Qeios

Peer Review

Review of: "Emergent Minkowski-like spaces of many-observers relational event universes"

Emad Ghalenoei[1]

1. Geomatics Engineering, University of Calgary, Canada

The manuscript presents a novel and conceptually ambitious exploration of the latent causal structure in large language models (LLMs), drawing analogies to Minkowski-like spacetime. The authors propose and empirically investigate the existence of a pseudo-geometric structure in the representation space of LLMs that appears to exhibit causal properties akin to those in special relativity. Through synthetic experiments and interpretability tools, they suggest that certain directions in embedding space can be associated with “timelike” and “spacelike” relationships, thereby enabling a causal-like ordering of tokens and potentially revealing deeper organizational principles of LLMs.

Overall, this is a thought-provoking and original piece of work. The analogy to spacetime geometry is intellectually appealing and could open up new ways of reasoning about internal representations and dynamics in LLMs. The authors are commended for their interdisciplinary approach, blending ideas from physics, geometry, and AI interpretability, and for making an effort to validate their claims through controlled experiments and visualization techniques.

That said, several aspects of the manuscript would benefit from clarification and further development:

Conceptual Framing:

While the spacetime analogy is stimulating, it occasionally feels overstretched or loosely defined. Terms such as “causal cone,” “timelike,” and “metric” are imported from physics but not always rigorously adapted to the machine learning context. A clearer definition of the mapping between physical quantities and embedding-space constructs would improve the accessibility and precision of the claims. It may also help to explicitly state the limits of the analogy.

doi.org/10.32388/S8CJ36

Empirical Evidence and Robustness:

The experimental results are intriguing but relatively narrow in scope. Most demonstrations rely on synthetic or toy tasks, which are necessary for control but may not fully reflect the behavior of LLMs in more realistic settings. It would strengthen the paper to provide evidence that the proposed causal-like structures persist (or meaningfully manifest) in real-world language modeling contexts or downstream interpretability scenarios.

Reproducibility and Methodological Detail:

Although the authors describe their procedures at a high level, important implementation details are sometimes omitted or briefly mentioned (e.g., how "metric tensors" are approximated, how causal cones are visualized, or how null-like directions are determined). Providing more methodological transparency, possibly in an appendix or supplementary material, would enhance reproducibility and reader confidence.

Relation to Prior Work:

The authors do cite some related literature on interpretability, linear probes, and geometry in neural networks, but the connection to existing causal representation learning, information geometry, and manifold learning could be elaborated further. Clarifying what is genuinely new in their approach relative to prior notions of "geometry in embeddings" would help position the contribution more firmly.

Writing and Organization:

The paper is generally well-written, but at times it assumes familiarity with both theoretical physics and LLM internals, which may limit accessibility for readers coming from one side of the interdisciplinary divide. Some sections could benefit from reorganization or summarizing key takeaways to guide the reader through the argument more clearly.

Summary

In summary, this manuscript offers an imaginative and potentially impactful perspective on the internal structure of LLMs. While some of the analogies remain speculative, the work stimulates valuable discussion and may inspire further research bridging physics-inspired concepts and deep learning interpretability. I encourage the authors to tighten their definitions, expand empirical validation, and clarify methodology. With these improvements, the paper has the potential to make a unique contribution to our understanding of large-scale neural models.

Declarations

Potential competing interests: No potential competing interests to declare.