

When Is a Glass Full?

An Experiment on Semantic Precision in Large Language Models

Rudolf Stepan

2025

Version 1.0 | First public release

Abstract

This study explores how a large language model (LLM) interprets the concept of 'full' when generating visual content. Although the instruction appears unambiguous to a human observer, the model initially applied a colloquial interpretation rather than an exact physical one. Through the introduction of explicit reference images and iterative corrective feedback, the model's representation of fullness converged toward a technically precise definition. The experiment demonstrates that semantic accuracy in LLM systems is not static but can be actively shaped through interaction, grounding, and constraint reinforcement.

This experiment relates to established discussions on semantic grounding and interpretability in LLM systems *Harnad, 1990; Bender et al., 2021*.

Keywords

LLM semantics, visual grounding, human–AI interaction, interpretability, prompt correction

1. Introduction

Language-based AI systems operate on probabilistic associations learned from vast corpora of human communication. While this enables remarkable fluency, it also means that technical expressions are often interpreted through culturally embedded usage.

2. Initial Incorrect Image Generation

In the first iteration, the model generated images in which the liquid level ended slightly below the rim, revealing reliance on colloquial semantics.

3. Visual Reference and Correction

A reference image depicting a truly filled glass was introduced, resulting in technically correct outputs.

4. Model Self-Diagnosis

The model recognized it had applied an everyday definition rather than a physical one.

5. Generalization Test

The corrected concept remained stable across different liquids.

6. Conclusion

Semantic precision in LLMs improves through grounding and interaction design.

Figures



Figure 1: Incorrect semantic interpretation of fullness (wine).



Figure 2: Second incorrect attempt.



Figure 3: Correct (much better) interpretation after grounding.



Figure 4: Technical correctness verified with milk.

References

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021).
On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.
2. Harnad, S. (1990).
The Symbol Grounding Problem.
Physica D: Nonlinear Phenomena, 42(1-3), 335-346.
3. Radford, A., Kim, J. W., Hallacy, C., et al. (2021).
Learning Transferable Visual Models From Natural Language Supervision.
Proceedings of ICML.
4. Bommasani, R., et al. (2021).
On the Opportunities and Risks of Foundation Models.
Stanford CRFM Report.
5. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017).
Building Machines That Learn and Think Like People.
Behavioral and Brain Sciences.
6. Mitchell, M. (2019).
Artificial Intelligence: A Guide for Thinking Humans.
Farrar, Straus and Giroux.