearer alignment and understanding, rather than merely remembering the simple answer, thereby capturing more generalizable patterns. Additionally, the proposed GCoT provides higher-quality reasoning data by eliminating the noisy information that may exist in the distilled CoT data. To better demonstrate generalizability, we conducted a cross-model evaluation, as shown in Tab. 1. The results indicate that the model trained on ChartQA is capable of generalizing improvements to other datasets. Notably, our proposed GCoT method shows even greater enhancements.

Table 1. The model is trained on 128 ChartQA datasets and demonstrates strong generalization capabilities across different datasets with proposed GCoT method.

| Method | ChartQA | TabMWP | DVQA |
|---|---|---|---|
| Zero-shot | 10.12 | 19.60 | 13.95 |
| CoT | 20.44 | 21.01 | 12.34 |
| GCoT | 25.18 | 25.55 | 15.73 |

### A.2. Evaluation of Chain-of-Thought Accuracy

To address the scarcity of detailed annotations for intermediate reasoning steps and to more intuitively validate our process verification approach, we manually examined 64 ChartQA CoT samples, as shown in Table 2. The results support the findings from our ablation study: although the distilled CoT contains certain errors, incorporating box verification substantially reduces these mistakes, resulting in more accurate training data. Moreover, if process verification is omitted and only the final answers are checked, the self-augmented data exhibit a higher error rate. In contrast, GCoT effectively alleviates this issue, yielding significantly more reliable training data.

Table 2. Box-based Verification can effectively eliminate errors and provide more accurate training samples.

| | CoT | GCoT | w/o Box | GCoT+Aug |
|---|---|---|---|---|
| Acc/All | 49/64 | 57/64 | 33/64 | 162/192 |

### A.3. Visualization

This section presents illustrative examples to clarify the details of GCoT. We demonstrate the self-verified GCoT with various sources of distillation data and different datasets, highlighting the improvements brought about by GCoT.

#### A.3.1. Different source of Distillation CoT

In the main paper, we have shown some examples to compare distillation CoT data from different sources and analyze why their performance varies significantly under training. Additionally, we will contrast these with our self-generated GCoT data, with the results illustrated in Fig. 1. In contrast to the distillation CoT, the GCoT markedly reduces inaccuracies in the reasoning process, resulting in higher-quality training data.

#### A.3.2. Chart Understanding Task Case

Figs. 2–4 illustrate GCoT's adaptability across different datasets in the context of chart understanding tasks. Firstly, in comparison to the distillation data, GCoT significantly reduces inaccuracies and standardizes the format. Furthermore, GCoT is capable of generating a diverse range of CoT data while maintaining high accuracy. These examples highlight GCoT's ability to produce reliable and varied datasets through self-generation. This capability not only expands the originally limited datasets but also enhances training efficiency, leading to improved performance across diverse tasks.

**TabMWP**

| Fan letters received | |
|---|---|
| **Day** | **Number of fan letters** |
| Wednesday | 204 |
| Thursday | 204 |
| Friday | 337 |
| Saturday | 463 |
| Monday | 271 |

**Question:** An actor was informed how many fan letters he received each day. How many fan letters total were received on Thursday and Monday?
**Answer:** 475

**Prompt:** Based on the following question: An actor was informed how many fan letters he received each day. How many fan letters total were received on Thursday and Monday? Your task is to give a explanation for the question. Give step by step reasoning to get the answer, and when you're ready to answer, please use the format '*Answer*:'

**GCoT (LLaMA)**

To determine the total number of fan letters received by the actor on Thursday and Monday, we can take the sum of those two days' fan letter counts. The fan letter count on Thursday [0.000, 0.450, 0.200, 0.515] was 204. The fan letter count on Monday [0.019, 0.819, 0.275, 0.895] was 271 [0.619, 0.819, 0.719, 0.865]. Thus, the total number of fan letters received on Thursday and Monday is: 204 + 271 = 475.
*Answer*: 475.

**GCoT (Claude)**

1. First, let's understand the problem: The actor was told how many fan letters he received.
2. Potential fan letters:   - The number of fan letters [0.411, 0.195, 0.998, 0.271] could be 204 received on Thursday [0.000, 0.455, 0.281, 0.531]   - 271 received on Monday [0.000, 0.715, 0.255, 0.871].
3. Total fan letters calculation: - 204 + 271 = 475

*Answer*: 475

**GCoT (Gemini)**

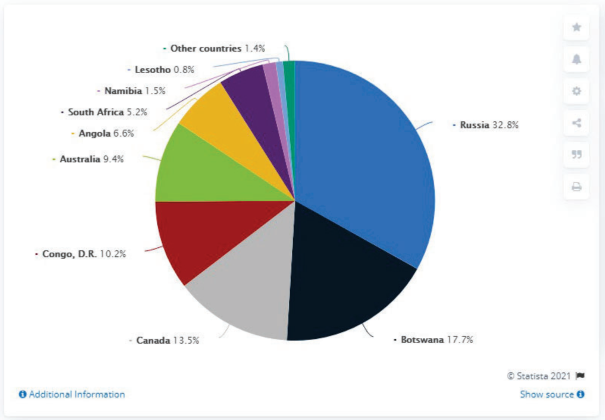
We need to find the total number of fan letters received on Thursday and Monday. The table shows the number of fan letters for each day. The actor received 204 fan letters on Thursday [0.095, 0.455, 0.250, 0.515] and 271 [0.619, 0.795, 0.750, 0.855] on Monday. Calculate the total number of fan letters:** 204 + 271 = 475
*Answer*: 475

**GCoT (Qwen)**

To determine the total number of fan letters received by the actor on Thursday [0.000, 0.455, 0.200, 0.519] and Monday, we can take the sum of those two days' fan letter counts. The fan letter count on Thursday was 204. The fan letter count on Monday was 271 [0.617, 0.793, 0.751, 0.859]. Thus, the total number of fan letters received on Thursday and Monday is: 204 + 271 = 475.
*Answer*: 475.

Figure 1. Self-generated GCoT data eliminating external language discrepancies and aligning closely with the model's linguistic tendencies.

**ChartQA**



**Question:** Is the sum of Canada and Botswana more then Russia?
**Answer:** No

**LLaMA**

To determine if the sum of Canada and Botswana is more than Russia, we need to calculate the total percentage of Canada and Botswana and compare it with Russia's percentage.
Step 1: Calculate the sum of Canada and Botswana's percentages.
  - Canada = 10.2%
  - Botswana = 17.7%
  Sum of Canada and Botswana = 10.2% + 17.7% = 27.9%
Step 2: Compare the sum of Canada and Botswana with Russia's percentage.
  Russia = 32.8%
Step 3: Determine if the sum of Canada and Botswana is more than Russia.
  Since 27.9% (sum of Canada and Botswana) is less than 32.8% (Russia), the answer is no.
*Answer*: No.

**GCoT**

To answer this question, we need to look at the provided pie chart and identify the values for each country. The chart shows that the sum of Canada [0.211, 0.701, 0.351, 0.715] and Botswana [0.651, 0.699, 0.791, 0.715] is 31.2%, while Russia has a value of 32.8%. Comparing these values, we can see that the sum of Canada [0.211, 0.701, 0.351, 0.715] and Botswana (31.2%) is less than Russia (32.8 %). *Answer*: No.

**GCoT**

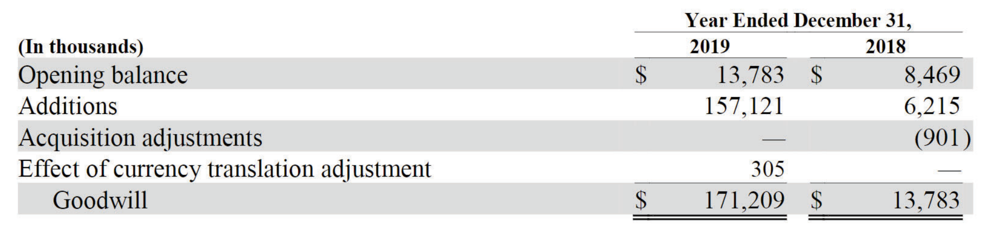
To answer this question, we need to add the values of Canada [0.211, 0.701, 0.351, 0.715] and Botswana together and then compare it to the value of Russia. The values are:
  Canada [0.211, 0.701, 0.351, 0.715] = 13.5
  Botswana [0.651, 0.701, 0.791, 0.715] = 17.7
  Total = 13.5 + 17.7 = 31.2
  Russia = 32.8
Since the sum (31.2) is less than the value of Russia (32.8), the answer is no.
*Answer*: No.

**GCoT**

To determine whether the sum of Canada [0.211, 0.701, 0.331, 0.715] and Botswana is more than Russia, we need to add these values together. Canada [0.211, 0.701, 0.331, 0.715] has a value of 13.5 [0.263, 0.699, 0.334, 0.715]. Botswana has a value of 17.7 [0.731, 0.688, 0.798, 0.723]. When we add these values together, we get: 13.5 + 17.7 = 31.8. Now, let‘s compare this value to Russia’s value: Russia has a value of 32.8 [0.791, 0.338, 0.868, 0.373]. When we compare the values, we can see that Russia‘s value is larger than the sum of Canada [0.211, 0.701, 0.331, 0.715] and Botswana. *Answer*: No.

Figure 2. Example of GCoT applied to ChartQA dataset, illustrating its ability to generate reliable and diverse training CoT data.

**TAT-QA**



| (In thousands) | Year Ended December 31, 2019 | 2018 |
|---|---|---|
| Opening balance | $ 13,783 | $ 8,469 |
| Additions | 157,121 | 6,215 |
| Acquisition adjustments | — | (901) |
| Effect of currency translation adjustment | 305 | — |
| Goodwill | $ 171,209 | $ 13,783 |

**Question:** From 2018 to 2019, how many years was the Opening balance more than $5,000 thousand?
**Answer:** 2

**LLaMA**

To determine the number of years the Opening balance was more than $5,000 thousand from 2018 to 2019, we need to compare the Opening balance for each year with $5,000 thousand.
In 2018, the Opening balance was $8,469 thousand. Since $8,469 is greater than $5,000, the Opening balance was more than $5,000 thousand in 2018.
In 2019, the Opening balance was $13,783 thousand. Since $13,783 is also greater than $5,000, the Opening balance was more than $5,000 thousand in 2019 as well.
Therefore, the Opening balance was more than $5,000 thousand for both years from 2018 to 2019.
*Answer*: 2

**GCoT**

To determine the number of years the Opening balance [0.021, 0.421, 0.111, 0.439] was more than $5,000 thousand from 2018 [0.821, 0.385, 0.911, 0.397] to 2019 [0.642, 0.385, 0.722, 0.397], we need to examine the Opening balances for these two years. In 2018, the Opening balance is $8,469 thousand. Since this value is greater than $5,000 thousand, the Opening balance was more than $5,000 thousand in 2018. In 2019, the Opening balance is $13,783 thousand. Since this value is also greater than $5,000 thousand, the Opening balance was more than $5,000 thousand in 2019 as well. Therefore, the Opening balance was more than $5,000 thousand in both 2018 and 2019, indicating that it was more than $5,000 thousand for two consecutive years from 2018 to 2019. *Answer*: 2.

**GCoT**

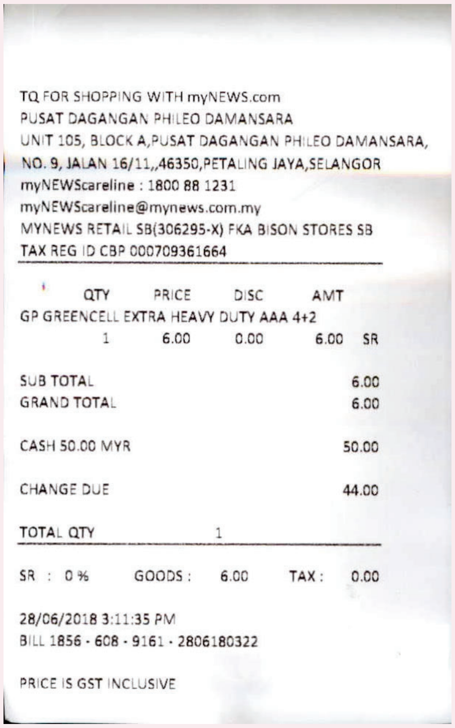
To find out the number of years from 2018 to 2019 [0.621, 0.379, 0.717, 0.391] with the Opening balance more than $5,000 thousand, we need to look at the table provided and find the row that corresponds to the Opening balance. The column that corresponds to the Opening balance is the first column labeled as \"Opening Balance.\" We then look at the the first column to see the value of the Opening balance for 2018 and 2019. In 2018, the Opening balance is $8,469 thousand. In 2019, the Opening balance is $13,783 thousand. The Opening balance is more than $5,000 thousand in both years.*Answer*: 2.

**GCoT**

To find the number of years the Opening balance was more than $5,000 thousand from 2018 to 2019, we need to look at the Opening balance over those two years. Opening Balance 2018 = $8,469 [0.833, 0.399, 0.923, 0.422] thousand, Opening Balance 2019 = $13,783 thousand. Since the Opening balance was more than $5,000 thousand in both 2018 and 2019, the answer is 2. *Answer*: 2.

Figure 3. Example of GCoT applied to TAT-QA dataset, illustrating its ability to generate reliable and diverse training CoT data.

### SROIE



TQ FOR SHOPPING WITH myNEWS.com
PUSAT DAGANGAN PHILEO DAMANSARA
UNIT 105, BLOCK A,PUSAT DAGANGAN PHILEO DAMANSARA,
NO. 9, JALAN 16/11,,46350,PETALING JAYA,SELANGOR
myNEWScareline : 1800 88 1231
myNEWScareline@mynews.com.my
MYNEWS RETAIL SB(306295-X) FKA BISON STORES SB
TAX REG ID CBP 000709361664

| | QTY | PRICE | DISC | AMT | |
|---|---|---|---|---|---|
| GP GREENCELL EXTRA HEAVY DUTY AAA 4+2 | | | | | |
| | 1 | 6.00 | 0.00 | 6.00 | SR |

| | |
|---|---|
| SUB TOTAL | 6.00 |
| GRAND TOTAL | 6.00 |
| CASH 50.00 MYR | 50.00 |
| CHANGE DUE | 44.00 |
| TOTAL QTY | 1 |

SR : 0 % GOODS : 6.00 TAX : 0.00

28/06/2018 3:11:35 PM
BILL 1856 - 608 - 9161 - 2806180322

PRICE IS GST INCLUSIVE

**Question:** What is the total amount in the invoice shown in the picture?
**Answer:** 6.00

### LLaMA

To find the total amount in the invoice, we need to look at the "SUB TOTAL" and "GRAND TOTAL" sections. The "SUB TOTAL" is 6.00, and the "GRAND TOTAL" is also 6.00. This means that there are no additional charges or taxes added to the subtotal, so the grand total remains the same as the subtotal. *Answer*: 6.00.

### GCoT

The invoice shows the total [0.271, 0.541, 0.325, 0.555] amount as 6.00. *Answer*: 6.00.

### GCoT

To determine the total amount in the invoice, we need to look at the bottom of the receipt. The total amount is listed as 6.00 [0.671, 0.542, 0.725, 0.561]. *Answer*: 6.00.

### GCoT

To find the total amount in the invoice, we need to look at the bottom of the receipt. The total amount [0.274, 0.539, 0.325, 0.553] is listed as 6.00 [0.671, 0.542, 0.725, 0.561]. This means that the total amount due for the transaction is 6.00. *Answer*: 6.00.

Figure 4. Example of GCoT applied to SROIE dataset, illustrating its ability to generate reliable and diverse training CoT data.