International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

Leveraging AI for Data Mapping in Life Insurance System Conversions

Anshul Kaushik

Abstract: ***Data conversion in the life insurance industry is a complex and time-intensive process due to the long history of legacy systems, product variations, and nuanced regulatory requirements. Traditional data mapping methods require significant manual effort, making large-scale system conversions both risky and inefficient. This paper presents a novel approach that leverages artificial intelligence (AI), specifically vector search powered by fine-tuned language models, to streamline the data mapping process. By developing a custom AI model tailored to life insurance terminology and workflows, and integrating vector embeddings for semantic field matching, we achieved a 70% reduction in manual mapping time. The results indicate strong potential for AI to accelerate digital transformation in insurance by simplifying one of its most resource-intensive components.***

Keywords: Life Insurance, Data Mapping, System Conversion, Artificial Intelligence, Vector Search, BERT, Word2Vec, GloVe, Semantic Matching, Data Migration, Insurance Technology, Legacy Systems, Prompt Engineering, Fine-tuned Language Models

1. Introduction

The life insurance industry is undergoing a significant digital transformation, driven by the need for agility, regulatory compliance, and enhanced customer experience. However, one of the most complex challenges in this transformation is the migration of data from legacy systems to modern core platforms. Life insurance products are inherently intricate, involving long-term policies, multi-tiered contractual structures, actuarial calculations, and strict regulatory oversight—all encoded within decades-old systems.

According to industry research, over 70% of global insurers continue to rely on at least one legacy system, typically using outdated technologies that hinder data interoperability. As organizations move toward modern systems, accurate data mapping becomes a mission-critical task. It involves aligning thousands of fields across vastly different schemas, often requiring expert knowledge of both the legacy and target systems.

Traditional data mapping approaches are largely manual, iterative, and error-prone, consuming significant project time and increasing the risk of migration failures. This paper explores the application of AI—particularly AI-powered vector search—to automate and improve this process for life insurance system conversions.

Our approach involves training a custom AI model that captures the nuances of life insurance data structures and product logic. By leveraging vector embeddings and semantic similarity scoring, the model can intelligently map fields between systems, reducing manual effort and enhancing accuracy. This research contributes to the growing field of applied AI in insurance by addressing a long-standing operational challenge with a scalable and adaptive solution.

2. Literature Review

Data migration and system conversion are recognized as critical phases in enterprise transformation initiatives. Traditional approaches rely heavily on manual data mapping, subject-matter expertise, and deterministic rule-based matching. While effective in limited contexts, these methods struggle to scale across complex systems with large and inconsistent data schemas.

Early research by Rahm and Bernstein (2001) provided a taxonomy of schema-matching techniques, while more recent work by Doan et al. (2012) introduced semi-automated integration using machine learning. However, these efforts were largely focused on structured enterprise data with relatively consistent semantics.

In insurance, most AI research has centered around risk assessment, fraud detection, and customer analytics. Data mapping and migration remain less explored. Consulting reports (e. g., McKinsey, 2020; Accenture, 2021) emphasize AI's transformative potential in insurance operations, but few practical implementations exist for automating data conversions—particularly in the life insurance domain.

The introduction of language models such as BERT and domain-adapted embeddings (e. g., GloVe, Word2Vec) has enabled new approaches to semantic matching. These models can generate vector representations of text that capture contextual meaning, making them well-suited for schema mapping tasks where field names differ syntactically but not semantically.

However, off-the-shelf models often lack the contextual understanding required for life insurance terminology. Terms like "cash value maturity" or "death benefit rider" require domain awareness that generic models do not possess. This highlights the need for **custom-trained AI models** that embed insurance-specific semantics into their representations.

Our research addresses this gap by creating a fine-tuned AI model that applies vector similarity search for intelligent field mapping during life insurance data conversions.

3. Methodology

This study developed an AI-powered data mapping framework tailored for life insurance system conversions. The methodology included five core stages: data preparation, model customization, vector embedding, semantic search, and validation.

3.1 Data Collection and Preparation

Metadata was collected from both legacy and target systems, including field names, descriptions, types, and sample values. Sensitive customer data was excluded. Preprocessing involved token normalization and the creation of domain dictionaries reflecting life insurance terminology (e. g., “Policy Issue Date, ” “Benefit Term, ” “Modal Premium”).

3.2 Model Selection and Customization

We used a hybrid approach combining:
- **Word2Vec** for lexical proximity,
- **GloVe** for global co-occurrence semantics,
- **BERT** for deep contextual embeddings.

Each model was fine-tuned on life insurance-specific metadata, documents, and annotated field mappings. Additionally, **prompt engineering** was applied by injecting domain context (e. g., “This field relates to a policy’s issue details”) into input prompts, improving semantic alignment.

3.3 Embedding and Vector Generation

Source and target fields were embedded into vector space using each model. These high-dimensional representations captured semantic relationships across systems, enabling meaningful comparisons beyond surface-level naming.

3.4 Vector Search and Mapping Execution

Cosine similarity was used to compute semantic distance between vectors. For each source field, the top-k closest target fields were ranked. Results from all three models were aggregated for robustness, and mappings with the highest composite scores were selected.

3.5 Validation and Threshold Scoring

To ensure mapping quality, a threshold score (e. g., cosine similarity > 0.85) was applied. Fields falling below the threshold were flagged for SME review. This hybrid validation mechanism allowed high-confidence automation while preserving human oversight where needed.

4. Results and Discussion

The framework was evaluated on a full system migration involving 1, 250 source fields and 980 target fields. Results demonstrate significant improvements in speed and efficiency:

Metric	Manual Baseline	AI-Assisted Approach	Improvement
Avg. Mapping Time/Field	12 min	3.5 min	~70% faster
Total Mapping Effort	~250 person-hours	~75 person-hours	–175 hours
First-Pass Match Accuracy	—	89%	n/a
Rework (Post-SME Review)	—	6%	n/a

Key insights include:
- **Time Savings**: The AI model reduced manual mapping time by ~70%, substantially shortening the overall project timeline.
- **Model Performance**: BERT delivered the highest accuracy on conceptually complex fields, while Word2Vec and GloVe improved recall on abbreviations and legacy naming.
- **Prompt Engineering Impact**: Injecting domain context into prompts increased mapping accuracy by ~6% across the board.
- **Validation Efficacy**: The 0.85 similarity threshold proved effective, requiring minimal manual corrections and ensuring high trust in AI-generated mappings.

These results affirm that AI-driven mapping with domain customization can significantly improve the speed, accuracy, and scalability of data conversions in the insurance industry.

5. Conclusion and Future Work

Data mapping is a longstanding challenge in life insurance system conversions due to the complexity of legacy schemas and regulatory constraints. This research demonstrated that a hybrid AI approach using fine-tuned embeddings and vector search can dramatically reduce mapping effort while maintaining accuracy and contextual fidelity.

The use of multiple models (BERT, GloVe, Word2Vec), combined with prompt engineering and threshold-based validation, yielded strong first-pass accuracy and substantial time savings. Our approach proves that AI, when adapted to industry-specific semantics, can deliver real operational value.

6. Future Work

The next phase of this work involves building a **customizable AI model that can be readily deployed across the insurance industry**. This model will be pretrained on life insurance datasets, integrated with field-level regulatory rules, and designed to work with modern data platforms via APIs. The goal is to provide insurers with a scalable, out-of-the-box solution that accelerates digital transformation and lowers the barrier to modernization.

References

[1] Rahm, E., & Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *The VLDB Journal*, 10 (4), 334–350.
[2] Doan, A., Halevy, A., & Ives, Z. (2012). *Principles of Data Integration*. Elsevier.
[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint* arXiv: 1810.04805.
[4] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint* arXiv: 1301.3781.
[5] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In *EMNLP 2014*, 1532–1543.

[6] McKinsey & Company. (2020). Insurance 2030—The impact of AI on the future of insurance. https: //www.mckinsey. com
[7] Accenture. (2021). Technology Vision for Insurance: Leaders Wanted. https: //www.accenture. com
[8] Vaswani, A., et al. (2017). Attention Is All You Need. *NeurIPS 2017*, 5998–6008.
[9] Zhang, Y., Qiu, X., & Huang, X. (2020). A Survey on Knowledge-Enhanced Pretrained Models. *arXiv preprint* arXiv: 2004.13239.