Do Vision-Language Models See Visualizations Like Humans?

Alignment in Chart Categorization

Péter Ferenc Gyarmati[1], Manfred Klaffenböck[2,1], Laura Koesten[3,4,1], Torsten Möller[1]

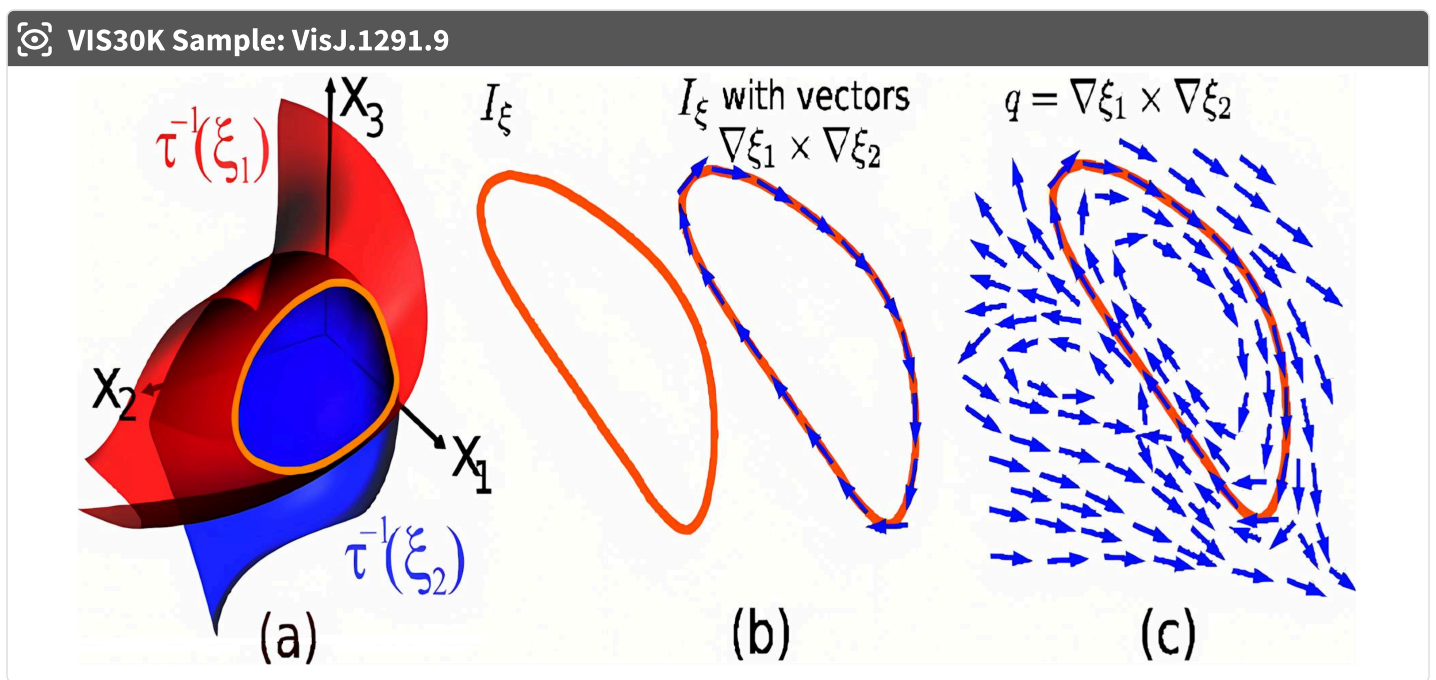
[1]University of Vienna, [2]Vienna University of Technology, [3]Mohamed bin Zayed University of Artificial Intelligence, [4]Austrian Institute of Technology

One Image, Two Realities

Human Expert Perception

Purpose	✓ visualization
Encoding	✓ line ✓ surface-volume ✓ glyph
Dimensionality	✓ 3D

VIS30K Sample: VisJ.1291.9

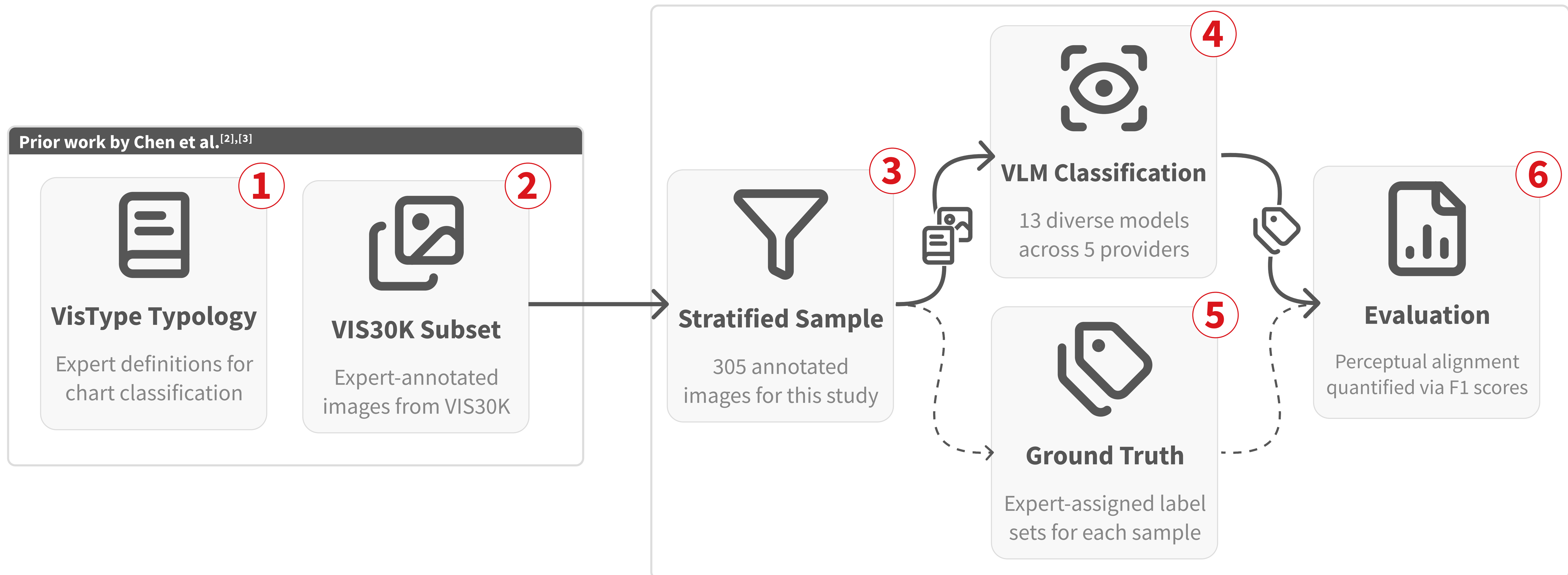


VLM Perception of GPT-4.1

Purpose	✗ schematic
Encoding	✓ line ✓ surface-volume ✗ point
Dimensionality	✗ 2D ✓ 3D

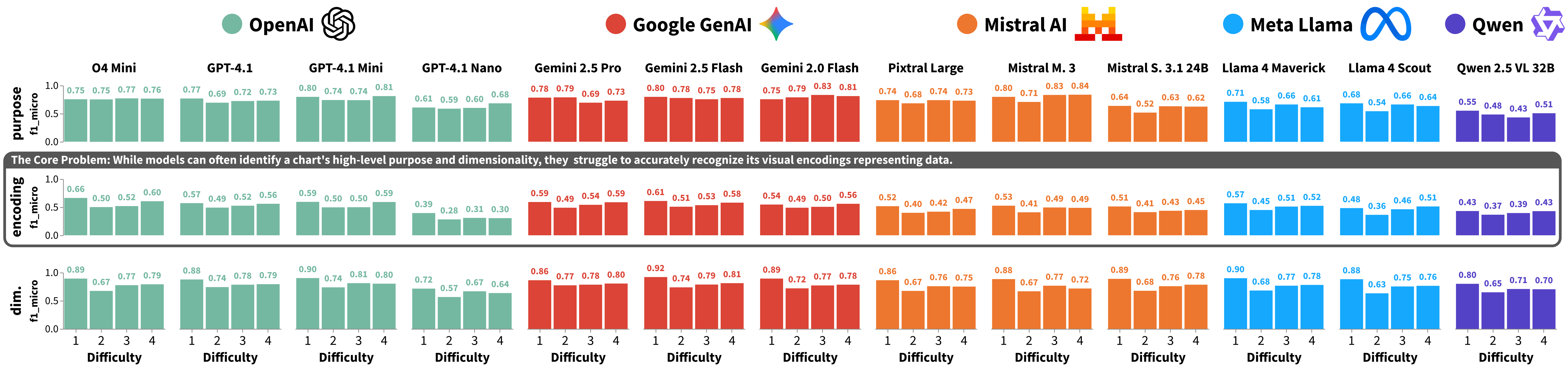
To become a useful partner, an AI must first perceive charts like a human. Here, that foundational skill fails at every level: the figure's **purpose** is misidentified, a core **encoding** is mistaken, and a **dimension** is hallucinated. To build reliable AI, we must first understand this gap, making systematic measurement essential.

Benchmarking Perceptual Alignment



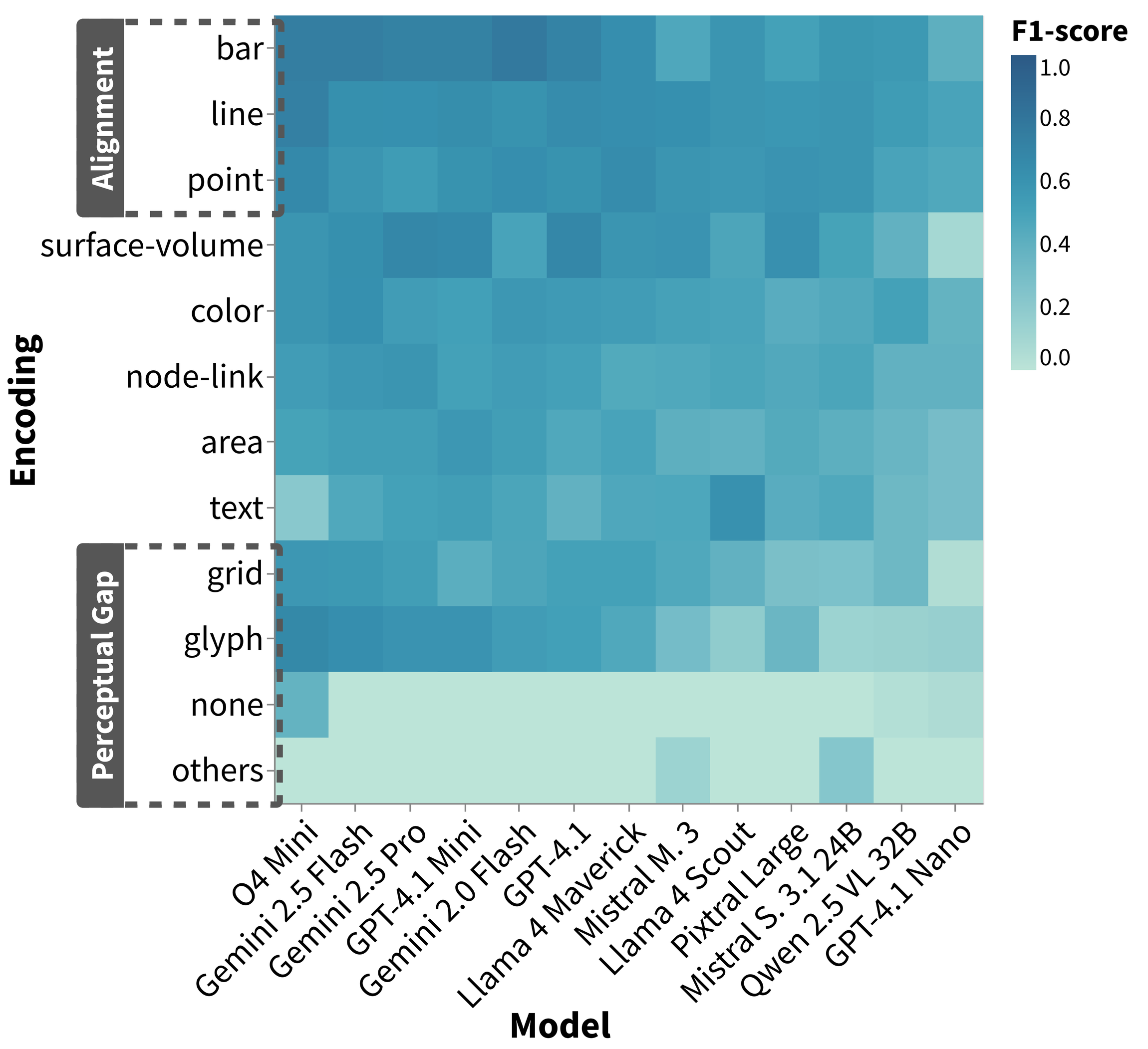
1. **Guiding Framework:** The VisType typology[2] provided a shared, objective classification framework for both human experts and VLMs.
2. **Expert-Annotated Dataset:** A foundational dataset of 6,000+ images from VIS30K[3], annotated by human experts according to the VisType typology[2].
3. **Stratified Sample:** A 305-image stratified sample was drawn from the expert dataset to ensure a balanced and representative test set.
4. **Perceptual Test:** In a strict zero-shot task, 13 diverse VLMs classified each image, guided only by the raw visual input and the VisType[2] definitions provided as their system prompt.
5. **Benchmark:** The expert labels for the 305-image sample—specifying each image's *purpose*, *encoding*, and *dimensionality*—served as the ground truth for evaluating human-VLM alignment.
6. **Alignment Measurement:** Perceptual alignment was quantified by comparing VLM predictions against the ground truth using the multi-label F1-Score.

The Anatomy of Misalignment



Encoding Performance Breakdown

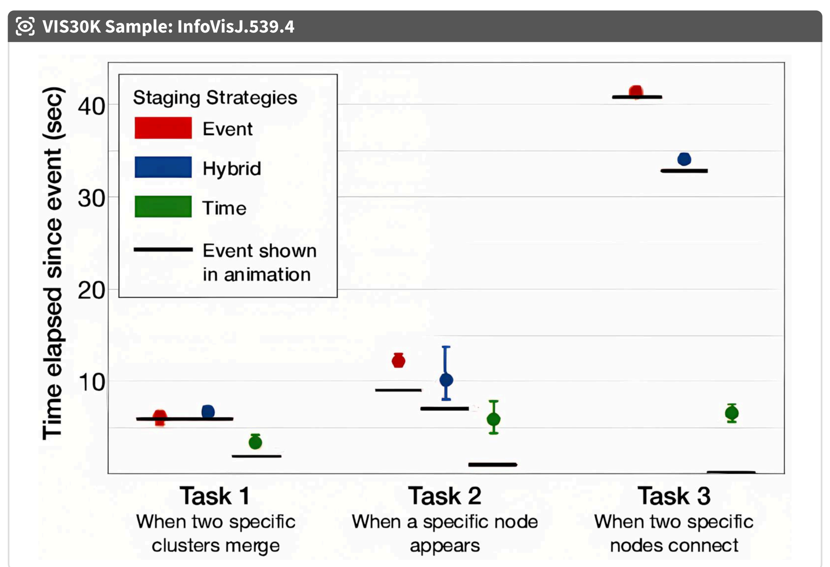
This map of F1 scores reveals a clear performance divide: while alignment is strong for simple encodings, a significant perceptual gap exists for ambiguous types and when human experts abstained from labelling.



A Failure of Consensus

Flagship vision-language models not only fail on the same complex image, but they also disagree on the incorrect alternative, revealing a lack of shared perceptual understanding between VLMs.

VIS30K Sample: InfoVisJ.539.4

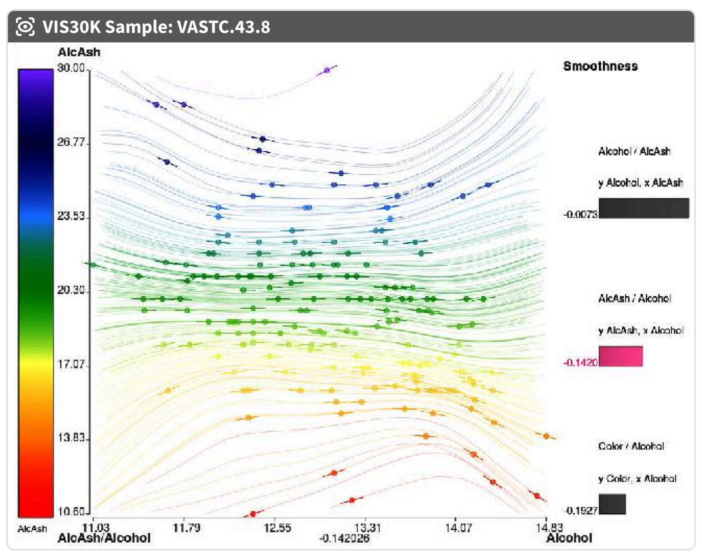


	Perceived Encoding	
Human Expert	✓ bar ✓ point	
Gemini 2.5 Pro	✗ line ✓ point	
Pixtral Large	✗ line ✗ text	✓ point
Qwen 2.5 VL 32B	✗ line ✗ color	✓ point

A Challenge Beyond Scale

This counter-intuitive trend, where smaller models can outperform larger ones on the encoding task, suggests the perceptual gap is a deep architectural challenge, not merely a problem of model scale.

VIS30K Sample: VASTC.43.8



	Perceived Encoding	
Human Expert	✓ color ✓ line	
Llama 4 Scout	✓ color ✓ line	
Llama 4 Maverick	✓ line ✗ point	
Gemini 2.5 Pro	✗ bar ✓ line	✗ point

References

[1] A. Bendeck and J. Stasko. An Empirical Evaluation of the GPT-4 Multimodal Language Model on Visualization Literacy Tasks. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):1105–1115, 2025. doi: 10.1109/TVCG.2024.3456155

[2] J. Chen, P. Isenberg, R. S. Laramee, T. Isenberg, M. Sedlmair, T. Möller, and R. Li. An Image-based Typology for Visualization, 2025. arXiv preprint.

[3] J. Chen, M. Ling, R. Li, P. Isenberg, T. Isenberg, M. Sedlmair, T. Möller, R. S. Laramee, H.-W. Shen, K. Wünsche, and Q. Wang. VIS30K: A Collection of Figures and Tables From IEEE Visualization Conference Publications. *IEEE Transactions on Visualization and Computer Graphics*, 27(9):3826–3833, 2021. doi: 10.1109/TVCG.2021.3054916

[4] J. Hong, C. Seto, A. Fan, and R. Maciejewski. Do LLMs have visualization literacy? an evaluation on modified visualizations to test generalization in data interpretation. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–13, 2025. doi: 10.1109/TVCG.2025.3536358

1 universität wien

2 TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology

3 Mohamed bin Zayed University of Artificial Intelligence

4 AIT AUSTRIAN INSTITUTE OF TECHNOLOGY