The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Language Models Do Not Embed Numbers Continuously (Student Abstract)

Alex Davies*, **Roussel Nzoyem Ngueguin***, **Nirav Ajmeri, Telmo de Menezes E Silva Filho**
University of Bristol, Bristol BS8 1UB, UK
{alexander.davies, rd.nzoyemngueguin}@bristol.ac.uk

Abstract

We evaluate how well large language model embeddings represent continuous numerical values across different precisions and ranges. Using linear models and principal component analysis on models from major providers, we show that while embeddings can reconstruct numbers with high fidelity ($R^2 \geq 0.95$), they introduce substantial noise, with principal components explaining less than 40% of embedding variance. Performance degrades with increasing decimal precision and mixed-sign values, revealing fundamental limitations in how these models encode numerical information.

Code — https://github.com/alexodavies/LanguageNumbers

Introduction

Large Language Models (LLMs), trained on internet-wide data, demonstrate some ability to manipulate numbers and perform arithmetic operations. They are deployed in safety-critical scenarios requiring mathematical reasoning, such as accounting (Yoo 2024), medical calculations (Khandekar et al. 2024), and radiotherapy planning (Wang et al. 2025). These deployments pose serious concerns, emphasizing the need to investigate how LLMs represent numbers. Recent findings have underlined representational space as critical for arithmetic (Zhu, Dai, and Sui 2024). Current literature focuses heavily on integers within specific arithmetic tasks (Kantamneni and Tegmark 2025), finding that numbers are encoded using per-digit circular representations and trigonometric algorithms. While these studies reveal complex geometric structures, they leave a fundamental question unanswered: *do embedding models actually embed continuous values continuously?*

Rather than decoding specific geometric structures, we propose metrics (linear R^2, PCA correlation, explained variance) that directly quantify how well embeddings capture the one-dimensional nature of scalar values. Our finding that explained variance is consistently low reveals that complex representations identified in other studies hinder the ability to represent continuous spaces.

*Student, Primary Author

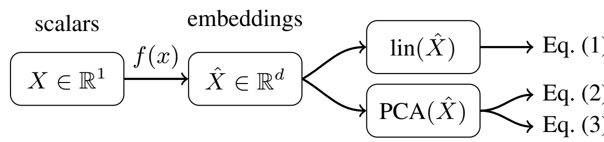

Figure 1: Framework for measuring numerical embeddings.

Metrics and Experiments

Consider a scalar $x \in \mathbb{R}^1$ embedded as $\hat{x} \in \mathbb{R}^d$. For perfect continuous representations, we expect three properties:

Linear Reconstruction Numbers should be reconstructible via linear models:

$$\mathrm{corr}(X', X) \simeq 1, \quad \text{where } X' = \mathrm{lin}(\hat{X}) \tag{1}$$

PCA Direction The primary variation direction should align with numerical ordering:

$$|\mathrm{corr}(\mathrm{PCA}_0, X)| \simeq 1 \tag{2}$$

Explained Variance Given $\mathrm{Rank}(X) = 1$, the first component should explain all variance:

$$\mathrm{VR} = \frac{\lambda_0}{\sum_{i=0}^{d} \lambda_i} \simeq 1 \tag{3}$$

We evaluate three datasets with increasing precision or magnitude: (A) positive decimals $x \in [0, 1]$ with 1 through 20 decimal places, (B) mixed-sign decimals $x \in [-1, 1]$ with same precisions, and (C) mixed-sign integers with increasing magnitudes up to 10^{20}. With these datasets we test the 'flagship' embedding models from three providers: OpenAI, Google, and Voyage AI.

Models maintain $R^2 \geq 0.95$ for positive decimals, confirming numerical information preservation. However, performance degrades significantly with mixed-sign values and higher precision. OpenAI models show sensitivity to decimal precision when negative numbers are included.

For positive integers, embeddings show reasonable numerical ordering in the first component, though the second introduces systematic variation unrelated to magnitude (see Figure 3). For mixed-sign integers, the first principal component encodes sign rather than continuous values, relegating

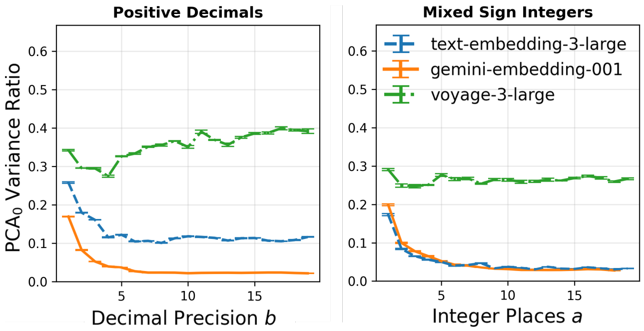

Figure 2: Precision or magnitude plotted against the explained variance ratio of the first principal component.

magnitude to secondary components. Explained variance ratios are consistently low. Even for simple positive decimals, the first principal component explains at most 40% of total variance, indicating the majority of principle components capture information orthogonal to numerical content. Performance degrades further with decimal precision, mixed signs and magnitude. The multi-dimensional patterns identified in other works (Kantamneni and Tegmark 2025) therefore act as noise in representing continuous values.

Provider	Linear R^2	PCA R^2	PCA Variance
+ Decimals			
Google	0.96-1.00	0.04-0.83	0.03-0.20
OpenAI	0.96-1.00	0.46-0.85	0.03-0.17
Voyage	1.00-1.00	0.92-0.97	0.25-0.29
± Decimals			
Google	0.94-1.00	0.01-0.82	0.03-0.14
OpenAI	0.93-1.00	0.72-0.79	0.15-0.23
Voyage	0.99-1.00	0.73-0.80	0.32-0.40
± Integers			
Google	0.48-0.99	-0.03-0.72	0.02-0.10
OpenAI	0.91-1.00	0.69-0.84	0.10-0.19
Voyage	0.95-1.00	0.73-0.83	0.28-0.44

Table 1: Performance metrics across datasets (Min-Max)

Our results reveal a fundamental trade-off in current embedding models: while numerical information is preserved sufficiently for linear reconstruction, embeddings introduce substantial noise that obscures numerical structure. The finding that mixed-sign representations encode sign in the first principal component is particularly concerning, with models representing numbers in terms of signs and values, instead of one continuous space. This leads us to several practical implications: **1.** Rounding numbers to fewer decimal places may improve downstream numerical task performance **2.** Applications involving both positive and negative values should expect degraded embedding quality **3.** Current embedding models may not be optimal for applications requiring precise numerical relationships.

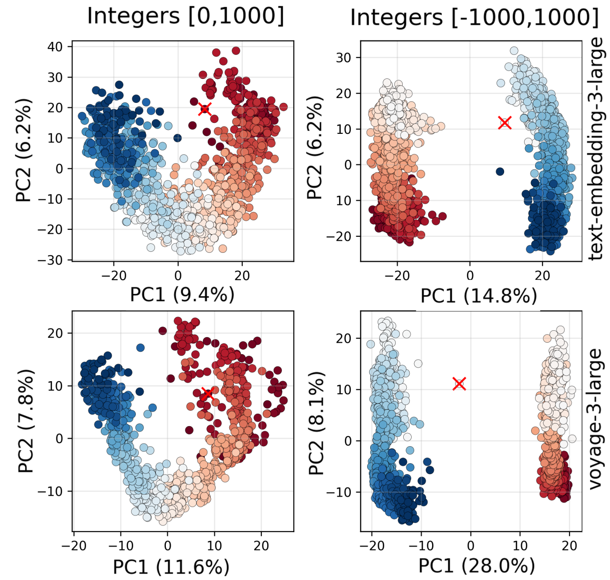

Figure 3: PCA_1 and PCA_2 for $x \in [0, 1k]$ and $x \in [-1k, 1k]$. Colour indicates integer value, red lower, blue higher.

Conclusion

While embedding models can preserve numerical information for reconstruction, they fail to represent numbers as truly continuous values. The majority of embedding dimensions encode information unrelated to numerical content, and the sign of a number is a separate feature.

Our lightweight framework provides a foundation for evaluating numerical representations in embedding models, addressing a critical gap between complex geometric structures identified in prior work and the fundamental requirement for continuous numerical understanding. Future work should focus on specialized architectures that better isolate numerical information from other sources of variation.

References

Kantamneni, S.; and Tegmark, M. 2025. Language models use trigonometry to do addition. *arXiv:2502.00873*.

Khandekar, N.; Jin, Q.; Xiong, G.; Dunn, S.; Applebaum, S.; Anwar, Z.; Sarfo-Gyamfi, M.; Safranek, C.; Anwar, A.; Zhang, A.; et al. 2024. Medcalc-bench: Evaluating large language models for medical calculations. *Advances in Neural Information Processing Systems*, 37: 84730–84745.

Wang, Q.; Wang, Z.; Li, M.; Ni, X.; Tan, R.; Zhang, W.; Wubulaishan, M.; Wang, W.; Yuan, Z.; Zhang, Z.; et al. 2025. A feasibility study of automating radiotherapy planning with large language model agents. *Physics in Medicine & Biology*, 70(7): 075007.

Yoo, M. 2024. How Much Should We Trust LLM-Based Measures for Accounting and Finance Research? *SSRN*.

Zhu, F.; Dai, D.; and Sui, Z. 2024. Language models know the value of numbers. *arXiv:2401.03735*.