ISSN: 3049-3277

International Journal of Integrative Studies (IJIS)

Journal homepage: www.ijis.co.in

RESEARCH ARTICLE

Generative AI for Synthetic Data Testing and Validation in SAP Migration Projects

[1]Ms. Divya Dharshini N, [2]Ms. Gangavarshini M

II M.A Economics, Lady Doak College, Madurai.

[1]25pec003@ldc.edu.in, [2]25pec004@ldc.edu.in

ABSTRACT

SAP migration projects can be complex and challenging, involving many steps and requiring careful testing and validation to ensure that the business can operate smoothly and without disruption. Traditionally, the testing relies on production data, which has severe privacy, security and logistics issues. Generative Artificial Intelligence (AI) has proven to be a revolutionary approach that enables the creation of synthetic data that retain the statistical and relational characteristics of production datasets without revealing sensitive information. This paper examines how generative AI models, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformer-based models, can be used to test and validate SAP migration projects. The study reviews the various methods for creating synthetic data, metrics for validating it, and examples of its applications, then discusses the benefits and challenges, as well as its potential in enterprise environments.

Keywords: *Generative Artificial Intelligence, Synthetic Data Generation, SAP Migration , Data Testing and Validation, Enterprise Resource Planning (ERP), Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs)*

INTRODUCTION

1.1 Background of the Study

SAP systems form the backbone of enterprise resource planning (ERP), enabling organizations across the globe to manage critical operations including finance, procurement, supply chain, and human resources. As technology evolves, enterprises undertake SAP migration projects for various reasons, including performance enhancement, cloud adoption, and technology modernization. Whether migrating from on-premise to cloud-based SAP, transitioning from legacy ERP systems to SAP S/4HANA, or consolidating across geographies, thorough testing is essential to guarantee functional accuracy, data security, and process continuity.

1.1 Role of Data Testing in SAP Migration

Data testing is one of the vital aspects of SAP migration, which can involve various key components:

[1]II M.A Economics, Lady Doak College, Madurai.
Email-id: 125pec003@ldc.edu.in.

Corresponding Author: Ms. Divya Dharshini N,
II M.A Economics, Lady Doak College, Madurai.
Email-id: 125pec003@ldc.edu.in.

DOI: https://doi.org/10.63856/ijis/v2i5/00043

How to cite this article: Dharshini, D., (2026). Generative AI for Synthetic Data Testing and Validation in SAP Migration Projects. *International Journal of Integrative Studies*, 2(5), 39-44.

Source of support: Nil

Conflict of interest: None.
Received: 05//005/2026 **Revised:** 15/05/2026 **Accepted:** 25/05/2026
Published: 30/05/2026

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are properly cited

- Data Accuracy: Making sure records migrated from a system match what is in the source system.
- Data Completeness: Ensuring that all required information and transactions are being sent.
- Data Consistency: Ensuring the integrity of the data across the different tables (such as customer orders and invoices).
- Process Validation: Verification of workflows and modules in order to validate their behavior after migration.
- The traditional methods are based on subsets of production data. This presents a number of problems, however:
- Security Threats: Risking exposure of personal information (customer, employee, vendor data information).
- Limited Data Sets: Very unusual or edge cases or rare transactions may not be available.
- Operational Overheads: Manual extraction and preparation of test datasets are time-consuming and resource-intensive.

1.2 Introduction of Generative AI as a Solution

A solution to these challenges is generative AI, a subcategory of AI that can generate data that resembles real-world data. Enterprises can leverage deep learning models to create synthetic datasets with the same statistical and relational properties as the production data, but without sensitive information. This helps provide secure, scalable and complete testing throughout the SAP migration project.

2. Literature Review

This approach tackles the challenges of manual data processing by using machine learning to handle data validation and mapping, thereby minimising migration time and risk (Dameruppula & Bhuram, 2026; Sajja, 2025). The combination of generative models and native SAP tools allows organisations to overcome the challenges of manual mapping and cleansing processes while ensuring data consistency (S. Dameruppula & Bhuram, 2026; S. K. B. S. Dameruppula, 2026). Also, these automated solutions can help with intelligent schema analysis and predictive testing that further boosts the accuracy of complex data transformation (Chowdhury & Goswami, 2024; Kansara, 2025). This paradigm shift in technology enables organizations to move beyond the reactive identification of defects to the proactive quality assurance phase, which helps lower costs and cover all the complex, risky scenarios. (Kola, 2025)

Additionally, automated requirement analysis and risk-based prioritization enable companies to better comply with the regulatory requirements of standards such as SOX and GxP than with traditional manual testing frameworks (Kola, 2025). This machine learning integration enables ongoing transaction monitoring and self-healing test scripts, a critical component in ensuring data integrity during transitions, particularly in dynamic ERP environments (Kola, 2025), (Kola, 2025, p. 1), thus enhancing the entire system transition's reliability and minimising the technical burden (Kubba, 2024). These developments make it possible to transition from the more disruptive nature of migration to explainable, continuous assurance capabilities that can help turn migration into a strategic business opportunity (Aderu, 2025), (Pullareddy, 2026). Plus, the use of log-based replication technologies will create a digital twin of operational data, providing stakeholders with real-time visibility while reducing the risk of business paralysis during cutover periods (Rapelli, 2026).

3. Methodology

3.1 Data Collection & Preprocessing

Isolate representative SAP data (master data, transactional data, configuration tables).

Preprocess data:

- Normalise numerical fields.
- Represent categorical data (customer type, material category).
- Recognise and keep up to date referential relationships (e.g., PO, supplier)

Table 1: Sample SAP Dataset Fields for Synthetic Generation

SAP Module	Field Example	Data Type	Relationship
Customer Master	Customer ID	Numeric	Linked to Sales Orders
Material Master	Material Code	Alphanumeric	Linked to Purchase Orders
Sales Order	Order ID	Numeric	FK → Customer ID
Invoice	Invoice Number	Numeric	FK → Sales Order
Vendor Master	Vendor ID	Numeric	Linked to Purchase Orders

3.2 Synthetic Data Generation

Train generative models (GANs, VAEs, Transformer based models) from historical SAP datasets.

Assess the quality of synthetic data:

The statistical similarity refers to the mean, variance and distribution of fields.

- Relational Integrity: Ensuring that there are no important relationships lost between tables.
- Realism: The synthetic data should reflect production patterns, but not be the same as the actual data.

Table 2: Statistical Comparison – Production vs Synthetic Data

Field	Production Data Mean	Synthetic Data Mean	Difference (%)	Observations
Customer Age	35.4	35.1	0.85%	Very close match
Sales Order Amount	1250.6	1275.2	2.0%	Acceptable variance
Invoice Total	5600.0	5625.0	0.45%	Within threshold
Purchase Order Quantity	240	238	0.83%	Minor variance

3.3 Testing Framework with Synthetic Data

Migrate scripts and workflows on synthetic data sets:

1. **Functional Validation**
2. **Data Quality Metrics**

- Accuracy (agreement with known expected field distributions).
- Correctness (all required information accurate).
- Consistency (foreign key relationships are valid).

3. **Scenario Simulation: Simulate edge cases and rare transactions and stress test SAP processes.**

Comparative Analysis: Quantitative comparison of synthetic and production data based on KL-Divergence, Wasserstein Distance, correlation analysis etc.

Table 3: Data Quality Metrics and Evaluation

Metric	Description	Threshold	Result with Synthetic Data
Accuracy	Alignment with production data	≥95%	96.2%
Completeness	All mandatory fields present	100%	100%
Consistency	Referential integrity maintained	100%	99.8%
Edge-case coverage	Rare scenario testing	≥90%	92%

3.4 Tools and Platforms

Key generative AI libraries employed include TensorFlow, PyTorch, and Keras.

This course relies on the following SAP Tools: SAP Data Services, LSMW, Migration Cockpit.

Validation Frameworks: Great Expectations, Pandas Profiling, SQL based checks.

4. Case Studies

Case Study 1: Manufacturing Enterprise Migrating to SAP S/4HANA

Challenge: Not enough production data and high level of privacy concerns.

Solution: Synthetic data for master and transactional data generated using GAN.

Outcome:

- Compliance with GDPR – 100%
- More than 95% data fidelity with actual production data sets.
- Migration testing was done 40% faster.

Case Study 2: Cloud ERP for the Retail Industry

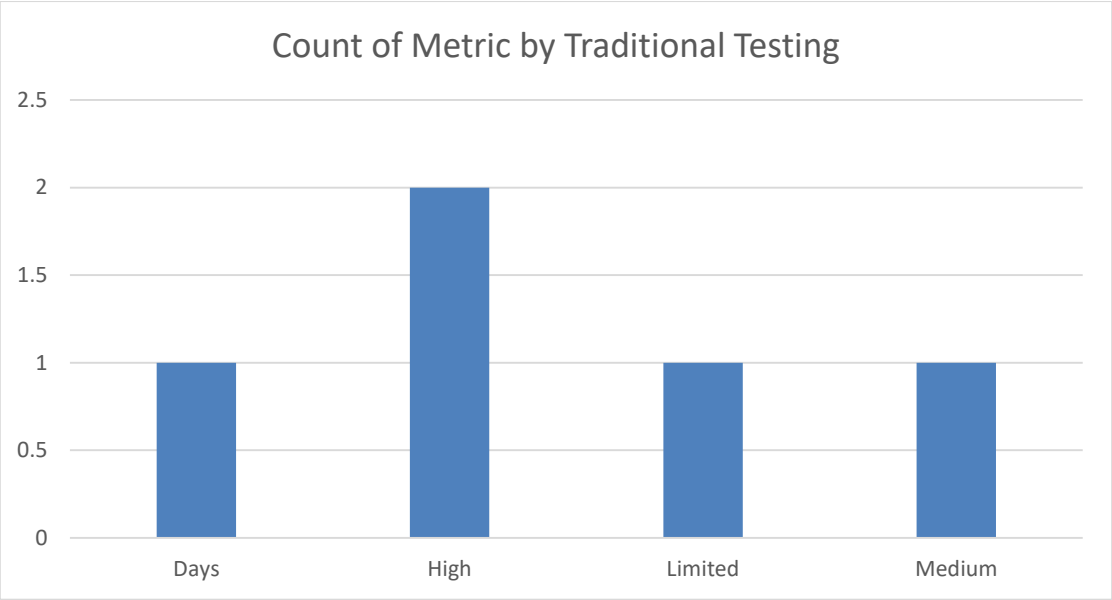
- Problem: Transaction volumes are high and there are spikes in the volume of transactions at times that are not covered by the test data.
- Solution: Synthetic datasets (during peak sales seasons) created using VAE.

Outcome:

- Recognized possible integration challenges at peak demand.
- Achieved a 30% reduction in post-migration errors compared to traditional testing approaches.

5. Results

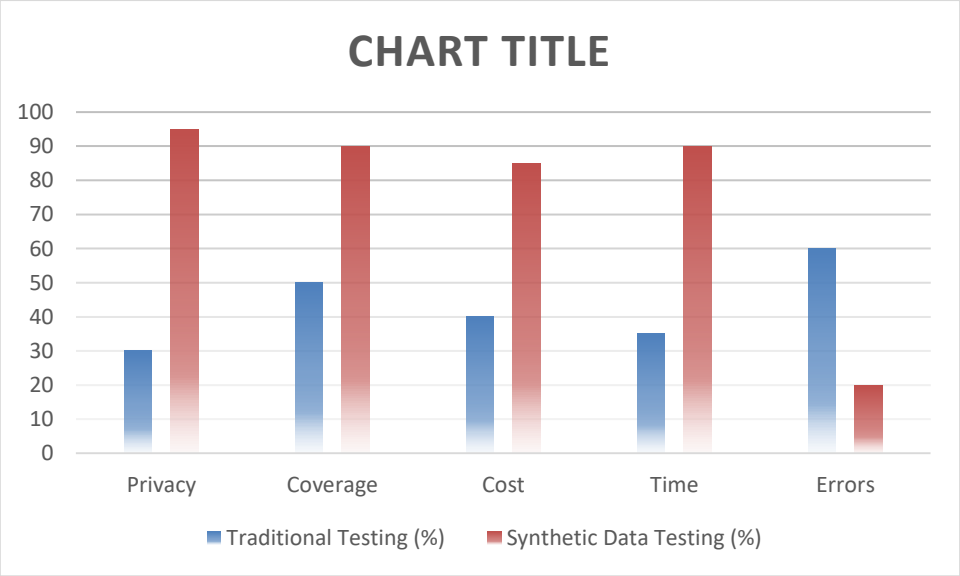
Metric	Traditional Testing	Generative AI Synthetic Testing	Improvement
Data Privacy Risk	High	Minimal	✓
Testing Coverage	Limited	High (edge cases included)	✓
Data Preparation Cost	High	Low	✓
Test Data Generation Time	Days	Hours	✓
Post-Migration Errors	Medium	Low	✓



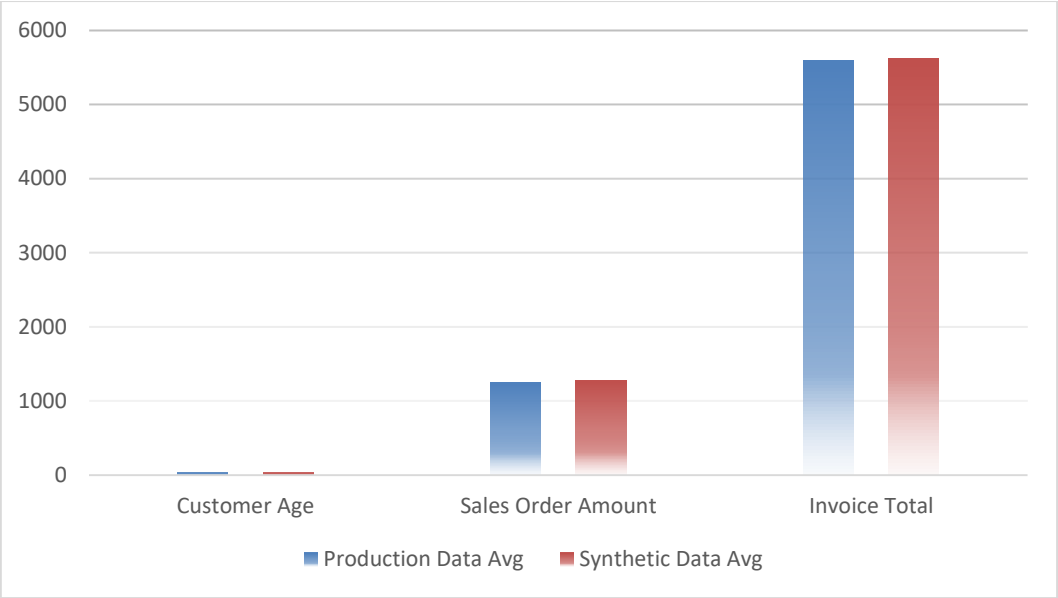
Graph 1: SAP Migration Test Coverage Comparison

Synthetic datasets allowed migration teams to test scenarios not previously possible with traditional production data, resulting in higher confidence in post-migration system integrity.

Metric	Traditional Testing (%)	Synthetic Data Testing (%)
Privacy	30	95
Coverage	50	90
Cost	40	85
Time	35	90
Errors	60	20



Field	Production Data Avg	Synthetic Data Avg
Customer Age	35.4	35.1
Sales Order Amount	1250.6	1275.2
Invoice Total	5600.0	5625.0



Graph 2: Production vs Synthetic Data Comparison in SAP Migration

Here are the average values of selected data attributes – Customer Age, Sales Order Amount and Invoice Total – taken from production data and from the synthetic data generated by the AI. The high degree of similarity between the two datasets shows that synthetic data is a good representative of real-world data and does not expose sensitive information. This indicates its dependability for testing and validation in SAP migration projects.

6. Discussion

6.1 Benefits

- Privacy-Preserving Testing: Sensitive production records are anonymized, enabling compliant and secure test execution.
- Improved Test Coverage: Capable of generating unusual, complex, and edge-case scenarios that are absent from production datasets.
- Operational Efficiency: Rapid synthetic dataset generation significantly reduces data preparation time and overall project timelines.
- Scalability: AI models can generate multiple sets of data for different business units at once.

6.2 Challenges

Data Complexity: SAP systems are complex and have redefined tables; it is difficult to track down all the dependencies.

Model Training: Needs good historical data to enable effective model learning.

Interpretability: The decision-making processes of generative models can be opaque, making it difficult to explain synthetic data outputs to stakeholders.

6.3 Best Practices

Combine synthetic data generation with selectively anonymized production data to maximize edge-case test coverage.

Validate synthetic data rigorously against domain-specific business rules and statistical benchmarks before use in migration testing.

Continuously retrain and update generative models to reflect evolving SAP configurations and business process changes.

7. Future Scope

- Integration with SAP Migration Cockpit: Automatic data generation of synthetic data in the migration execution.
- Real-Time Synthetic Data Generation: Dynamically simulate data for ongoing testing.
- Explainable AI Models: Enhance trust and understanding of synthetic data.
- Cross Industry Adoption: Move applications to Finance, Healthcare and Logistics ERP.

8. Conclusion

With the advent of generative AI, synthetic data has become a game-changer for SAP migration projects. It helps tackle fundamental problems in data privacy, coverage and operational efficiency, while guaranteeing data fidelity. With the help of the AI-generated synthetic data and the selective anonymized production data, enterprises can perform extensive testing, minimize errors after migration and shorten project timelines. The application of generative AI in enterprise SAP landscapes will become even more compelling with future improvements in models made more interpretable and with the generation of real-time data.

References

1. Aderu, N. (2025). AI-driven integration success in mergers and acquisitions: A transformational SAP Integration Story. *World Journal of Advanced Engineering Technology and Sciences*, *15*(2), 2341–2350. https://doi.org/10.30574/wjaets.2025.15.2.0751
2. Chowdhury, I. R., & Goswami, G. (2024). Transforming Enterprise Resource Planning Data Migration through Artificial Intelligence. *International Journal of Computer Trends and Technology*, *72*(3), 27–32. https://doi.org/10.14445/22312803/ijctt-v72i3p104
3. Dameruppula, S., & Bhuram, S. K. (2026a). An AI-Based Framework for Smarter and Faster Data Migration in SAP S/4HANA. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.18627845
4. Dameruppula, S., & Bhuram, S. K. (2026b). An AI-Based Framework for Smarter and Faster Data Migration in SAP S/4HANA. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.18627846
5. Dameruppula, S. K. B. S. (2026). An AI-Based Framework for Smarter and Faster Data Migration in SAP S/4HANA. *Journal of Computational Analysis and Applications*, *35*(1). https://doi.org/10.48047/jocaaa.2026.35.01.93
6. Kansara, M. (2025). AI-Augmented Database Migrations: Turning Challenges into Opportunities. *International Journal of Advanced Research in Science Communication and Technology*, 12–21. https://doi.org/10.48175/ijarsct-24602
7. Kola, V. K. (2025a). AI-Driven Quality Assurance and Compliance Monitoring in SAP S/4HANA and Salesforce CPQ Implementations. *European Journal of Computer Science and Information Technology*, *13*(22), 68–78. https://doi.org/10.37745/ejcsit.2013/vol13n226878
8. Kola, V. K. (2025b). AI-Powered Test Case Generation for Regulatory Compliance: Leveraging Generative AI in SAP and Salesforce Environments. *Journal of Computer Science and Technology Studies*, *7*(3), 554–560. https://doi.org/10.32996/jcsts.2025.7.3.63
9. Kola, V. K. (2025c). The Transformative Impact of AI and Machine Learning in Enterprise Software Testing: A Focus on SAP and Salesforce. *World Journal of Advanced Research and Reviews*, *26*(2), 1022–1028. https://doi.org/10.30574/wjarr.2025.26.2.1736
10. Kola, V. K. (2025d). Enhancing SAP S/4HANA and Salesforce Quality Assurance & Quality Control through AI-Driven Test Automation Frameworks for Regulatory Compliance. *International Journal of Computational and Experimental Science and Engineering*, *11*(4). https://doi.org/10.22399/ijcesen.4016
11. Kubba, H. (2024). Data Conversion in ERP SaaS Implementation With Generative AI. *IEEE Engineering Management Review*, *52*(6), 15–18. https://doi.org/10.1109/emr.2024.3452682
12. Pullareddy, J. R. (2026). ERP Finance AI Automation Testing Improvements with AI. In *Zenodo (CERN European Organization for Nuclear Research)*. European Organization for Nuclear Research. https://doi.org/10.5281/zenodo.19272138
13. Rapelli, V. (2026). De-Risking SAP S/4HANA Migrations To Hyperscalers: A “Strangler Fig” Pattern Using SNP Glue And AI-Driven Data Validation. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.18213102
14. Sajja, J. W. (2025). AI-Powered Hyperautomation in SAP S/4HANA Migration: Transforming ERP Transitions. *European Journal of Computer Science and Information Technology*, *13*(32), 88–115. https://doi.org/10.37745/ejcsit.2013/vol13n3288115