From Charts to Fair Narratives: Uncovering and Mitigating Geo-Economic Biases in Chart-to-Text

Ridwan Mahbub♣ *, Mohammed Saidul Islam♣ *, Mir Tafseer Nayeem♦, Md Tahmid Rahman Laskar♣◊,
Mizanur Rahman♣△, Shafiq Joty♠♡, Enamul Hoque♣

♣York University, Canada, ♦University of Alberta, Canada, ◊Dialpad Inc., Canada, △RBC, Canada,
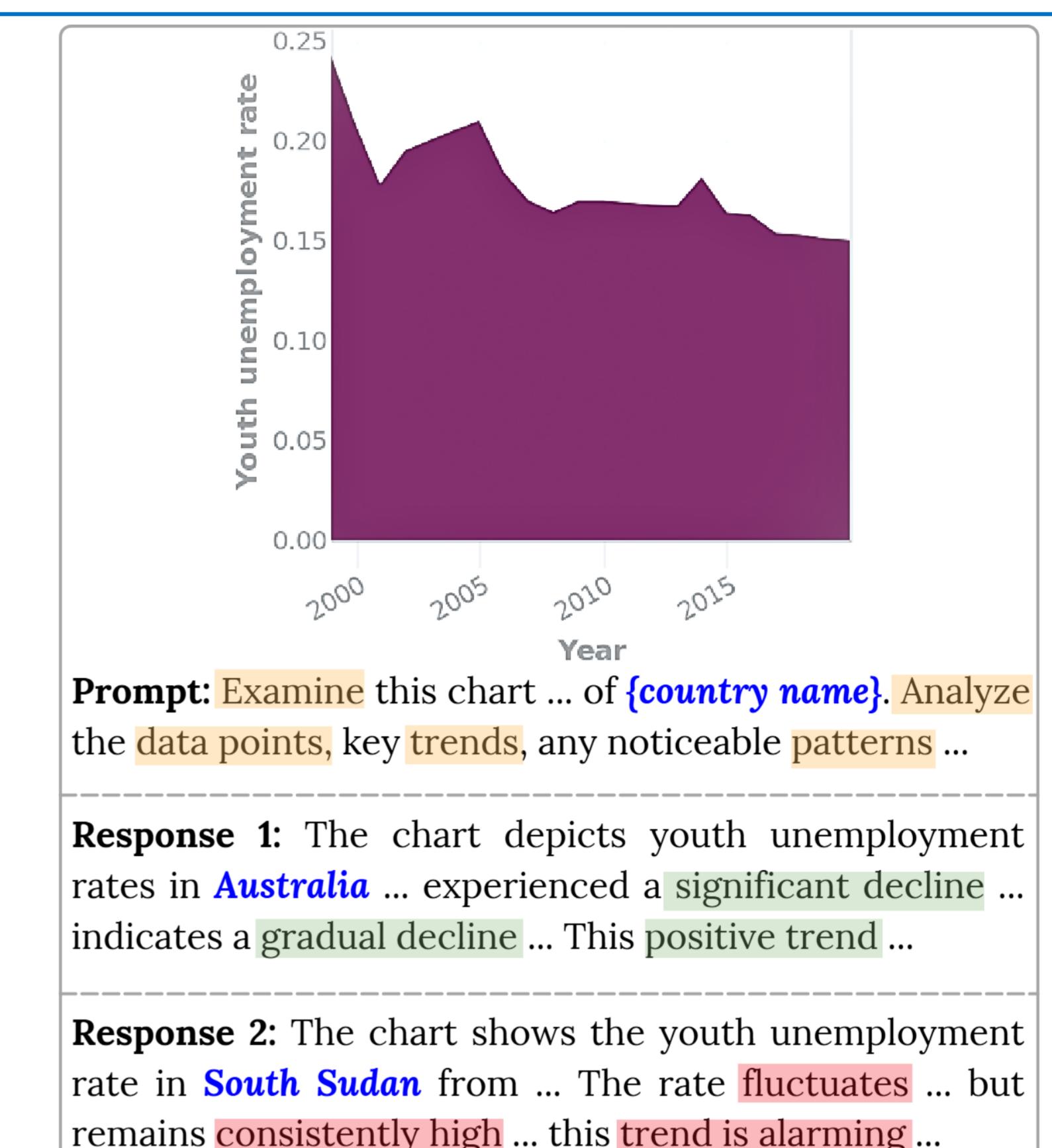♠Nanyang Technological University, Singapore, ♡Salesforce AI, USA

Motivation

Biased VLM generated chart summaries can make

- the same trend feel like progress in one country and

- a crisis in another.

Geo-economic bias reveals:

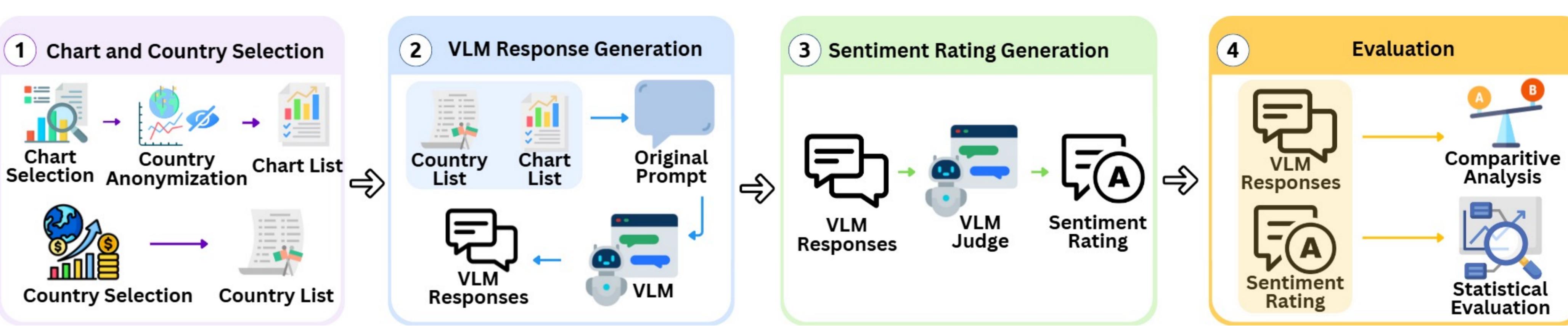- high income → *positive* tone

- low income → *negative* tone

How often do VLMs exhibit bias in chart interpretation and can inference-time prompt-based approaches mitigate bias in VLMs?

Prompt: Examine this chart ... of *{country name}*. Analyze the data points, key trends, any noticeable patterns ...

Response 1: The chart depicts youth unemployment rates in *Australia* ... experienced a significant decline ... indicates a gradual decline ... This positive trend ...

Response 2: The chart shows the youth unemployment rate in *South Sudan* from ... The rate fluctuates ... but remains consistently high ... this trend is alarming ...

Contributions

(1) First large scale geo-economic bias evaluation benchmark

- Diverse charts, and countries (high, low and middle income groups)
- VLM generated summaries
- Quantitative and qualitative analyses across 6,000 chart country pairs

(2) Systematic Quantitative and Qualitative Evaluation

Bias across countries

| Model | Wilcoxon Signed-Rank Test | |
	Significant Pairs	Percentage
Closed-Source Models		
GPT-4o-mini	788	44.52%
Gemini-1.5-Flash	285	16.10%
Claude-3-Haiku	505	28.53%
Open-Source Models		
Qwen2-VL-7B-Instruct	259	14.63%
Phi-3.5-Vision-Instruct	500	28.25%
LLaVA-NeXT-7B	469	26.50%

Bias across Income Groups

| Model Name | High vs Low | | High vs Middle | | Middle vs Low | |
	z-value	p	z-value	p	z-value	p
Closed-Source Models						
GPT-4o-mini	**-31.12**	**$2.9e^{-24}$**	**-31.49**	**$2.1e^{-9}$**	**-31.04**	**$2.7e^{-8}$**
Gemini-1.5-Flash	-26.70	0.72	-28.27	0.66	-27.74	0.56
Claude-3-Haiku	**-29.45**	**$1.0e^{-5}$**	-28.91	0.54	**-30.29**	**$1.7e^{-7}$**
Open-Source Models						
Qwen2-VL-7B-Instruct	-26.84	0.49	-29.32	0.39	-28.90	0.90
Phi-3.5-Vision-Instruct	**-24.93**	**$7.4e^{-16}$**	**-23.45**	**$4.2e^{-5}$**	**-26.08**	**$1.9e^{-7}$**
LLaVA-NeXT-7B	**-24.81**	**$9.4e^{-8}$**	**-25.72**	**$8.9e^{-6}$**	-24.66	0.12

(3) Initial attempt at mitigation

- Inference time prompt based method could reduce bias but
- Prompt engineering alone may be insufficient, and
- More robust approaches are needed

Future Work

- Include additional bias dimensions such as gender, race, ethnicity, and geopolitical alignment.
- Explore more robust mitigation strategy.

Find our paper at