SIM: Statically Implied Models for Transparent and Interoperable Embedding of Spatiotemporal Knowledge

A Foundation for Prompt-Readable, Vector-Native Reasoning

Sean Tucker

Abstract

We introduce **SIM (Statically Implied Models)**, a new class of deterministic, interpretable vector embedding systems designed for spatial-temporal knowledge representation and reasoning. Unlike learned embeddings, SIMs are algebraically structured, reproducible, and prompt-compatible. They support native vector math, symbolic decomposition, and transparent inference across infrastructure and planning domains. We formally define the properties of SIMs, motivate their need in geospatial and infrastructure reasoning, and introduce **STEM** as the first canonical SIM instance ("SIM-0"). SIMs enable what we call **Prompt-Readable Vector Logic**—the ability for both humans and language models to interpret and reason over embeddings directly, without opaque training. We conclude by previewing **MOSES**, a modular GeoAI reasoning system that composes SIMs into live vector-native pipelines.

---

1. Introduction

Most spatial embeddings today—whether derived from transformers (e.g., GeoBERT) or graph models—are opaque, non-reproducible, and difficult to interpret [Bommasani et al., 2021; Mai et al., 2022]. This limits their use in high-stakes reasoning domains such as infrastructure, transit, disaster planning, or regulatory systems, where **traceability and symbolic compatibility** are essential [Cai & Wang, 2020].

We propose a new embedding model class: **Statically Implied Models (SIMs)**.
SIMs encode structured domain knowledge—such as geospatial, ontological, or temporal information—into **deterministic, algebra-compatible vectors**. These vectors can be:

- Directly compared
- Algebraically manipulated
- Referenced in prompts
- Used across systems without retraining

SIMs restore **interpretability, interoperability, and control** to AI-based spatial reasoning, creating a new interface between vector logic and symbolic inference [Bengio et al., 2021; Garnelo & Shanahan, 2019].

To our knowledge, no prior work has proposed a generalized, deterministic, and interpretable vector embedding *model class*—one that supports algebraic manipulation, cross-domain integration, and symbolic reasoning. While individual components of SIMs resemble ideas from positional encodings [Vaswani et al., 2017], geospatial reference systems [Kuhn, 2003], and symbolic AI frameworks [Bengio et al., 2021], SIMs unify these strands into a coherent and reusable framework. By introducing a model-theoretic foundation for **statistically implied embeddings**, SIMs offer a missing bridge between structured domain logic and vector-native AI systems—enabling direct reasoning across space, time, type, and function.

SIMs are not proposed as a replacement for learned semantic embeddings or statistical models in domains where abstraction, generalization, or high-dimensional concept learning is critical. Rather, SIMs are designed to **complement** such systems—especially in settings where **static knowledge**, **structured environments**, or **infrastructure constraints** demand referential clarity and logical accountability. When integrated into AI architectures such as **vector databases with RAG**, SIMs act as stable anchors that allow learned models to reason over interpretable, domain-grounded vectors. This positions SIMs as a **missing substrate for any statically-influenced systems**, such as **spatially aware, symbol-compatible AI**—filling the gap between ontologies and neural reasoning.

---

2. Defining SIMs

We define a **Statically Implied Model (SIM)** as an embedding model that satisfies the following properties:

Property	**Description**
Determinism	Same input always yields the same output vector
Decomposability	Components have known, interpretable meanings
Algebra-Compatibility	Supports vector math (e.g., differences, scaling)
Cross-Domain Reasoning	Operable across ontological/geospatial boundaries

Property	Description
Prompt-Readable	Vector operations can be referenced in LLM prompts or symbolic logic

This sets SIMs apart from learned vector models that are dense, opaque, and domain-fixed [Marcus, 2020].

---

3. SIMs and Prompt-Readable Vector Logic

SIMs support a new interaction mode for LLMs and AI systems:

Prompt-Readable Vector Logic — vectors whose meaning and operations can be directly referenced in natural language reasoning [Kojima et al., 2022; Liu et al., 2021].

Examples:

- “What’s the elevation difference between location A and B?”
- “Which areas are within 500m and 15 minutes of a flood zone?”
- “Return all coffee shops uphill from the station and open now.”

SIMs enable **symbolic + vector fusion**—a language model can reason about vectors **arithmetically**, **spatially**, or **relationally**, using interpretable components.

This makes SIMs usable in:

- Prompt injection pipelines
- Function-calling logic chains
- RAG frameworks over spatial vector databases [Tucker, 2024].

---

4. Case Study: STEM as SIM-0

We introduce **STEM (Spatiotemporal Embedding Model)** as the first SIM implementation [Tucker, 2025b].

Given inputs:

- Latitude
- Longitude

- Altitude
- Time of day

STEM produces a vector of the form:

STEM(lat,lon,alt,t)=[sin(lat),cos(lat),sin(lon),cos(lon),alt/h,sin(2πt/T),cos(2πt/T)]

This vector is:

- Reproducible and interpretable
- Sensitive to time and elevation
- Usable in both vector databases (e.g., Weaviate [Yan et al., 2021]) and AI prompts

Use case:

A transit planner queries “stations within 15 minutes and uphill of any Zone 1 flood zone.” Using STEM vectors, the system computes this without GIS layers—via **vector proximity and delta-altitude math**.

We define STEM as **SIM-0**: the first canonical, fully interoperable spatiotemporal SIM [Mai et al., 2020].

---

5. Comparison: SIMs vs Learned Embeddings

Capability	SIM (e.g., STEM)	Learned Embeddings (e.g., GeoBERT)
Deterministic	✅	❌
Interpretable	✅	❌
Algebraic	✅	⚠️ Partial
Prompt-Readable	✅	❌
Cross-domain	✅	⚠️ Limited
Reusable across systems	✅	❌

SIMs are ideal where **transparency, control, and reproducibility** are more important than statistical abstraction—such as public infrastructure, environmental systems, or AI explainability [Goodchild, 2009; Janowicz et al., 2020].

---

6. Toward MOSES: Modular Reasoning with SIMs

In upcoming work, we present **MOSES**: the **Modular Onto-logic Spatial Embedding System** [Tucker, 2025c].

MOSES composes multiple SIMs into a **live AI reasoning engine**:

- Uses STEM for space-time anchoring
- Adds GPLE and GOE for place logic and type embeddings [Smith & Mark, 2001; Kuhn, 2003]
- Integrates with vector search (e.g., Weaviate [Yan et al., 2021]) and LLMs
- Enables symbolic RAG with vector-native logic [Modi & Titov, 2021]

MOSES represents the system-level integration of SIM logic for real-time infrastructure applications—an executable architecture for modular GeoAI reasoning [Anselin, 2019].

---

7. Conclusion

SIMs represent a shift in AI vector modeling—from black-box embeddings to **deterministic, algebra-compatible, and interpretable representations**. They provide a foundation for **modular, reusable reasoning systems**, and enable **Prompt-Readable Vector Logic** for transparent and traceable AI.

We position SIMs as the beginning of a new class of AI infrastructure—designed not just to encode meaning, but to make that meaning accessible, composable, and accountable.

References

- Anselin, L. (2019). *Spatial data science: Contributions and challenges*. Spatial Economic Analysis, 14(1), 11–16.
- Bengio, Y. et al. (2021). *A review of symbolic inductive bias in deep learning*. Nature Machine Intelligence, 3(10), 845–855.
- Bommasani, R. et al. (2021). *On the Opportunities and Risks of Foundation Models*. arXiv:2108.07258.
- Cai, L., & Wang, J. (2020). *A survey of explainable AI: Interpretability, fairness, and robustness*. arXiv:2006.11371.
- Garnelo, M., & Shanahan, M. (2019). *Reconciling deep learning with symbolic reasoning: Representing objects and relations*. Current Opinion in Behavioral Sciences, 29, 17–23.
- Goodchild, M. F. (2009). *Geographic information systems and science: today and tomorrow*. Annals of GIS, 15(1), 3–9.
- Janowicz, K. et al. (2014). *Five stars of linked data vocabulary use*. Semantic Web, 5(3), 173–176.
- Janowicz, K. et al. (2020). *GeoAI: Spatially Explicit Artificial Intelligence Techniques for Geographic Knowledge Discovery and Beyond*. IJGIS.
- Kojima, T. et al. (2022). *Large Language Models are Zero-Shot Reasoners*. arXiv:2205.11916.
- Kuhn, W. (2003). *Semantic reference systems*. International Journal of Geographical Information Science, 17(5), 405–409.
- Liu, P. et al. (2021). *Pre-train Prompt: Deep Transfer Learning by Pre-Training Prompts*. arXiv:2107.13586.
- Mai, G. et al. (2020). *Multi-scale representation learning for spatial feature distributions using grid cells*. AAAI.
- Mai, G., Janowicz, K., & Yan, B. (2022). *Towards a theory of spatial information observability for deep learning*. Transactions in GIS, 26(2), 867–884.
- Marcus, G. (2020). *The next decade in AI: Four steps toward robust artificial intelligence*. arXiv:2002.06177.

- Modi, A., & Titov, I. (2021). *Modular Approaches to Learning Meaningful Representations*. ACL.
- Smith, B., & Mark, D. M. (2001). *Geographic categories: An ontological investigation*. IJGIS, 15(7), 591–612.
- Tucker, S. (2024). A systematic review of geospatial location embedding approaches in large language models: A path to spatial AI systems. arXiv:2401.10279.
- Tucker, S. (2025b). STEM: A Deterministic Spatiotemporal Embedding Model for Infrastructure-Aware Reasoning. arXiv Preprint.
- Tucker, S. (2025c). MOSES: A Modular Onto-logic Spatial Embedding System for Reasoning with SIMs. In preparation.
- Yan, B., Mai, G., & Janowicz, K. (2021). *Vector-based geographic information systems (V-GIS): A conceptual framework*.