Geometric Manifold Alignment for Sparse B2B Demand Forecasting: A Pattern-Based Approach

Pit Ogermann

pit.ogermann@gmail.com

February 15, 2026

Abstract

The digitization of Business-to-Business (B2B) commerce, particularly within the specialized ecosystem of the German Mittelstand, presents unique predictive challenges that conventional forecasting methodologies fail to address. Unlike data-rich B2C environments, B2B transactions are characterized by extreme data sparsity, intermittent demand, and complex functional interdependencies between Stock Keeping Units (SKUs). This paper introduces a novel Foundation Model architecture centered on **ALiT (Alignment via Linear Transformation)**. By shifting from token-based inductive biases to pattern-based geometric alignment, we demonstrate how Distributional Semantics and Orthogonal Procrustes Analysis can overcome the "Cold Start" problem in sparse datasets. Furthermore, we provide a critical analysis of current State-of-the-Art architectures, specifically distinguishing our autoregressive approach from frequency-enriched parallel decoding models (e.g., SAFERec), arguing that the latter fails to capture the complementary logic essential for industrial baskets.

1 Introduction

The application of Deep Learning to recommender systems has traditionally been driven by the needs of B2C giants (e.g., e-commerce, streaming). These environments are defined by dense user-item interaction matrices and high-frequency consumption. However, the industrial B2B sector—typified by the German "Hidden Champions"—operates under a fundamentally different paradigm.

We identify three critical pathologies in B2B data that render standard "End-to-End" Deep Learning ineffective:

1. **Extreme Sparsity:** An SME may have thousands of technical SKUs but only a few hundred clients, making the interaction matrix too sparse for neural networks to converge on stable embeddings from random initialization.

2. **Disjoint Vocabularies:** Unlike Natural Language Processing (NLP) where the token "dog" is universal, B2B identifiers are proprietary. This prevents traditional Transfer Learning.

3. **Functional Complementarity:** B2B baskets are logical assemblies (e.g., a pump and its specific seal), not merely preferential collections.

1.1 The German Mittelstand Paradox

A pertinent question arises: given the obvious efficiency gains of AI, why has the German Mittelstand been slow to adopt these technologies? Research suggests a "Pressure Paradox" [7]. These "Hidden Champions" often hold near-monopolies in niche hardware markets. This dominance has historically shielded them from the existential market pressures that force rapid digital adaptation in B2C sectors [8]. Consequently, there is little internal pressure to pool data or modernize legacy ERP systems. Standard AI solutions that require centralized data lakes are often rejected due to strict data sovereignty concerns. Our proposed **Pattern-Based Foundation Model** addresses this specific barrier by enabling transfer learning via geometric abstraction, respecting the sovereign "silos" of the Mittelstand while leveraging collective intelligence.

2 Theoretical Framework

2.1 The Failure of Random Initialization

Standard sequential recommenders, such as SASRec [5], initialize item embeddings randomly ($E \sim \mathcal{N}(0, \sigma)$) and rely on Backpropagation to learn semantic proximity. In B2B settings, where an item may appear in only a handful of sequences, the gradient signal is too sparse. The model essentially "memorizes" noise rather than learning structure, leading to severe overfitting.

2.2 Distributional Semantics & PMI

To solve the sparsity issue, we ground our approach in the Distributional Hypothesis [1]. Rather than learning from scratch, we explicitly calculate global corpus statistics using Pointwise Mutual Information (PMI).

Let w be a product and c be a context product in a basket. The PMI is defined as:

$$PMI(w, c) = \log \frac{P(w, c)}{P(w)P(c)} \tag{1}$$

In industrial contexts, purchasing behaviors drift over time due to technological obsolescence. Therefore, we introduce a **Time-Decayed Co-occurrence Matrix**. The weight of an interaction between items i and j is scaled by an exponential decay function:

$$W_{i,j} = \sum_k e^{-\lambda \Delta t_k} \tag{2}$$

where Δt_k is the time elapsed since the transaction. This ensures the model captures the *current* technological landscape rather than historical artifacts.

3 The ALiT Architecture

The core innovation of the proposed Foundation Model is the **ALiT (Alignment via Linear Transformation)** mechanism. This approach moves beyond "tokens" (which are disjoint across companies) to "patterns" (which are universal).

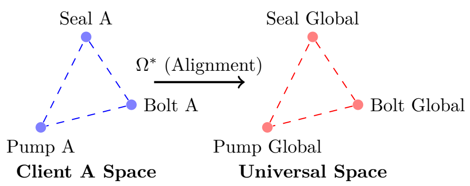

Figure 1: **Manifold Alignment.** Distinct vocabularies are aligned geometrically. The structural relationship between a Pump and a Seal is invariant across companies, allowing transfer learning via rotation Ω^*.

3.1 Geometric Embedding Construction

We utilize Singular Value Decomposition (SVD) on the smoothed PMI matrix to generate dense vector representations.

$$M \approx U \Sigma V^\top \tag{3}$$

Crucially, we employ *Eigenvalue Weighting* ($U\Sigma^{0.5}$), creating a symmetric factorization that balances the representation of high-frequency "head" items and low-frequency "tail" items.

3.1.1 North Star Attributes

To stabilize the geometry of the embedding space across different clients, we augment the latent vectors with **Universal Anchors**, or "North Star" attributes. These are statistical properties independent of product identity:

- **Global Popularity (*pop*):** The normalized frequency of the item.
- **Re-buy Entropy (H):** A measure of the regularity of purchase intervals.
- **Loop Coefficients:** The tendency of an item to be re-ordered immediately.

These scalars are concatenated to the SVD vectors. They act as "lighthouses" during the alignment process; even if "Pump A" and "Pump B" have different IDs, their high popularity and low entropy allow the model to recognize them as functionally equivalent nodes in the manifold.

3.2 The Point Alignment Head

A full Procrustes alignment requires matching all points A to B. In a live inference setting, this is computationally expensive and noisy. We introduce the **Point Alignment Head** based on the "Sensor Hypothesis." Instead of aligning the entire catalog, we sample a small, statistically representative subset of items (the "Sensor," $S \subset A$). A lightweight attention mechanism computes the optimal rotation matrix T based solely on this sensor:

$$T = \text{Attention}(S_{sensor}) \tag{4}$$

This allows the model to orient a new client's dataset in milliseconds, effectively "snapping" their proprietary data into the universal orientation of the Foundation Model.

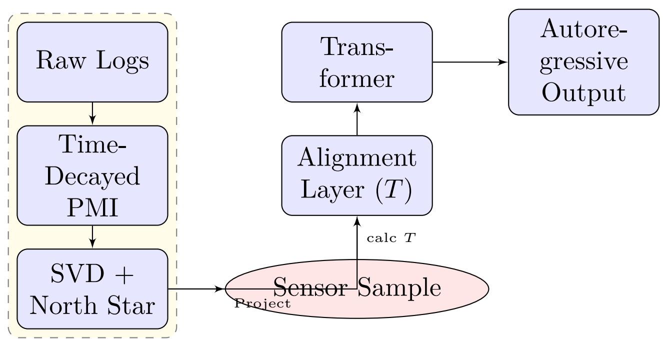

Figure 2: **System Architecture.** The pipeline decouples structural learning (PMI/SVD) from dynamic learning (Transformer). The Alignment Layer uses a "Sensor" sample to project proprietary data into the universal space.

4 The Transformer Paradigm: Attention without Representation

The seminal paper "Attention is All You Need" [4] introduced the Self-Attention mechanism:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \quad (5)$$

In standard implementations, the model must simultaneously learn the representation of the items (the embeddings) and the dynamics of their interaction (the weights W_Q, W_K, W_V). Our architecture fundamentally alters this contract. By fixing the item representations via ALiT (which are already semantically rich due to SVD), the Transformer is relieved of the burden of *representation learning*. It focuses exclusively on *sequence dynamics*. This explains why our model converges on sparse B2B data where others fail: the "geometry" of the products is pre-calculated using rigorous algebra (SVD), not learned via stochastic gradient descent. The Transformer effectively acts as a "physics engine," learning the universal laws of how industrial baskets evolve, while the ALiT layer ensures the "objects" in the simulation are correctly defined.

5 Sequence Modeling & Inference

5.1 Autoregressive vs. Parallel Decoding

A critical distinction of our architecture is the adherence to **Autoregressive Inference** for basket generation. Recent State-of-the-Art models, such as SAFERec [6], employ parallel decoding to maximize inference speed. SAFERec calculates probabilities for all items simultaneously:

$$P(Basket|History) \approx \prod_{i \in Basket} P(i|History) \quad (6)$$

This assumes *conditional independence* between items in the target basket. While effective for grocery recommendations, this is catastrophic for industrial B2B. In B2B, items are **complements**. If a "Drill Body" is selected, the probability of "Drill Bit" must increase, and the probability of a rival "Drill Body" must decrease. Our model generates baskets sequentially:

$$P(i_t|i_{t-1}, ..., i_0, History) \quad (7)$$

This ensures that the predicted basket is a coherent technical assembly, not a list of redundant substitutes.

5.2 Hard Constraints & Frequency

To further enforce logical consistency, the inference engine applies hard constraints during beam search to prevent the repetition of capital goods within a single basket, while allowing high-frequency consumables (bolts, lubricants) based on their learned entropy in the ALiT embedding space.

6 Conclusion

This whitepaper outlined a custom AI Foundation Model designed specifically for the data realities of the B2B sector. By combining Time-Decayed PMI with Orthogonal Procrustes Alignment, we solve the cold-start problem inherent in sparse industrial datasets. Furthermore, our critique of parallel decoding models like SAFERec highlights the necessity of autoregressive coherence in technical procurement. This architecture provides a scalable, privacy-preserving pathway for German SMEs to leverage collective industrial intelligence.

References

[1] O. Levy and Y. Goldberg, "Neural word embedding as implicit matrix factorization," in *Advances in neural information processing systems*, pp. 2177–2185, 2014.

[2] O. Levy, Y. Goldberg, and I. Dagan, "Improving distributional similarity with lessons learned from word embeddings," in *Transactions of the Association for Computational Linguistics*, vol. 3, pp. 211–225, 2015.

[3] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[5] W.-C. Kang and J. McAuley, "Self-attentive sequential recommendation," in *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 197–206, IEEE, 2018.

[6] D. Lashinin et al., "Saferec: Self-attention and frequency enriched model for next basket recommendation," *arXiv preprint arXiv:2412.14302*, 2024.

[7] H. Simon, "Hidden Champions of the Twenty-First Century: The Success Strategies of Unknown World Market Leaders," *Springer Science & Business Media*, 2009.

[8] Bertelsmann Stiftung, "The German Mittelstand and Digital Transformation," *Policy Brief*, 2018.