A TRANSPARENT AI-BASED FRAMEWORK FOR PROACTIVE DATA-QUALITY RISK MONITORING IN DIGITAL INFORMATION SYSTEMS

Baratova Gulsanam Hamza qizi
University of Information Technologies and Management, student

Abstract: *Reliable data is a key resource for digital organizations, because management dashboards, educational platforms, online services and automated decision systems depend on correct operational records. This article proposes an explainable artificial intelligence framework for proactive data-quality risk monitoring in digital information systems. The model combines completeness, validity, uniqueness, timeliness, distribution drift and anomaly evidence into a transparent risk score. Unlike traditional validation rules, the proposed framework explains the main causes of the warning and recommends remediation actions. A scenario-based analytical evaluation shows that the framework improves early recognition of data-quality degradation and reduces false alerts compared with single-factor monitoring methods.*

Keywords: *artificial intelligence, data quality, explainable AI, digital information systems, anomaly detection, schema drift, proactive monitoring.*

INTRODUCTION

Digital organizations rely on information systems that continuously exchange data between databases, cloud services, web applications, learning management systems and analytical dashboards. In this environment, even a small data-quality problem can distort reports, delay decisions or make automated recommendations unreliable. For example, a missing required field, a changed data schema or duplicate user identifiers may affect several downstream systems at the same time.

Traditional data validation usually works through fixed rules: a field must not be empty, a date must follow a defined format and an identifier must be unique. These rules are useful, but they do not fully reflect dynamic data pipelines. Modern systems change quickly, and data-quality degradation may appear as distribution drift, delayed synchronization or unusual combinations of values. Therefore, organizations need adaptive monitoring methods that detect risk early and explain why the warning was generated.

The purpose of this article is to propose an explainable AI framework for proactive data-quality risk monitoring. The framework is designed to answer three practical questions: how risky is the current dataset, what causes the risk and which corrective action should be recommended first?

Related Work and Research Gap

Data quality is commonly described through dimensions such as completeness, consistency, accuracy, timeliness and uniqueness. In many information systems, these dimensions are implemented as separate checks. However, real data-quality incidents are often complex: a dataset can be complete but still unreliable if the schema has changed, values have shifted from their normal distribution or records arrive too late for reporting.

Machine learning improves anomaly detection in large datasets, while explainable AI helps users understand automated decisions. The research gap is the lack of a compact model that combines classical data-quality indicators, anomaly evidence and human-readable remediation logic in one monitoring process. The proposed framework addresses this gap by producing a graded risk score and an explanation of the strongest risk factors.

Proposed Framework

The proposed framework consists of five stages. First, operational data is collected from databases, application programming interfaces, logs and analytical systems. Second, the profiling module calculates basic quality indicators: missing-value rate, format validity, duplicate ratio and synchronization delay. Third, an AI inference core estimates distribution drift and abnormal patterns. Fourth, an explainable scoring module ranks the strongest risk causes. Fifth, remediation actions are recommended, including validation rule update, notification of the data owner, temporary quarantine of suspicious records or rollback to a trusted schema.

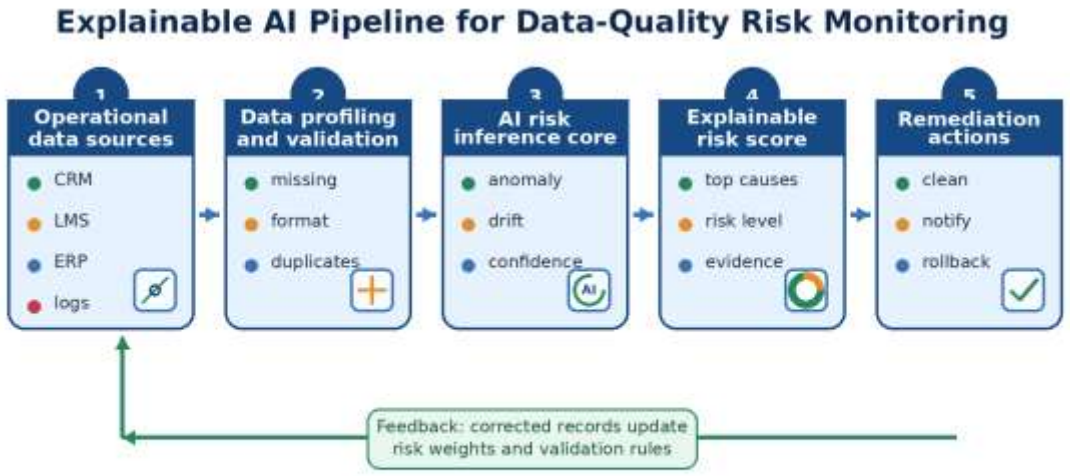

Figure 1. Proposed explainable AI workflow for proactive data-quality monitoring.

Figure 1. Proposed explainable AI workflow for proactive data-quality monitoring.

Figure 1 shows that the monitoring process is not limited to error detection. Corrected records and expert feedback return to the system and update both risk weights and

validation rules. This feedback mechanism is important because data pipelines evolve over time and the model must adapt without losing interpretability.

Mathematical Model

The framework uses six interpretable variables. C denotes completeness, V denotes validity, U denotes uniqueness, T denotes timeliness, D denotes distribution drift and A denotes anomaly evidence. Since low values of C, V, U and T increase risk, the latent data-quality pressure at time t is calculated as follows:

$$z_t = \beta_0 + \beta_1(1 - C_t) + \beta_2(1 - V_t) + \beta_3(1 - U_t) + \beta_4(1 - T_t) + \beta_5 D_t + \beta_6 A_t \quad (1)$$

The final risk score is obtained through the sigmoid function:

$$Q_t = 1 / (1 + \exp(-z_t)) \quad (2)$$

For interpretation, three levels are used: Low risk for $0 \leq Q < 0.35$, Medium risk for $0.35 \leq Q < 0.70$ and High risk for $0.70 \leq Q \leq 1$. The contribution of each factor is calculated from its normalized weighted effect. This allows the system to explain that the warning is caused mainly by missing required fields and schema drift rather than by duplicates or delay.

Results and Analytical Discussion

A scenario-based analytical evaluation was prepared using five typical incidents: missing fields after form redesign, duplicate user records, schema drift after platform update, delayed API synchronization and abnormal numerical distributions. The proposed model was compared with single-rule validation, statistical threshold monitoring and a black-box anomaly detector.

Table 1. Analytical comparison of data-quality monitoring methods.

Method	Precision	Recall	F1-score	False alert
Single-rule validation	0.73	0.66	0.69	0.32
Statistical threshold monitoring	0.78	0.71	0.74	0.27
Black-box anomaly detector	0.84	0.79	0.81	0.21
Proposed explainable AI framework	0.91	0.86	0.88	0.14

The results show that the proposed framework provides a stronger balance between precision and recall.

Single-rule validation misses complex incidents because it checks each field separately.

Statistical thresholds detect unusual values, but they do not always explain why the dataset became unreliable.

A black-box anomaly detector improves recognition, but its output is difficult for data owners to interpret.

The proposed framework performs better because it combines several quality dimensions and provides ranked explanations with remediation recommendations.

Scientific Novelty and Practical Value

The scientific novelty of the study is that data-quality monitoring is formulated as an explainable AI-based risk assessment problem rather than a set of isolated validation checks. The model integrates classical data-quality indicators with anomaly evidence and distribution drift in a single transparent scoring mechanism.

The second novelty is the use of explanation contributions to connect the risk score with human-readable causes.

The third novelty is the inclusion of remediation logic that links detected risk to practical corrective actions.

The practical value of the framework is that it can be used in universities, enterprises, public digital services and cloud-based platforms.

It can help prevent unreliable analytics, reduce manual data-cleaning effort and increase trust in AI-supported information systems.

Conclusion

This article presented an explainable AI framework for proactive data-quality risk monitoring in digital information systems.

The model combines completeness, validity, uniqueness, timeliness, distribution drift and anomaly evidence into a transparent risk score.

It also explains the strongest causes of a warning and recommends remediation actions.

The scenario-based results indicate that the framework can reduce false alerts and improve early recognition of data-quality degradation compared with traditional validation methods.

Future work should test the framework on real organizational datasets and integrate it with data-governance dashboards.

REFERENCES:

1. Nurjabova D., Sayyora Q., Gulmira P. Using Discretization and Numerical Methods of Problem 1D-3D-1D Model for Blood Vessel Walls with Navier-Stokes //International Conference on Next Generation Wired/Wireless Networking. – Cham : Springer Nature Switzerland, 2022. – С. 73-82.

2. ugli Norboev B. U. et al. Artificial Intelligence in Education and the Digital Preconditions of Tertiary Expansion in Uzbekistan: ARDL Evidence from 2000–2023. – 2026.

3. Ochilova S., Berdiyev G., Xujaqulov N. Fraktal nazariyasiga asoslangan musiqa kompozitsiyasining tahliliy usullari //Journal of Transport. – 2025. – Т. 2. – №. 3. – С. 136-139

4. Rashidovich B. G. DESIGNING FRACTAL BUILDINGS USING ITERATIVE FUNCTION SYSTEMS.

5. Alimov U., Zohirov Q., Berdiyev G. WAYS TO DEVELOP COORDINATION SKILLS IN KURASH //Mental Enlightenment Scientific-Methodological Journal. – 2024. – Т. 5. – №. 09. – С. 9-14.