Graph Ingestion Engine (Fall 2024)

Mohamed Essam Moustafa Kamel Fayed, Bruce Walker

mohamed.fayed@gatech.edu, bruce.walker@psych.gatech.edu

Abstract—Multimodal Large Language Models have shown impressive visual capabilities in many Visual Question Answering tasks. In this paper, we aim to test them on Chart-to-Table task.

CONTENTS

1 INTRODUCTION

2 RELATED WORK

Extracting data points from charts has witnessed attention from research community. There has been work to summarize charts end-to-end. Another direction was related to converting charts to tables. There has been a lot of efforts in summarizing charts, answering questions Masry et al., Masry et al., 2022, 2024 and converting them into tables Liu et al., 2022. Recently, there has been efforts to analyze the performance of Large Visual Models (LVMs) in many all of those tasks. In our work, we aim to pay closer attention to Chart-to-Table task.

3 METHODOLOGY

3.1 Datasets

3.2 Models

· Gemini 1.5 Flash
· ChartGemma
· Deplot

3.3 Evaluation

· Relative Mapping Similarity (RMS)
· Qualitative

4 RESULTS AND DISCUSSION

4.1 Text Recognition

For the sample we analyzed, there has been no errors in recognizing text in the images, e.g. columns names. However, table **??** shows that ChartGemma has a tendency to labelize even if there are no labels in the input image. [1] [2] For both models, the tables layouts were perfectly generated into table in json format for Gemini and markdown for ChartGemma.

---

1 the prediction of Gemini and Ground Truth have no labels for x-axis, but ChartGemma made years as labels.

2 In some cases, the ground truth is mislabelled. The reference has no values for x-axis, but the image includes them as in 5.

4.2 Values Extraction

For PlotQA and ICPR22 samples, it is frequent to find errors like:

1. rounding errors, e.g. $15.42 -> 15$ and $15.6 -> 15$.
2. Precision Errors: we have noticed that the model can not predict more than 3 digits for each value, e.g. $126765000.0 -> 156000000$.
3. In case of near values, e.g. $24.18, 24.09$, there might be some errors, e.g. predicting 23 instead of 24. For that kind of error, it may result in changing trend, e.g. steady performance may seam as decreasing. [3]
4. Gemini can differntiate outputs based on scale, e.g. $156000000 \& 50.2$ for instance. However, ChartGemma sometimes change scale, e.g. table 4 where the model returned values multiplied by 10.
5. Occasionally, both models swap two columns as shown in table 8. As a result, RMS score is significantly lower (f1=0.34) than its fixed version (f1=0.83).

In the following subsubsections, we illustrate issues related to each kind of graphs.

4.2.1 Bar Charts

1. Tables **??** and **??** show that both models are very good in extracting data points from bar charts. [4]

4.2.2 Line Charts

1. There are some graphs, like 2, the Gemini API just fails with no clear response message (till now). However, it is suspected that the very large number of data points might be the reason.
2. Table 7 shows that ChartGemma may fail in extracting data points from slightly complex graphs. It fails in both extracting correct values as well as mapping them to the correct label.

---

3 It is worth noting that we have not seen cases where increasing is replaced by decreasing trends or vice versa.

4 A small notice, that needs more examples to approve/disapprove, is that ChartGemma has lower margin of error while having less precision. The numbers of Canada, for instance, are correctly approximated to 52. This may indicate almost steady value, which sounds reasonable conclusion for that country, especially when looking to the whole graph at a glance.

Tables 1 and **??** include Gemini 1.5 Flash and ChartGemma predictions for figure 1

	Country	2005 Cost of computers, communications and other servic
1	Belarus	31
2	Belize	15
3	Bosnia and Herzegovina	33
4	Brazil	45
5	Cabo Verde	11
6	Canada	52

respectively.
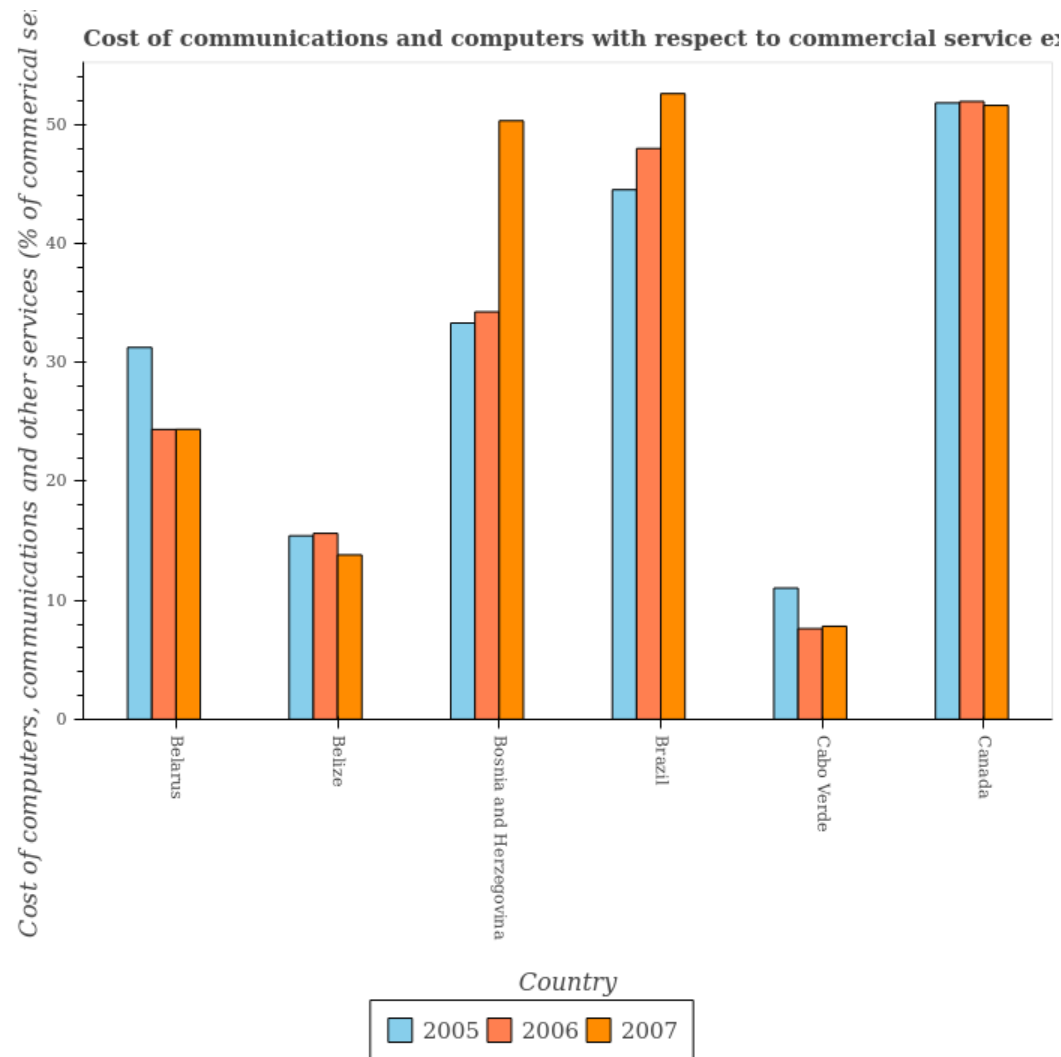
Figure 1—Vertical Bar Chart example from PlotQA testset.

	Country	2005	2006	2007
0	Belarus	31	24	23
1	Belize	15	15	13
2	Bosnia and Herzegovina	33	34	52
3	Brazil	44	47	52
4	Cabo Verde	10	7	8
5	Canada	51	52	51

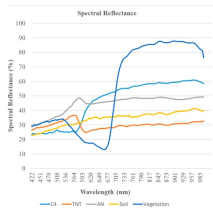*Table 1*—Gemini 1.5 Flash predictions on Vertical Bar # 25905

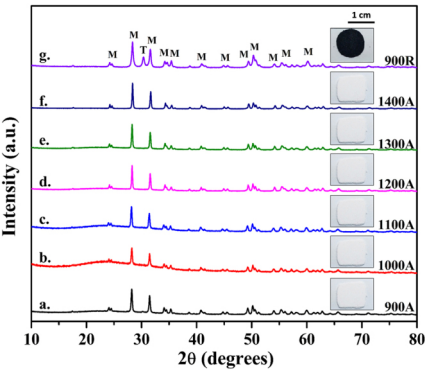*Figure 2*—Example for charts that causes the API to fail.

Figure 3—A good example for graph in the wild that causes Gemini 1.5 Flash to fail.

	name	color	label	bboxes
0	Portfolio Investment	#BA55D3	Portfolio Investment	['y': 51, 'x': 132, 'w': 466, 'h': 413, 'y': 51, '

Table 2—Reference for Line Chart from PlotQA #21673 Portfolio Investment

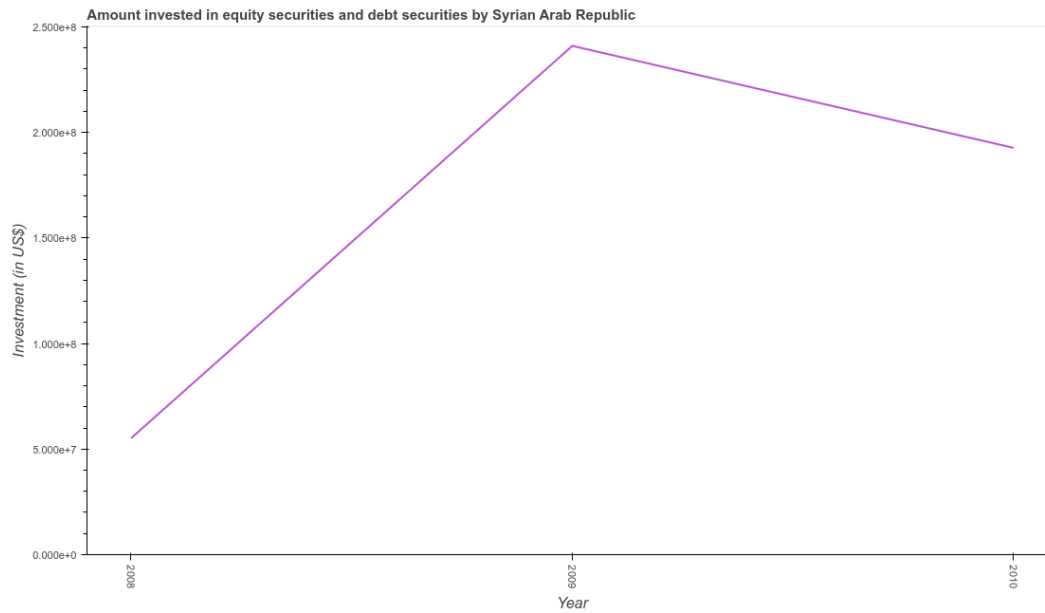**Amount invested in equity securities and debt securities by Syrian Arab Republic**

Figure 4—Example for Line Chart from PlotQA testset # 21673 about Portfolio Investment

	Year	Investment (in USD)
0	2008	54000000
1	2009	240000000
2	2010	200000000

Table 3—Predicted data points by Gemini 1.5 Flash for Line Chart from PlotQA #21673 Portfolio Investment

	Year	Investment (in USD)
1	2008	500000000
2	2009	2400000000
3	2010	1900000000

Table 4—ChartGemma: prediction for PlotQA line chart #21673

6
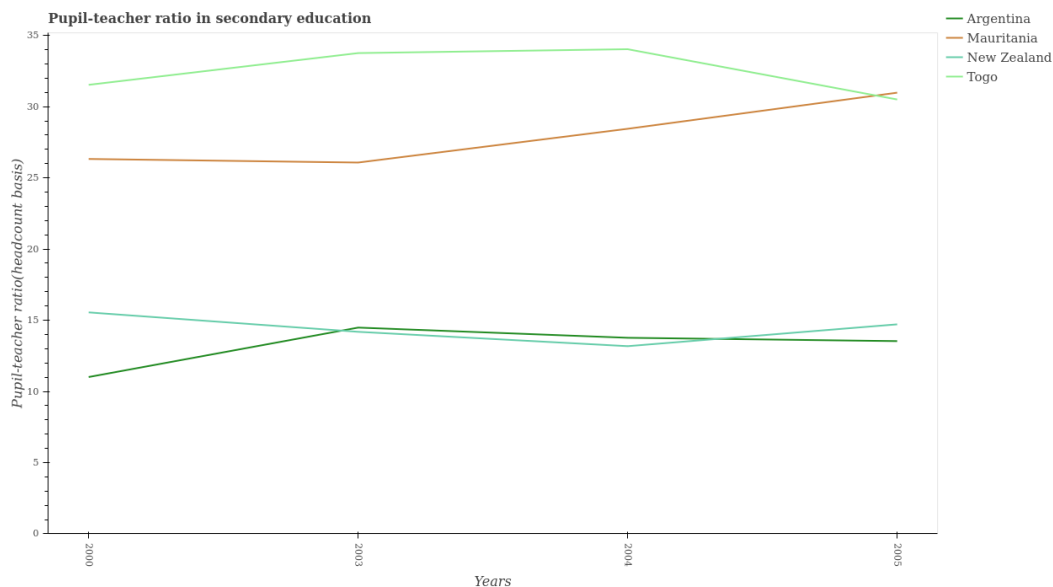
Figure 5—PlotQA # 20049: Line chart containing 4 lines.

	name	color	label	bboxes
0	Argentina	#228B22	Argentina	['y': 386, 'x': 101, 'w': 320, 'h': 58, 'y': 386, 'x': 421, 'w':
1	Mauritania	#CD853F	Mauritania	['y': 186, 'x': 101, 'w': 320, 'h': 4, 'y': 150, 'x': 421, 'w':
2	New Zealand	#66CDAA	New Zealand	['y': 368, 'x': 101, 'w': 320, 'h': 23, 'y': 391, 'x': 421, 'w':
3	Togo	#90EE90	Togo	['y': 60, 'x': 101, 'w': 320, 'h': 38, 'y': 56, 'x': 421, 'w': 32

Table 5—Reference table for PlotQA line chart # 20049

	Year	Argentina	Mauritania	New Zealand	Togo
0	2000	10.600000	26.000000	15.800000	31.400000
1	2003	14.200000	25.800000	14.000000	33.200000
2	2004	13.600000	28.000000	13.000000	33.800000
3	2005	13.400000	30.200000	14.600000	30.000000

Table 6—Gemini 1.5 Flash prediction for PlotQA line chart # 20049

	Years	Argentina	Mauritius	New Zealand	Togo
1	2000	21	15	22	11
2	2003	21	14	23	14
3	2004	22	13	23	13
4	2005	21	14	21	14

Table 7—ChartGemma prediction for PlotQA line chart # 20049. The model fails in mapping lines with values, e.g. Togo column seams more likely to be Argentina. For values, it is obvious that ChartGemma is very far away from correctly detecting values greater than 20!

	Australia	Turkmenistan
0	'Year': 2009.0, 'Subscribers per 100 People': 47.0	'Year': 2009.0, 'Subscribers per 100 People': 48.5
1	'Year': 2010.0, 'Subscribers per 100 People': 46.0	'Year': 2010.0, 'Subscribers per 100 People': 47.5
2	'Year': 2011.0, 'Subscribers per 100 People': 45.0	'Year': 2011.0, 'Subscribers per 100 People': 46.0
3	'Year': 2012.0, 'Subscribers per 100 People': 44.5	'Year': 2012.0, 'Subscribers per 100 People': 45.0
4	'Year': 2013.0, 'Subscribers per 100 People': 44.0	'Year': 2013.0, 'Subscribers per 100 People': 44.0

Table 8—Example for Gemini Flash predictions where it swapped the values of Turkmenistan and United States. The swapped table has score of $F_1 = 0.34$ and the corrected version has $F_1 = 0.83$.

5 CONCLUSION AND RECOMMENDATIONS

In this report, we document our quanitative analysis for LLMs behavior in Chart-to-Table task. Based on the selected sample, we observed that the model can accurately recognize the layout of the graph, but it is not very precise in recognizing small differences in values. For future work, we recommend combining both LLMs and Computer Vision algorithms to complement each other in accurately converting charts into tables. [5]

6 REFERENCES

1. Liu, Fangyu, Eisenschlos, Julian Martin, Piccinno, Francesco, Krichene, Syrine, Pang, Chenxi, Lee, Kenton, Joshi, Mandar, Chen, Wenhu, Collier, Nigel, and Altun, Yasemin (2022). "Deplot: One-shot visual language reasoning by plot-to-table translation". In: *arXiv preprint arXiv:2212.10505*.
2. Masry, Ahmed, Long, Do Xuan, Tan, Jia Qing, Joty, Shafiq, and Hoque, Enamul (2022). "Chartqa: A benchmark for question answering about charts with visual and logical reasoning". In: *arXiv preprint arXiv:2203.10244*.
3. Masry, Ahmed, Thakkar, Megh, Bajaj, Aayush, Kartha, Aaryaman, Hoque, Enamul, and Joty, Shafiq (2024). "ChartGemma: Visual Instruction-tuning for Chart Reasoning in the Wild". In: *arXiv preprint arXiv:2407.04172*.

---

5 Based on my expertise in using LLaMA 3.1 8B Instruct, we can convert among formats with almost no errors, e.g. convert prints from python code in latex table. It correctly follows instruction of to round numerical values or copy them as is.