

Chart-to-Table Conversion: A Survey

Mohamed Fayed, Kruthik Ravikanti, Bruce Walker
mohamed.fayed@gatech.edu, kravikanti3@gatech.edu,
bruce.walker@psych.gatech.edu

Abstract—Multimodal Large Language Models (MLLMs) have shown impressive visual capabilities in many Visual Question Answering tasks. In this paper, we aim to survey the recent advancements in Chart-to-Table task, score the performance of some MLLMs and highlight their strengths and weaknesses. Our quantitative and qualitative analysis shows that there is a room for improvement in Chart-to-Table conversion.

CONTENTS

1	Introduction	2
2	Chart-to-Table in Literature	2
3	Methodology	3
3.1	Datasets	3
3.2	Models	4
3.3	Evaluation	4
4	Results and Discussion	4
4.1	Relative Mapping Similarity Scores	4
4.2	Text Recognition	4
4.3	Values Extraction	4
4.3.1	Bar Charts	5
4.3.2	Line Charts	5
5	Conclusion and Recommendations	11

1 INTRODUCTION

Chart-to-Table is the task of extracting data points from an image of a chart into a table usually in markdownLiu et al., Masry et al., 2022a, 2024b. This task is important in the process of digitizing those charts into more space efficient format of text. Moreover, tables are more accessible mean of communicating data to people with disabilities who count on screen readers in interacting with digital world.

There has been a lot of efforts in summarizing charts, answering questions Masry et al., Masry et al., 2022, 2024b and converting them into tables Liu et al., 2022a. Recently, there has been efforts to analyze the performance of Multimodal Large Language Models (MLLMs) in many all of those tasks. In our work, we aim to pay closer attention to Chart-to-Table task. Our main contributions are:

1. Survey recent advancements in Chart-to-Table task,
2. Do quantitative analysis for some models on different benchmark datasets,
3. Do fine-grained qualitative analysis on various kinds of charts, and
4. highlight strengths, weaknesses and rooms for improvement of those models in performing this task

2 CHART-TO-TABLE IN LITERATURE

There has been work in operating in charts in a way or another. Some researchers aim at converting charts to tables (Chart-to-Table Conversion). Others have worked on Chart Question Answering and Chart Summarization. In this work, we focus on Chart-to-Table Conversion.

It all begins with computer vision (CV) based methods.

Fasting forward, researchers gave deep learning a try.

Pretraining and Large Language Models attracted researchers. Masry et al., 2023 gathered a large dataset of 6.9M questions and charts, and used it to pre-train UniChart model. Another group of researchers Liu et al., 2022b introduced Matcha model, which was a fine-tuned version of Pix2Struct Lee et al., 2023 on many tasks including Chart-to-Table. Later on, Liu et al., 2022a continued pre-training it on Chart-to-Table only to create Deplot. They forwarded the generated table, human

query and an example to FlanPaLM 540B Chung et al., 2024 to answer complex queries.

MLLMs has made significant progress in many tasks and Chart-to-Table was no exception. One direction is about utilizing general domain LLMs without any tuning. This direction includes Prompting and Retrieval Augmented Generation techniques Cao et al., Voigt et al., 2024, 2023 to improve the capabilities of general LLMs on chart related tasks.

Another direction is to fine-tune LLMs on Chart-specific Instruction following datasets. Masry et al., 2024a introduced an instruction following dataset and instruction tuned both LLaMA 2 7B Touvron et al., 2023 and Flan-T5 XL 3B Chung et al., 2024 on it. This instruction tuning strategy was proved to make a generalized model that can handle unseen tasks. Similarly, Masry et al., 2024b fine-tuned PaliGemma Beyer et al., 2024 to create a 3B ChartGemma. Another key distinguishing contribution was the method of instruction following data generation. They generated for predefined tasks, such as Chain of Thought and Chart-to-Tables in form of markdown, and open-ended tasks, such as justifying trends in charts and describing visual elements.

3 METHODOLOGY

3.1 Datasets

In our analysis, we focus on reporting scores on testsets of ICPR22 Rousseau and Kapralos, 2023 and PlotQA Methani et al., 2020 datasets. ICPR22 testset Rousseau and Kapralos, 2023 is gathered from research papers published on PubMed Central website.¹ Those publications are in biomedical and life sciences domains. It contains 443 charts splitted into 5 types: Line Charts, Horizontal and Vertical Bar Charts, Scatter Plot and Vertical Box Plot.

PlotQA Methani et al., 2020 was made by gathering data from various online sources, such as World Bank and Open Data, generate plots out of these data points, and ask annotators questions about those provided charts. In our work, we focus on the data points used in constructing the charts only. Its test set contains 33657 charts divided equally among dotted line charts, line charts, and vertical

¹ <https://pmc.ncbi.nlm.nih.gov>

and horizontal bar charts. In our analysis, due to limitations on API calls and time constraints, we ran computed scores for 3000 randomly selected charts.

3.2 Models

For our analysis, we selected the following models:

- Gemini 1.5 Flash Team et al., 2024: A general purpose lightweight MLLM.
- ChartGemma Masry et al., 2024b: A specialized model in chart summarization, question answering and reasoning about charts. It utilizes PaliGemma Beyer et al., 2024 as its backbone, and was tuned on Visual Chart Instructions dataset.
- Deplot Liu et al., 2022a:

3.3 Evaluation

- Relative Mapping Similarity (RMS)
- Qualitative

4 RESULTS AND DISCUSSION

4.1 Relative Mapping Similarity Scores

4.2 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.^{2 3} For both models, the tables layouts were perfectly generated into table in json format for Gemini and markdown for ChartGemma.

4.3 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. $126765000.0- > 156000000$.

² the prediction of Gemini and Ground Truth have no labels for x-axis, but ChartGemma made years as labels.

³ In some cases, the ground truth is mislabelled. The reference has no values for x-axis, but the image includes them as in [5](#).

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 seem as decreasing. ⁴
4. Gemini can differentiate outputs based on scale, e.g. 156000000&50.2 for instance. However, both models sometimes change scale, e.g. table 4 where ChartGemma 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 subsections, we illustrate issues related to each kind of graphs.

4.3.1 Bar Charts

1. Tables ?? and ?? show that both models are very good in extracting data points from bar charts. ⁵

4.3.2 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.

⁴ It is worth noting that we have not seen cases where increasing is replaced by decreasing trends or vice versa.

⁵ 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

respectively.

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

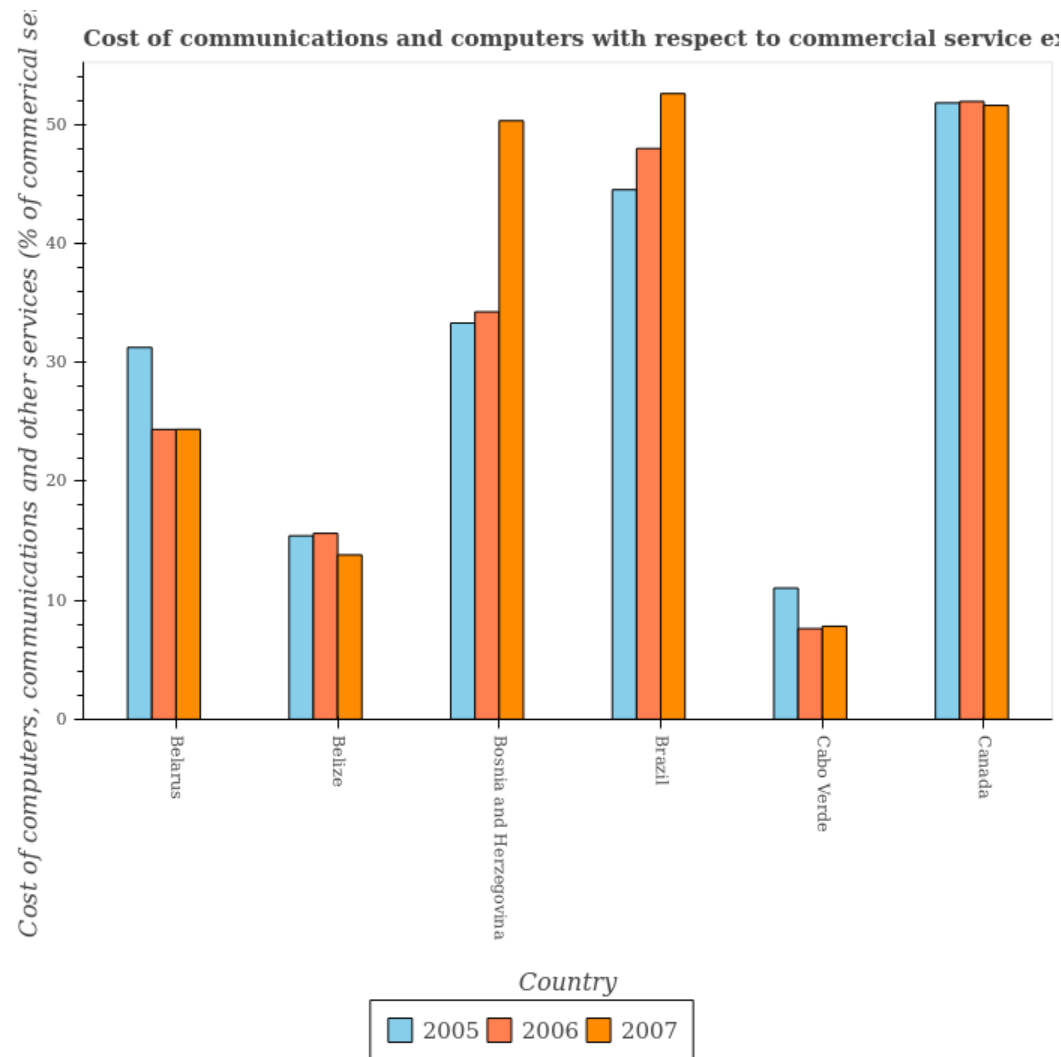


Figure 1—Vertical Bar Chart example from PlotQA testset.

	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

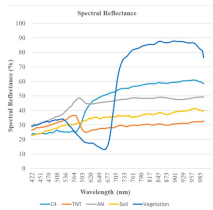


Figure 2—Example for charts that causes the API to fail.

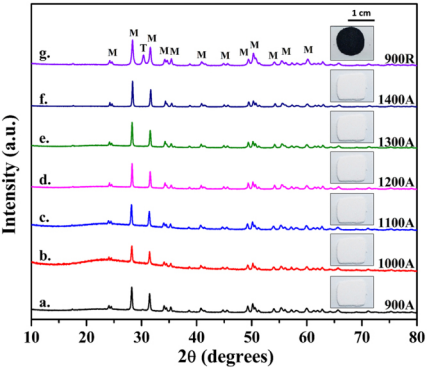


Figure 3—A good example for graph in the wild that causes Gemini 1.5 Flash to fail.

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

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

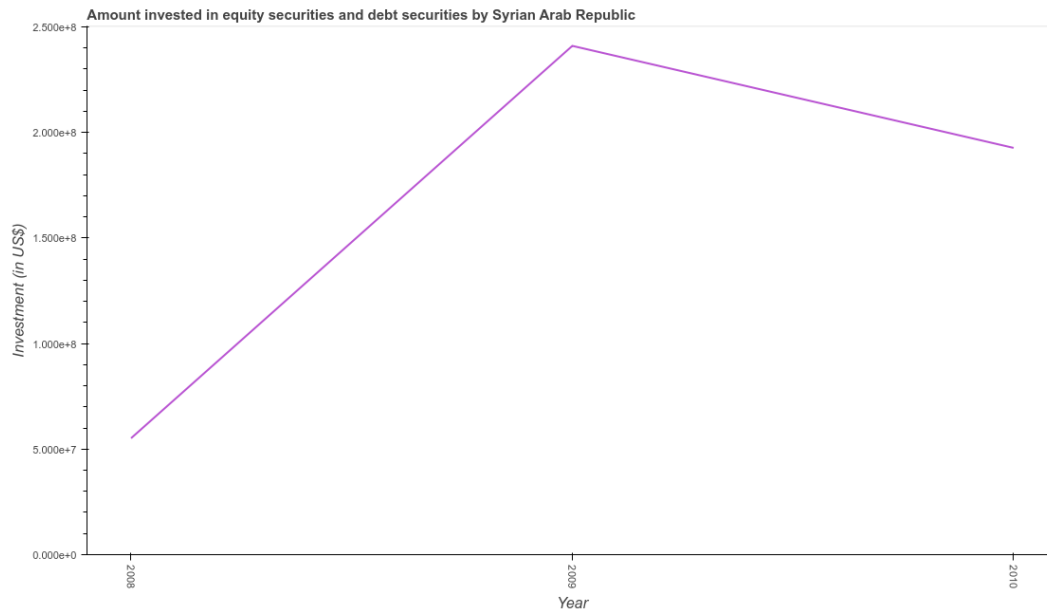


Figure 4—Example for Line Chart from PlotQA testset # 21673 about Portfolio Investment

	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

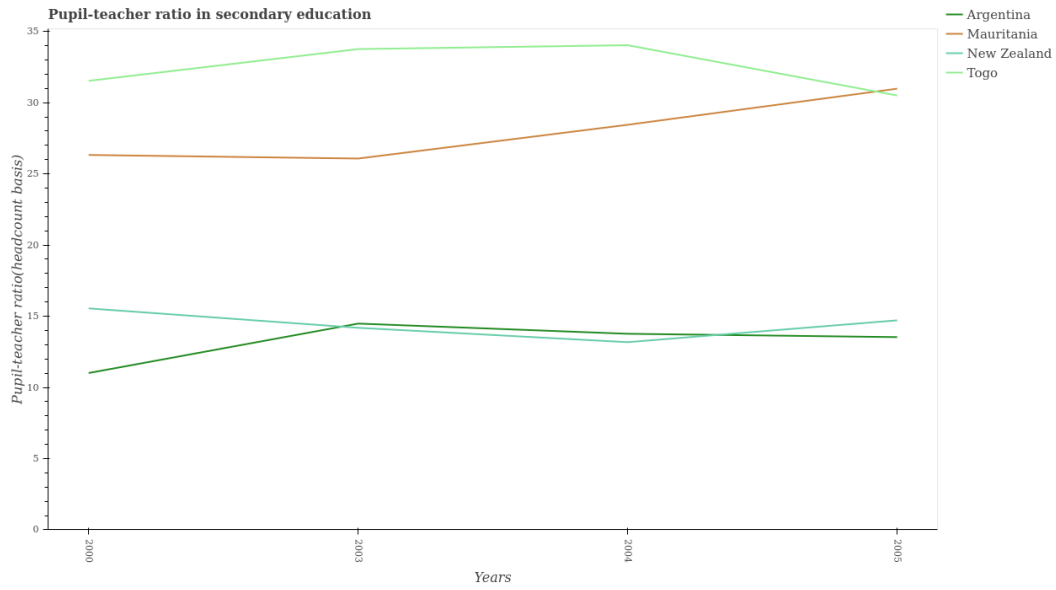


Figure 5—PlotQA # 20049: Line chart containing 4 lines.

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

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 seems 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 $F1 = 0.34$ and the corrected version has $F1 = 0.83$.

5 CONCLUSION AND RECOMMENDATIONS

In this report, we document our quantitative 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.⁶

6 REFERENCES

1. Methani, Nitesh, Ganguly, Pritha, Khapra, Mitesh M, and Kumar, Pratyush (2020). "Plotqa: Reasoning over scientific plots". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1527–1536.
2. Liu, Fangyu, Eisenschlos, Julian Martin, Piccinno, Francesco, Krichene, Syrine, Pang, Chenxi, Lee, Kenton, Joshi, Mandar, Chen, Wenhui, Collier, Nigel, and Altun, Yasemin (2022a). "Deplot: One-shot visual language reasoning by plot-to-table translation". In: *arXiv preprint arXiv:2212.10505*.
3. Liu, Fangyu, Piccinno, Francesco, Krichene, Syrine, Pang, Chenxi, Lee, Kenton, Joshi, Mandar, Altun, Yasemin, Collier, Nigel, and Eisenschlos, Julian Martin (2022b). "Matcha: Enhancing visual language pretraining with math reasoning and chart derendering". In: *arXiv preprint arXiv:2212.09662*.
4. 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*.
5. Lee, Kenton, Joshi, Mandar, Turc, Iulia Raluca, Hu, Hexiang, Liu, Fangyu, Eisenschlos, Julian Martin, Khandelwal, Urvashi, Shaw, Peter, Chang, Ming-Wei, and Toutanova, Kristina (2023). "Pix2struct: Screenshot parsing as pre-training for visual language understanding". In: *International Conference on Machine Learning*. PMLR, pp. 18893–18912.
6. Masry, Ahmed, Kavehzadeh, Parsa, Do, Xuan Long, Hoque, Enamul, and Joty, Shafiq (2023). "Unichart: A universal vision-language pretrained model for chart comprehension and reasoning". In: *arXiv preprint arXiv:2305.14761*.

⁶ 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.

7. Rousseau, Jean-Jacques and Kapralos, Bill (2023). *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges: Montreal, QC, Canada, August 21–25, 2022, Proceedings, Part I*. Vol. 13643. Springer Nature.
8. Touvron, Hugo, Martin, Louis, Stone, Kevin, Albert, Peter, Almahairi, Amjad, Babaei, Yasmine, Bashlykov, Nikolay, Batra, Soumya, Bhargava, Prajjwal, Bhosale, Shruti, et al. (2023). “Llama 2: Open foundation and fine-tuned chat models”. In: *arXiv preprint arXiv:2307.09288*.
9. Voigt, Henrik, Carvalhais, Nuno, Meuschke, Monique, Reichstein, Markus, Zarrie, Sina, and Lawonn, Kai (2023). “VIST5: An adaptive, retrieval-augmented language model for visualization-oriented dialog”. In: *The 2023 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 70–81.
10. Beyer, Lucas, Steiner, Andreas, Pinto, André Susano, Kolesnikov, Alexander, Wang, Xiao, Salz, Daniel, Neumann, Maxim, Alabdulmohsin, Ibrahim, Tschanen, Michael, Bugliarello, Emanuele, et al. (2024). “Paligemma: A versatile 3b vlm for transfer”. In: *arXiv preprint arXiv:2407.07726*.
11. Cao, Yukun, Han, Shuo, Gao, Zengyi, Ding, Zezhong, Xie, Xike, and Zhou, S Kevin (2024). “Graphinsight: Unlocking insights in large language models for graph structure understanding”. In: *arXiv preprint arXiv:2409.03258*.
12. Chung, Hyung Won, Hou, Le, Longpre, Shayne, Zoph, Barret, Tay, Yi, Fedus, William, Li, Yunxuan, Wang, Xuezhi, Dehghani, Mostafa, Brahma, Siddhartha, et al. (2024). “Scaling instruction-finetuned language models”. In: *Journal of Machine Learning Research* 25.70, pp. 1–53.
13. Masry, Ahmed, Shahmohammadi, Mehrad, Parvez, Md Rizwan, Hoque, Enamul, and Joty, Shafiq (2024a). “ChartInstruct: Instruction Tuning for Chart Comprehension and Reasoning”. In: *arXiv preprint arXiv:2403.09028*.
14. Masry, Ahmed, Thakkar, Megh, Bajaj, Aayush, Kartha, Aaryaman, Hoque, Enamul, and Joty, Shafiq (2024b). “ChartGemma: Visual Instruction-tuning for Chart Reasoning in the Wild”. In: *arXiv preprint arXiv:2407.04172*.
15. Team, Gemini, Georgiev, Petko, Lei, Ving Ian, Burnell, Ryan, Bai, Libin, Gulati, Anmol, Tanzer, Garrett, Vincent, Damien, Pan, Zhufeng, Wang, Shibo, et al. (2024). “Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context”. In: *arXiv preprint arXiv:2403.05530*.