Do Large Language Models Understand Data Visualization Rules?

Martin Sinnona[1,2] Valentin Bonas[1] Emmanuel Iarussi[1,3] Viviana Siless[1]

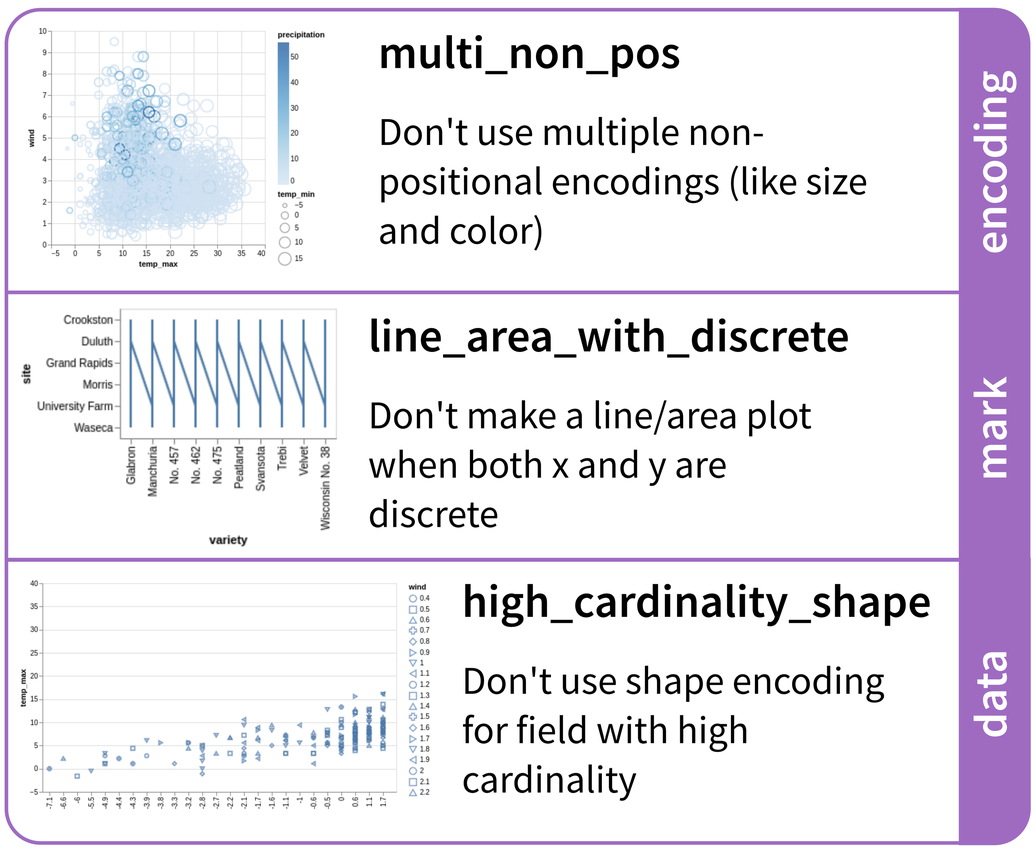
Universidad Torcuato Di Tella, **Argentina**[1] Universidad de Buenos Aires, **Argentina**[2] CONICET, **Argentina**[3]

Motivation

Visualization depends on well-established **design rules**, but **symbolic constraint-based** systems require expert upkeep.

Although **LLMs** can generate charts and detect misleading visuals, it is unclear whether they can **enforce** these **rules** directly from specifications.

Can **LLMs** validate these **rules**? Can they detect solver-verified **violations** in Vega-Lite charts?



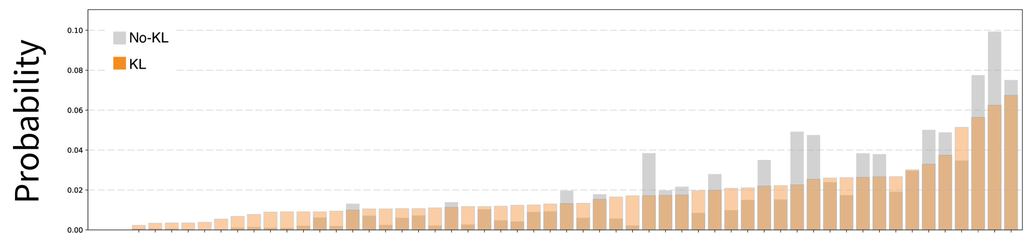
Examples of Visualization Problems

Dataset

Synthetic Dataset: **2,000 Vega-Lite Specifications**

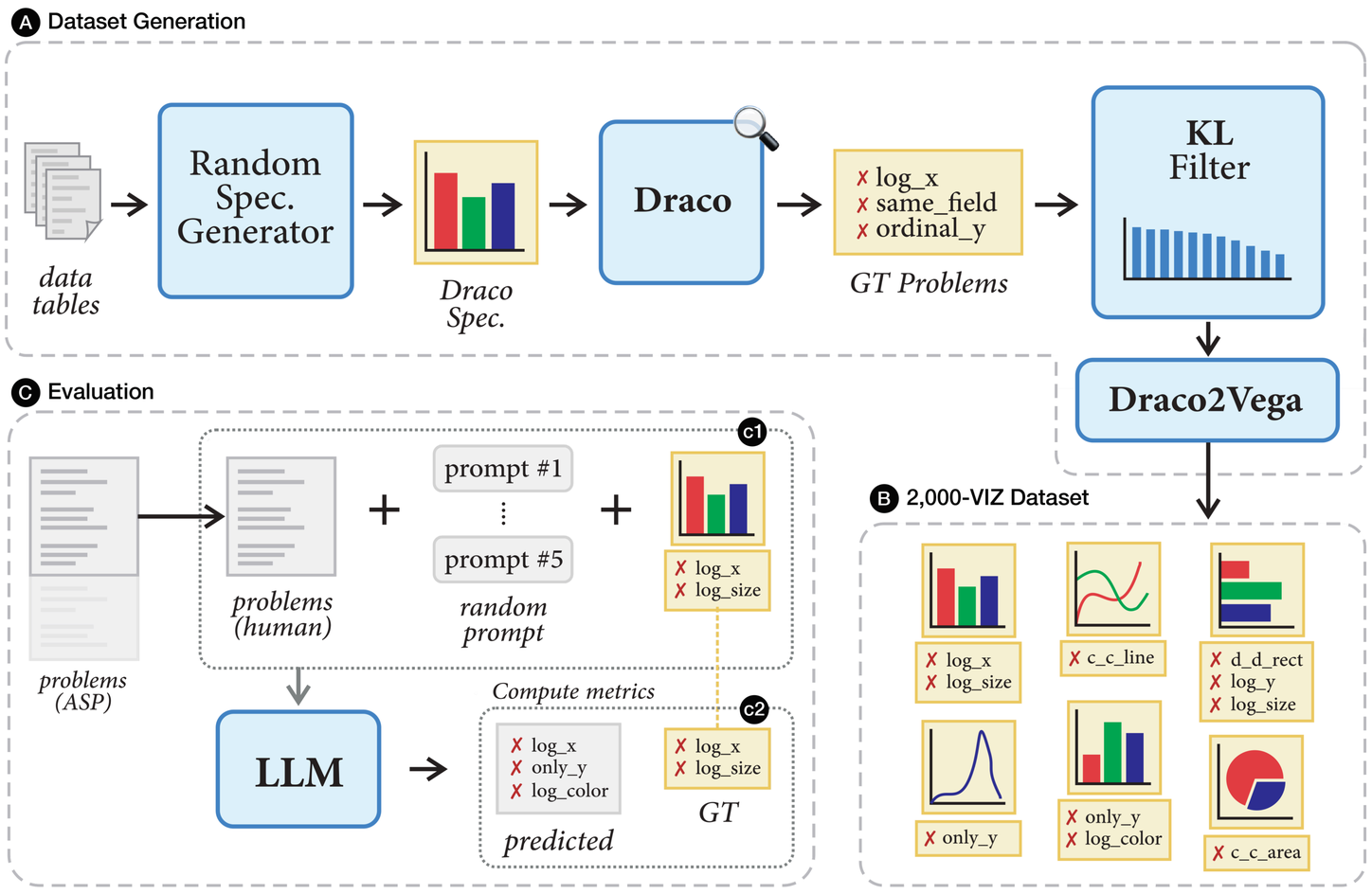
- Randomly generate Draco chart specs
- Use Clingo to identify explicit **rule violations**
- Apply **KL-divergence filtering** → balanced distribution of problem types
- Convert final specs to Vega-Lite

B 2,000-VIZ Dataset



A Distribution of Visualization Problems

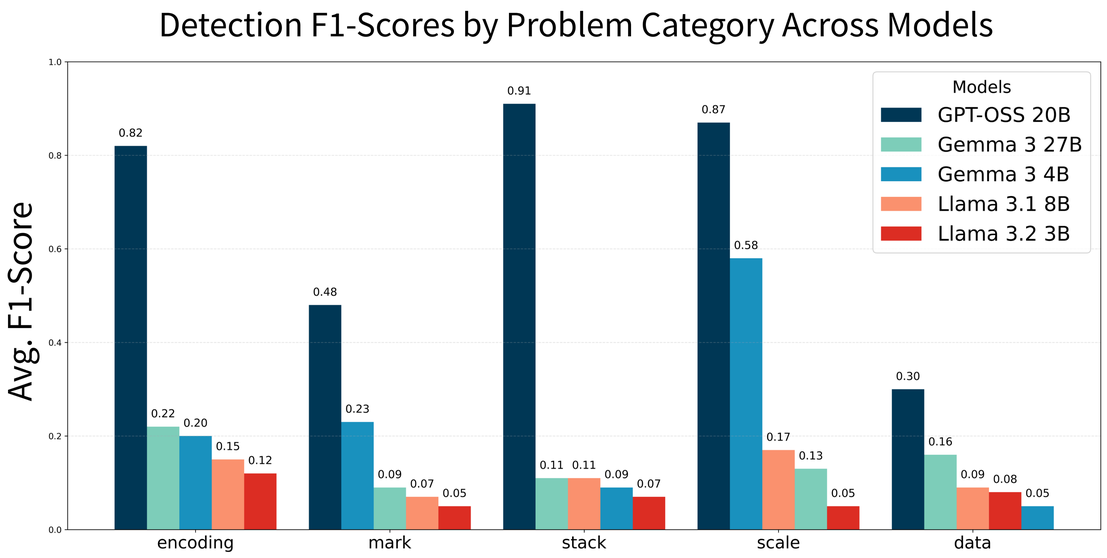
Method

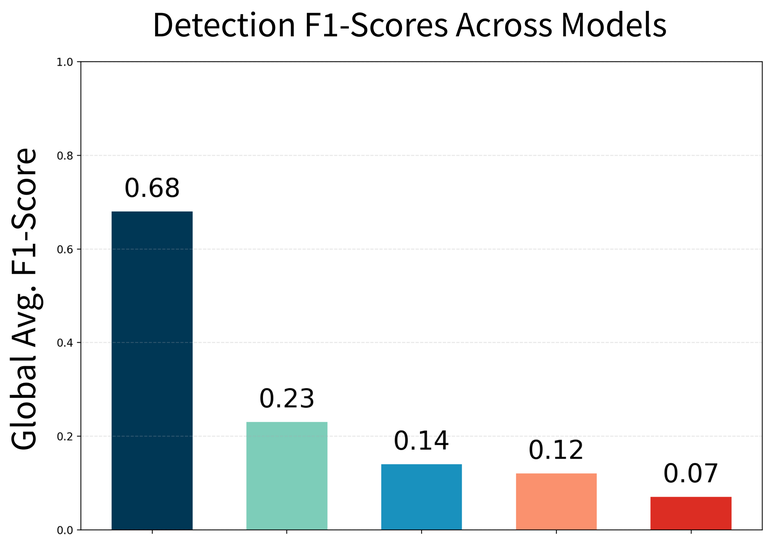


1. Prompt **LLMs**
   - 5 instruction variants
   - 5 repetitions per instance
2. Parse response, validate format
3. Compute metrics
   - **F1-score** (detection accuracy)
   - **Prompt adherence** (JSON validity)

C Evaluation Pipeline

Results





Gemma and **GPT-OSS** reached **100%** adherence. **Llama 3.1 8B** (**65%**) and **Llama 3.2 3B** (**87%**).

Natural-language rules improved performance over **ASP**.

GPT-OSS achieved the highest detection **F1-Score**. **Gemma 27B** followed. **Gemma 4B** outperformed both **Llama** models.