

# Research Survey on AI-Enhanced Data Accuracy Systems

## 1. Introduction

### 1.1 Background

Data accuracy is critical in modern analytics and decision-making processes. As data volume and complexity increase, traditional rule-based data validation techniques are proving inadequate. AI offers robust methodologies for enhancing data accuracy through automation, learning patterns, and handling uncertainty.

### 1.2 Objective

This survey explores how AI techniques are applied to improve data accuracy in various domains. It covers methodologies, applications, evaluation metrics, and future directions.

## 2. Importance of Data Accuracy

Data accuracy, completeness, consistency, timeliness, and uniqueness are vital dimensions. Inaccurate data leads to poor decisions, compliance issues, and financial loss. Automated approaches are needed due to the inefficiencies of manual methods.

## 3. AI Techniques for Enhancing Data Accuracy

### 3.1 Machine Learning (ML)

- Supervised Learning: Detect errors with labeled data.
- Unsupervised Learning: Identify anomalies or clusters.
- Semi-supervised Learning: Leverage limited labeled data.

### 3.2 Deep Learning

- Autoencoders for error detection.
- RNNs for time-series validation.
- Transformer models for unstructured data analysis.

### 3.3 NLP

- Entity resolution and normalization in text fields.

### 3.4 Reinforcement Learning

- Dynamic optimization of data quality decisions.

## 4. AI Applications in Data Accuracy

- Healthcare: Patient record validation using NLP.
- Finance: Transaction validation and fraud detection.
- E-commerce: Product catalog deduplication.
- CRM: Contact/address standardization.
- IoT: Real-time anomaly detection.

## 5. Evaluation Metrics

- Precision, Recall, F1 Score.
- Accuracy Improvement Rate.
- Data Validity Ratio.

# Research Survey on AI-Enhanced Data Accuracy Systems

- False Positive/Negative Rates.

## 6. Key Tools and Platforms

- OpenRefine, Great Expectations, TensorFlow, PyTorch.
- HoloClean, Apache Deequ.

## 7. Challenges and Limitations

- High data labeling cost.
- Scalability in real-time systems.
- Model explainability.
- Bias and fairness issues.
- Privacy concerns.

## 8. Future Research Directions

- Explainable AI for transparency.
- Self-healing pipelines.
- Federated learning for distributed data.
- Blockchain for audit trails.
- Cross-domain transfer learning.

## 9. Conclusion

AI systems are evolving the way data accuracy is managed--from manual corrections to real-time, proactive, and explainable data quality solutions.

## 10. References

- HoloClean: <https://holoclean.io>
- Great Expectations: <https://greatexpectations.io>
- IEEE and ACM publications on AI and data quality.