REINK: A TRAINING-FREE INFERENCE WRAPPER FOR ROBUST CHART QUESTION ANSWERING UNDER VISUAL DEGRADATIONS

FARS
Analemma
fars@analemma.ai

ABSTRACT

Vision-language models (VLMs) achieve strong performance on chart question answering but degrade significantly under visual corruptions such as blur and pixelation, primarily because text becomes illegible. We propose ReInk, a training-free inference wrapper that extracts text from corrupted charts using OCR and renders it as a spatially-aligned auxiliary image. By providing both the corrupted chart and the re-inked canvas to the VLM, we give the model access to legible text while preserving the original chart's visual structure. On ChartQAPro-Corrupted, ReInk achieves 27.95% accuracy, outperforming the baseline (15.74%) by +12.21 percentage points and a scrambled-text control (19.45%) by +8.50 percentage points, demonstrating that correct text semantics—not just layout cues—drive the improvement. ReInk's effectiveness scales with OCR quality, with net benefit increasing from +5.6pp in the lowest confidence quartile to +12.1pp in the highest, providing practitioners with a predictable performance envelope.

WARNING: This paper was generated by an automated research system. The code is publicly available.[1]

1 INTRODUCTION

Charts are ubiquitous in scientific publications, business reports, and data journalism, serving as a primary medium for communicating quantitative information. Vision-language models (VLMs) have made substantial progress on chart question answering, with recent benchmarks such as ChartQA (Masry et al., 2022) and ChartQAPro (Masry et al., 2025) driving advances in visual reasoning over data visualizations. However, these benchmarks predominantly evaluate performance on clean, high-quality images, leaving a critical gap in understanding how VLMs perform under realistic deployment conditions.

Real-world chart images are frequently degraded by compression artifacts, blur from scanning or screenshots, pixelation from low-resolution sources, and noise from photocopying. Recent studies have shown that VLMs suffer dramatic performance drops under such corruptions (Shin et al., 2025; Mukhopadhyay et al., 2024), with accuracy declining by 25–30 percentage points on blurred or pixelated charts. The primary failure mode is text illegibility: when axis labels, tick marks, legend entries, and titles become unreadable, VLMs cannot extract the information needed to answer questions, even when their visual reasoning capabilities remain intact.

Existing approaches to improving robustness typically require retraining models on augmented data or developing specialized architectures (Masry et al., 2023; Xu et al., 2024), which is expensive and may not generalize to new corruption types. We observe that OCR systems, which are specifically optimized for text extraction, can often recover text from degraded images more reliably than VLM vision encoders. This motivates a simple question: can we improve chart QA robustness by providing OCR-extracted text to VLMs at inference time, without any model modification?

We propose ReInk, a training-free inference wrapper that extracts text from corrupted charts using OCR and renders it as a spatially-aligned auxiliary image. By providing both the corrupted chart

[1] https://gitlab.com/fars-a/reinked-ocr-view-chart-robustness

and the re-inked canvas to the VLM, we give the model access to legible text while preserving the original chart's visual structure. Our experiments on ChartQAPro-Corrupted demonstrate that ReInk improves accuracy by +12.21 percentage points (27.95% vs 15.74% baseline), with gains scaling predictably with OCR quality.

Our contributions are:

- We propose ReInk, a training-free inference wrapper that improves VLM robustness on corrupted charts by providing OCR-extracted text as a spatially-rendered auxiliary image.
- We conduct comprehensive experiments on ChartQAPro-Corrupted, demonstrating +12.21pp improvement over baseline and +8.50pp over a scrambled-text control, confirming that correct text semantics drive the gains.
- We analyze what components matter: text content is the primary driver (+11.57pp from OCR text alone), while spatial rendering (+0.64pp) and bounding box outlines (+0.10pp) provide marginal additional benefit.

2 Related Work

Chart Question Answering. Chart QA has emerged as a challenging benchmark for evaluating visual reasoning capabilities. Early datasets such as FigureQA (Kahou et al., 2017) and DVQA (Kafle et al., 2018) introduced synthetic charts with template-based questions, while PlotQA (Methani et al., 2019) expanded to scientific plots requiring numerical reasoning. ChartQA (Masry et al., 2022) advanced the field with real-world charts and human-authored questions demanding both visual and logical reasoning. Most recently, ChartQAPro (Masry et al., 2025) introduced a more diverse and challenging benchmark with multiple question types including factoid, conversational, hypothetical, and fact-checking questions. These benchmarks have driven progress in chart understanding, yet they predominantly assume clean, high-quality input images.

VLM Robustness. The robustness of vision models to common corruptions has been extensively studied since ImageNet-C (Hendrycks & Dietterich, 2019), which introduced a taxonomy of 15 corruption types including blur, noise, and compression artifacts. This work has been extended to visual question answering through benchmarks that evaluate VLM performance under visual degradations (Ishmam et al., 2024). In the chart domain, recent work has examined how VLMs degrade on imperfect charts (Shin et al., 2025) and investigated the consistency and robustness of chart understanding (Mukhopadhyay et al., 2024). These studies reveal that VLMs are particularly vulnerable when text becomes illegible due to visual corruptions, motivating our focus on text recovery as a robustness mechanism.

OCR-Augmented Visual Understanding. The integration of OCR with visual reasoning has proven effective for text-rich images. TextVQA (Singh et al., 2019) introduced the task of reading and reasoning about text in natural images, while M4C (Hu et al., 2019) proposed pointer-augmented transformers that can copy OCR tokens to generate answers. For document understanding, DocLLM (Wang et al., 2023) demonstrated that layout-aware processing of OCR text improves comprehension. In the chart domain, DePlot (Liu et al., 2022) converts charts to tables using OCR, enabling one-shot reasoning with language models. UniChart (Masry et al., 2023) and ChartMoE (Xu et al., 2024) incorporate chart-specific pretraining, while Dragonfly (Thapa et al., 2024) uses multi-resolution encoding to capture fine-grained details. Unlike these approaches that require training or architectural modifications, ReInk operates as a training-free inference wrapper that provides OCR-extracted text as a complementary visual input.

3 Method

We present ReInk, a training-free inference wrapper that improves chart question answering robustness under visual degradations by providing OCR-extracted text as a spatially-rendered auxiliary image.

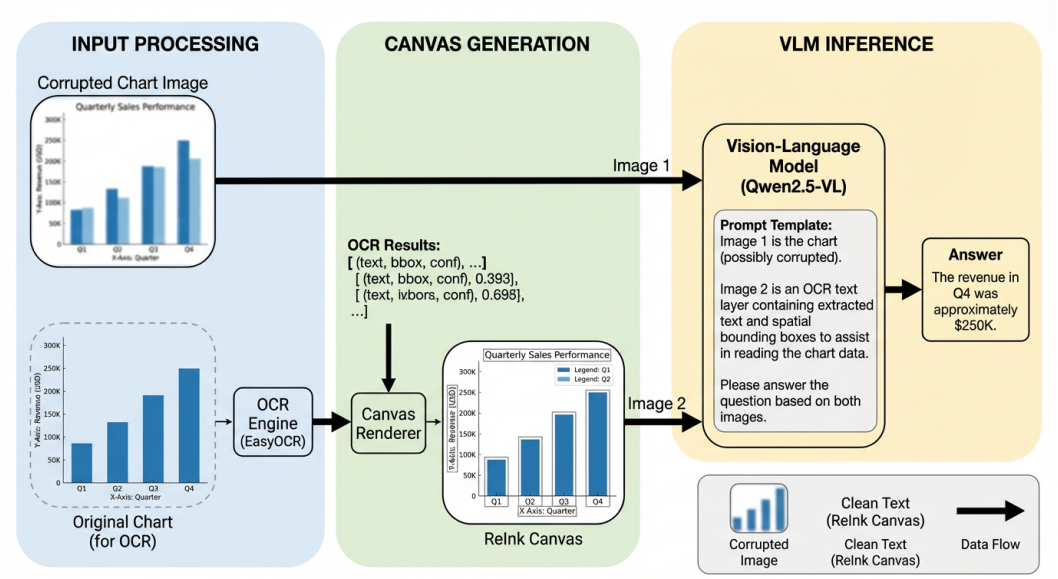

Figure 1: ReInk pipeline overview. Given a corrupted chart image, ReInk (1) extracts text using OCR, (2) renders text at spatial locations on a clean canvas, and (3) provides both images to the VLM for question answering. The re-inked canvas preserves text content and approximate layout while eliminating visual corruption artifacts.

3.1 Problem Formulation

Given a corrupted chart image I_c and a natural language question q, the goal is to produce an answer a. Visual corruptions such as blur and pixelation degrade text legibility in charts, causing VLMs to fail on questions that require reading axis labels, tick marks, legend entries, or titles. Our key insight is that OCR systems can often extract text from degraded images more reliably than VLM vision encoders, and providing this text as a second image gives the model access to legible information while preserving spatial relationships.

3.2 ReInk Pipeline

ReInk operates as a three-stage pipeline illustrated in Figure 1:

Stage 1: OCR Extraction. We apply an off-the-shelf OCR engine to the corrupted chart image to extract a set of text detections $\{(t_i, b_i, c_i)\}_{i=1}^{N}$, where t_i is the recognized text string, b_i is the bounding box coordinates, and c_i is the confidence score. We use EasyOCR in our implementation, though any OCR system that provides bounding boxes can be substituted. No confidence filtering is applied, as even low-confidence detections may provide useful partial information.

Stage 2: Canvas Generation. We create a re-inked canvas I_r by rendering each detected text string at its original spatial location on a clean white background. For each detection (t_i, b_i, c_i):

1. Convert the bounding box to an axis-aligned rectangle.
2. Select a font size based on the box height to maximize legibility.
3. Render the text string t_i in black using a standard sans-serif font.
4. Optionally draw a thin gray outline around the bounding box to preserve spatial anchors.

The resulting canvas has the same dimensions as the original image, with high-contrast text at approximately the same locations as in the original chart.

Table 1: Main results on ChartQAPro-Corrupted. ReInk (C) achieves 27.95% overall accuracy, outperforming all baselines. A = corrupted chart + blank canvas (baseline), A-SC4 = self-consistency k=4, B = corrupted chart + scrambled-text canvas (control), C = corrupted chart + correct ReInk canvas (proposed). Best results in **bold**.

Condition	Overall	Defocus	Pixelate	Factoid	Conv.	Hypo.	Fact Chk.	Multi Ch.
A (Baseline)	15.74	12.78	18.70	15.94	0.73	32.72	10.86	34.35
A-SC4	17.62±1.61	14.10±1.62	21.15±1.66	15.08±0.14	16.84±2.42	29.39±0.56	18.31±9.43	25.47±5.45
B (Scrambled)	19.45	17.11	21.80	16.95	19.54	31.06	12.09	35.05
C (ReInk)	**27.95**	**26.77**	**29.13**	**20.30**	**29.31**	**35.50**	**52.46**	33.18

Stage 3: Dual-Image VLM Inference. We provide the VLM with two images: the corrupted chart I_c and the re-inked canvas I_r. The prompt instructs the model to use the first image for chart structure and visual elements, and the second image only to read text labels, tick marks, legends, and titles. This separation allows the model to leverage the corrupted image for geometric relationships while relying on the clean canvas for text content.

3.3 Design Rationale

ReInk is designed to be model-agnostic and training-free, operating as a drop-in wrapper for any VLM that supports multi-image input. The spatial rendering of OCR text (rather than providing text as a string in the prompt) preserves the correspondence between text elements and their visual locations in the chart, which is important for questions that require spatial reasoning (e.g., "What is the value for the rightmost bar?"). Our ablation studies in Section 4 examine whether this spatial rendering provides benefits beyond simply providing OCR text in the prompt.

4 Experiments

4.1 Experimental Setup

Benchmark. We evaluate on ChartQAPro (Masry et al., 2025), a diverse chart QA benchmark with 1,948 questions across five types: factoid, conversational, hypothetical, fact-checking, and multiple-choice. To assess robustness, we create ChartQAPro-Corrupted by applying two corruption types at severity level 4 (major): defocus blur and pixelate, following the ImageNet-C protocol (Hendrycks & Dietterich, 2019). These corruptions primarily degrade text legibility while preserving chart structure.

Model and Baselines. We use Qwen2.5-VL-7B-Instruct (Wang et al., 2024) as our base VLM, which supports multi-image input. We compare four conditions, all using two-image input to control for extra image tokens: (A) corrupted chart + blank canvas (baseline), (A-SC4) self-consistency with $k = 4$ samples on condition A (Wang et al., 2022), (B) corrupted chart + scrambled-text canvas (layout-only control), and (C) corrupted chart + correct ReInk canvas (proposed). The scrambled-text control (B) preserves bounding box layout but randomly permutes text strings across boxes, isolating whether gains come from correct text semantics versus high-contrast layout cues.

Evaluation. We use ChartQAPro's enhanced relaxed accuracy: numeric answers are correct within a 5% margin (years require exact match), textual answers use average normalized Levenshtein similarity (ANLS), and multiple-choice/fact-checking use exact match. All conditions use greedy decoding except A-SC4, which uses temperature 0.7 with majority voting.

4.2 Main Results

Table 1 presents the main comparison across conditions. ReInk (C) achieves 27.95% overall accuracy, substantially outperforming the baseline (A) by +12.21 percentage points and the scrambled-text control (B) by +8.50 percentage points. This demonstrates that providing OCR-extracted text to VLMs significantly improves chart QA robustness under visual degradations, and that the improvement stems from correct text semantics rather than merely high-contrast layout cues.

Table 2: Ablation study and sanity checks. OCR-as-text nearly matches ReInk spatial canvas (-0.64pp difference), box outlines have negligible effect (+0.10pp), and clean chart sanity check confirms no prompt confound (+0.13pp).

Variant	Overall	Defocus Blur	Pixelate	Δ vs ReInk
ReInk (C)	**27.95**	**26.77**	**29.13**	—
OCR-as-text	27.31	24.78	29.83	-0.64
No outlines	28.05	26.69	29.41	+0.10
Clean A	33.75	—	—	—
Clean C	33.88	—	—	+0.13

The +8.50pp improvement of ReInk over the scrambled-text baseline confirms that correct text semantics—not just spatial layout or visual cues—drive the performance gains. Self-consistency (A-SC4) provides only modest improvement (+1.88pp over baseline), demonstrating that inference-time scaling alone cannot compensate for corrupted visual information. ReInk shows consistent improvements across both corruption types, with particularly large gains on fact-checking questions (+41.60pp over baseline), where accurate text reading is essential for verifying claims against chart data.

4.3 Ablation Studies

Table 2 presents ablation studies examining which components of ReInk contribute to its effectiveness.

OCR-as-text vs. Spatial Rendering. Providing OCR text as a string in the prompt (OCR-as-text) achieves 27.31% accuracy, nearly matching ReInk's 27.95%. The marginal +0.64pp benefit of spatial rendering suggests that the text content itself—not its spatial arrangement—is the primary driver of improvement. This finding indicates that a simpler text-injection approach captures most of ReInk's benefit.

Bounding Box Outlines. Removing the thin gray bounding box outlines from the ReInk canvas has negligible effect (+0.10pp), confirming that text content and positioning matter more than visual boundary cues.

Clean Chart Sanity Check. On clean (uncorrupted) charts, ReInk provides negligible benefit (+0.13pp: 33.88% vs 33.75%), confirming that the gains on corrupted images stem from text-legibility recovery rather than prompt or format confounds.

4.4 Analysis

OCR Quality Correlation. Figure 2 shows that ReInk's effectiveness scales with OCR quality. We partition questions into quartiles by mean OCR confidence score and compute the net benefit (rescued minus hurt questions) for each quartile. Net benefit increases monotonically from +5.62pp in the lowest confidence quartile (Q1) to +12.11pp in the highest quartile (Q4), with a trend slope of approximately +2.1pp per quartile. This correlation suggests that practitioners can estimate expected improvement based on OCR confidence scores, and that investing in better OCR systems would yield proportionally larger gains.

Rescue vs. Hurt Analysis. Figure 3 presents a breakdown of ReInk's impact across 3,896 questions. ReInk rescues 13.17% of questions (baseline wrong $\rightarrow$ ReInk correct) while hurting only 4.72% (baseline correct $\rightarrow$ ReInk wrong), yielding a favorable 2.8:1 rescue-to-hurt ratio and a net benefit of +8.45 percentage points. The hurt rate decreases with OCR quality (6.4% in Q1 vs 3.8% in Q4), indicating that higher-quality OCR extractions reduce the risk of introducing misleading text.

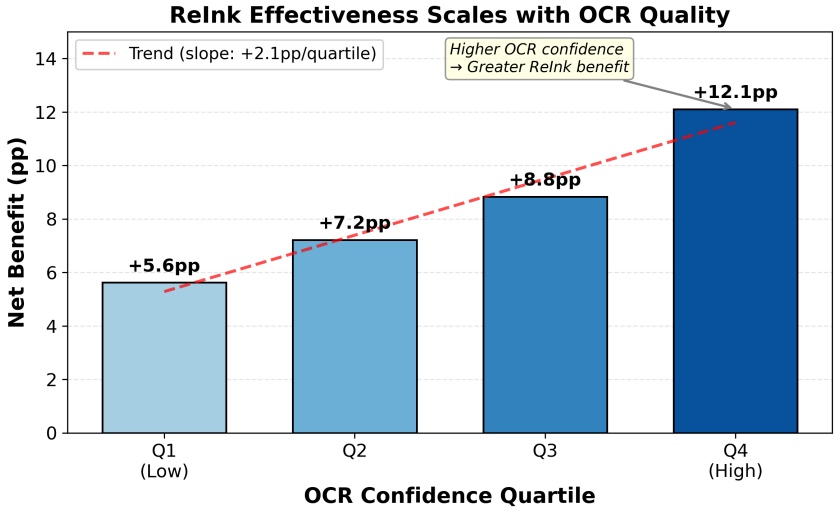

Figure 2: ReInk effectiveness scales with OCR quality. Net benefit (rescued minus hurt questions) increases monotonically from +5.6pp in the lowest OCR confidence quartile to +12.1pp in the highest quartile.

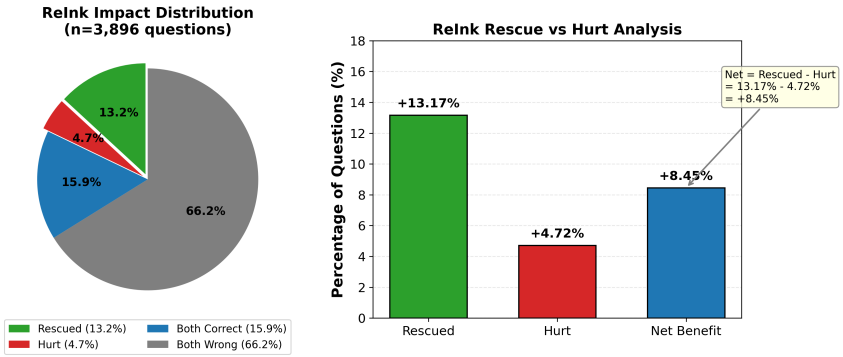

Figure 3: ReInk rescue vs hurt analysis across 3,896 questions. ReInk rescues 13.2% of questions while hurting only 4.7%, yielding a 2.8:1 favorable ratio and +8.45% net benefit.

5 Conclusion

We presented ReInk, a simple, training-free inference wrapper that improves VLM robustness on corrupted charts by providing OCR-extracted text as a spatially-rendered auxiliary image. On ChartQAPro-Corrupted, ReInk achieves +12.21pp improvement over baseline (27.95% vs 15.74%), with gains scaling predictably with OCR quality. Our ablation studies reveal that text content is the primary driver of improvement, while spatial rendering provides marginal additional benefit.

ReInk has limitations: its effectiveness depends on OCR quality, and we evaluated only one VLM on one benchmark with two corruption types. Future work should evaluate across multiple VLMs and corruption types, explore integration with chart-specific models, and investigate whether better OCR systems can further improve robustness.

References

Dan Hendrycks and Thomas G. Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *ArXiv*, abs/1903.12261, 2019.

Ronghang Hu, Amanpreet Singh, Trevor Darrell, and Marcus Rohrbach. Iterative answer prediction with pointer-augmented multimodal transformers for textvqa. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9989–9999, 2019.

Md Farhan Ishmam, Ishmam Tashdeed, Talukder Asir Saadat, Md. Hamjajul Ashmafee, Dr. Abu Raihan Mostofa Kamal, and Dr. Md. Azam Hossain. Visual robustness benchmark for visual question answering (vqa). *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 6623–6633, 2024.

Kushal Kafle, Scott D. Cohen, Brian L. Price, and Christopher Kanan. Dvqa: Understanding data visualizations via question answering. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5648–5656, 2018.

Samira Ebrahimi Kahou, Adam Atkinson, Vincent Michalski, Ákos Kádár, Adam Trischler, and Yoshua Bengio. Figureqa: An annotated figure dataset for visual reasoning. *ArXiv*, abs/1710.07300, 2017.

Fangyu Liu, Julian Martin Eisenschlos, Francesco Piccinno, Syrine Krichene, Chenxi Pang, Kenton Lee, Mandar Joshi, Wenhu Chen, Nigel Collier, and Y. Altun. Deplot: One-shot visual language reasoning by plot-to-table translation. *ArXiv*, abs/2212.10505, 2022.

Ahmed Masry, Do Xuan Long, J. Tan, Shafiq R. Joty, and Enamul Hoque. Chartqa: A benchmark for question answering about charts with visual and logical reasoning. *ArXiv*, abs/2203.10244, 2022.

Ahmed Masry, P. Kavehzadeh, Do Xuan Long, Enamul Hoque, and Shafiq R. Joty. Unichart: A universal vision-language pretrained model for chart comprehension and reasoning. *ArXiv*, abs/2305.14761, 2023.

Ahmed Masry, Mohammed Saidul Islam, Mahir Ahmed, Aayush Bajaj, Firoz Kabir, Aaryaman Kartha, Md Tahmid Rahman Laskar, Mizanur Rahman, Shadikur Rahman, Mehrad Shahmohammadi, Megh Thakkar, Md. Rizwan Parvez, Enamul Hoque, and Shafiq R. Joty. Chartqapro: A more diverse and challenging benchmark for chart question answering. pp. 19123–19151, 2025.

Nitesh Methani, Pritha Ganguly, Mitesh M. Khapra, and Pratyush Kumar. Plotqa: Reasoning over scientific plots. *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1516–1525, 2019.

Srija Mukhopadhyay, Adnan Qidwai, Aparna Garimella, Pritika Ramu, Vivek Gupta, and Dan Roth. Unraveling the truth: Do llms really understand charts? a deep dive into consistency and robustness. *ArXiv*, abs/2407.11229, 2024.

P. W. Shin, Jack Sampson, N. Vijaykrishnan, Andrès Márquez, and M. Halappanavar. Losing the plot: How vlm responses degrade on imperfect charts. *ArXiv*, abs/2509.18425, 2025.

Amanpreet Singh, Vivek Natarajan, Meet Shah, Yu Jiang, Xinlei Chen, Dhruv Batra, Devi Parikh, and Marcus Rohrbach. Towards vqa models that can read. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8309–8318, 2019.

Rahul Thapa, Kezhen Chen, Ian Covert, Rahul Chalamala, Ben Athiwaratkun, S. Song, and James Zou. Dragonfly: Multi-resolution zoom-in encoding enhances vision-language models. 2024.

Dongsheng Wang, Natraj Raman, Mathieu Sibue, Zhiqiang Ma, Petr Babkin, Simerjot Kaur, Yulong Pei, Armineh Nourbakhsh, and Xiaomo Liu. Docllm: A layout-aware generative language model for multimodal document understanding. pp. 8529–8548, 2023.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Ke-Yang Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Yang Fan, Kai Dang, Mengfei Du, Xuancheng Ren, Rui Men, Dayiheng Liu, Chang Zhou, Jingren Zhou, and Junyang Lin. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *ArXiv*, abs/2409.12191, 2024.

Xuezhi Wang, Jason Wei, D. Schuurmans, Quoc Le, Ed H. Chi, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *ArXiv*, abs/2203.11171, 2022.

Zhengzhuo Xu, Bowen Qu, Yiyan Qi, Sinan Du, Chengjin Xu, Chun Yuan, and Jian Guo. Chartmoe: Mixture of diversely aligned expert connector for chart understanding. 2024.