

# Harnessing AI to Safeguard Data Integrity in Legacy System Upgrades

Vijayasekhar Duvvur

Email: vijay.duvvur[at]gmail.com

**Abstract:** Migrating data from legacy systems to modern platforms presents significant challenges, particularly in maintaining data integrity. This article explores the application of Artificial Intelligence (AI) in ensuring data integrity during the complex process of legacy system migration. AI technologies, including machine learning and pattern recognition, can automate data validation, error detection, and correction processes, thereby reducing human error and enhancing the reliability of the migration process. We will delve into various AI-driven approaches and technologies that have proven effective in safeguarding data integrity during transitions.

**Keywords:** Legacy System Migration, Data Integrity, Artificial Intelligence, AI, Machine Learning, ML, Data Validation, Error Detection, Data Correction, Automated Migration.

## 1. Introduction

Legacy system migration is a crucial process for organizations looking to leverage advanced technological capabilities and improve their operational efficiency. However, maintaining the integrity of data during such migrations is a pivotal concern, as data is often an organization's most valuable asset. Data corruption, loss, or unauthorized alterations during migration can lead to significant operational disruptions and compliance issues. Artificial Intelligence offers sophisticated solutions to these challenges, employing advanced algorithms to manage and protect data throughout the migration process.

### Enhancing Data Integrity with AI

Data integrity is a foundational aspect of any system migration. When migrating from legacy systems to modern infrastructures, maintaining the accuracy, consistency, and reliability of the data is paramount. Artificial Intelligence (AI) and Machine Learning (ML) offer powerful tools to enhance data integrity by automating complex processes, improving error detection and correction, and ensuring seamless data transitions. Below, we delve deeper into how AI can be utilized to bolster data integrity during migrations [1-3,5].

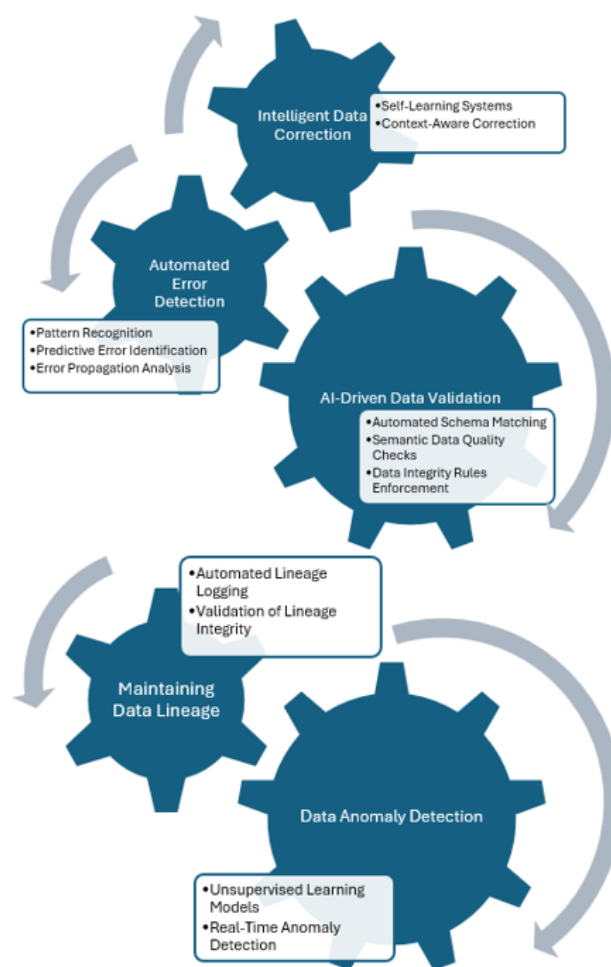


Figure 1: Steps to enhance data integrity with AI

### 1) AI-Driven Data Validation

Data validation is critical to ensure that data migrated from legacy systems maintains its accuracy and consistency in the new environment. AI can significantly enhance this process through:

- *Automated Schema Matching:* AI algorithms can automatically map data fields between the legacy and the new systems. This involves recognizing relationships between different data models and

aligning them accurately, ensuring that data fits into the new schema without loss or corruption.

- *Semantic Data Quality Checks:* Beyond structural validation, AI can perform semantic checks on data. For instance, ML models can learn from historical data patterns to identify entries that do not make sense contextually, such as improbable dates or mismatched geographical information.
- *Data Integrity Rules Enforcement:* AI-driven tools can enforce business and data integrity rules automatically during the migration process. Rule-based machine learning systems apply sets of predefined rules and also adapt these rules based on data patterns observed during the migration. For instance, if a certain type of data discrepancy recurs, the AI system can learn to apply specific checks preemptively in future operations.

## 2) Automated Error Detection

AI enhances the detection of errors that might occur during the data migration process. This capability is crucial for preempting data corruption before it affects business operations [6].

- *Pattern Recognition:* ML algorithms can detect deviations from known data patterns, flagging inconsistencies that suggest errors or corruption. For example, if transaction volumes suddenly drop or spike without a corresponding business reason, AI can alert administrators to investigate potential data loss or duplication[4].
- *Predictive Error Identification:* By analyzing historical migration projects and outcomes, AI models can predict where errors are likely to occur in the migration pipeline. This predictive insight allows teams to focus their efforts on high-risk areas, ensuring more robust data handling and validation strategies.
- *Error Propagation Analysis:* In complex data environments, errors in one part of the system can propagate to others. AI can trace and analyze the potential spread of these errors. Using graph theory and network analysis, AI models map out the relationships and dependencies within data sets to predict how errors might propagate. Understanding potential propagation paths helps in isolating errors and implementing targeted fixes without disrupting the entire system.

## 3) Intelligent Data Correction

Once errors are detected, correcting them is the next step. AI can automate and enhance the accuracy of this process[7].

- *Self-Learning Systems:* AI systems can learn from each migration, improving their ability to identify and correct similar errors in future projects. Over time, this reduces the overall error rate and improves migration efficiency.
- *Context-Aware Corrections:* AI can understand the context around data errors (such as related data points or previous corrections) and use this information to make more informed correction decisions, reducing the need for human intervention and the potential for additional errors.

## 4) Data Anomaly Detection

Detecting anomalies is crucial for maintaining data integrity, as they can indicate underlying issues such as data breaches or system malfunctions [4,8].

- *Unsupervised Learning Models:* These models can identify data anomalies by learning the normal distribution of data in the legacy system and flagging outliers that do not fit the pattern during and after migration.
- *Real-Time Anomaly Detection:* AI systems can monitor data continuously as it is migrated, immediately detecting and addressing anomalies. This real-time capability ensures that issues are resolved swiftly, minimizing their impact.

## 5) Maintaining Data Lineage

Maintaining a clear trace of where data originated and how it has been altered is essential for auditing and compliance, particularly in regulated industries [9].

- *Automated Lineage Logging:* AI tools can automatically track and log all data transformations and migrations, creating a detailed and accurate record that can be invaluable for compliance reviews and troubleshooting.
- *Validation of Lineage Integrity:* AI can continuously verify the integrity of data lineage records, ensuring that every piece of data can be accurately traced through its lifecycle across systems[8].

## 2. Conclusion

Artificial Intelligence significantly enhances the integrity of data during legacy system migrations. By automating complex processes such as data validation, error detection, and correction, AI reduces the risk of data corruption and loss. Moreover, AI's capacity to learn and adapt from each migration experience improves the overall efficiency and reliability of future migrations. Organizations embracing AI in their legacy system migrations can ensure a smoother transition to new technological platforms with minimal disruption to their operations.

## References

- [1] Vinjamuri, L. P., Aghanaashaa, A., Chahar, A., & Alisha, A. (2023). Artificial intelligence as a tool for enhanced data integrity and data security. 2023 International Conference on Artificial Intelligence and Smart Communication (AISC).
- [2] Robinson, C., & Dulcina, B. (2018, June). A machine learning approach to modeling human migration. Retrieved from <https://dl.acm.org/doi/abs/10.1145/3209811.3209868>
- [3] Hussein, A. E. A. (2021). Data migration need, strategy, challenges, methodology, categories, risks, uses with cloud computing, and improvements in its using with cloud using suggested proposed model. Journal of Information Security, 12(01). Retrieved from [https://www.scirp.org/html/4-7800724\\_106592.htm](https://www.scirp.org/html/4-7800724_106592.htm)
- [4] Cooper, S., Bhuiyan, M., & Arslan, E. (2020). Machine learning for data transfer anomaly detection. Department of Computer Science & Engineering,

- University of Nevada Reno. Retrieved from <https://par.nsf.gov/servlets/purl/10212851>
- [5] Boina, R., Achanta, A., Mandvikar, S. (2023). Integrating data engineering with intelligent process automation for business efficiency. (2022). International Journal of Science and Research (IJSR).
- [6] Paramsivapandi, S. K., & Nagarajan, S. (2021). Machine learning based hybrid recurrent data-driven flow algorithm to identify the risk of data flow error detection in cloud computing. International Journal for Research in Applied Science & Engineering Technology (IJRASET).
- [7] Warudkar, H. (2019). AI for data cleaning: How AI can clean your data and save your man hours and money. Retrieved from <https://www.expressanalytics.com/blog/ai-data-cleaning/>
- [8] Saiod, A. K., & van Greunen, D. (2021). The impact of deep learning on the semantic machine learning representation. Retrieved from <https://www.igi-global.com/chapter/the-impact-of-deep-learning-on-the-semantic-machine-learning-representation/271119>
- [9] Bertino, E., Bhattacharya, S., Ferrari, E., & Milojicic, D. (2023). Trustworthy AI and data lineage. Retrieved from <https://ieeexplore.ieee.org/abstract/document/10321961>