AI-Assisted Data Quality Assessment for Big Data Pipelines: Framework, Techniques, and Empirical Evaluation

Mohammed Imran Ahmed[1], Syed Saifuddin Ahmed Muzaffar[2]

Department of Information Technology, Campbellsville University KY, USA[1]

Department of Information Technology, Campbellsville University KY, USA[2]

https://orcid.org/0009-0004-1154-2638[2]

Abstract: Data is continuously coming in from many sources at a fast rate in today's big data ecosystems. Because of this, making ensuring the data is of a suitable quality is both crucial and challenging. Inaccurate analytics outputs and worse performance, fairness, and credibility of later AI models can result from poor data quality, which can show up as missing values, inconsistencies, anomalies, duplication, and delayed updates. The conventional methods of evaluating data quality—static rules, manual profiling, and preset limitations—do not work well with big data pipelines that are constantly changing.

An approach for AI-assisted data quality evaluation that can be used to both massive batch and streaming pipelines is covered in this paper. In order to automatically profile data, detect quality problems, and produce adaptive quality ratings for a range of attributes like completeness, consistency, correctness, timeliness, and validity, the suggested method makes use of machine learning and deep learning. Unsupervised models for anomaly detection include autoencoders and isolation forests. In contrast, supervised learning techniques estimate quality labels and scores based on expert feedback and historical data. The architecture provides a way for the system to keep learning, allowing it to adjust when pipeline configurations and data distributions change.

The AI-assisted approach beats conventional rule-based approaches at identifying intricate and previously unidentified data quality problems, according to an experimental evaluation using real-world big data workloads. The findings also demonstrate that downstream analytics and machine learning models perform better when trained on data validated by the suggested approach. All things considered, our study shows that AI-driven data quality assessment may be a scalable, flexible, and astute way to guarantee that future large data pipelines contain accurate data.

Keywords: AI-driven data quality and big data pipelines assessing data quality, applying machine learning, identifying anomalies, profiling, streaming, and data governance. clever engineering of data.

I. INTRODUCTION

The quick development of big data technologies has changed how companies collect, handle, and evaluate data to make decisions, automate processes, and offer intelligent services. Information is continuously gathered by contemporary big data pipelines from a range of sources, such as sensors, transactional systems, social media, and business apps [1]. Large-scale analytics and real-time insights are offered by these pipelines, but maintaining high data quality is also very challenging. Inaccurate, missing, inconsistent, or outdated data can be disseminated by analytics workflows, leading to inaccurate insights and subpar machine learning and artificial intelligence (AI) model performance [2].

Rule-based validation, schema enforcement, and human data profiling were traditional approaches to data quality [3]. Such systems rely on human talent and standards to identify problems such as missing digits or formatting errors. However, these technologies are more prone to malfunction and more challenging to manage in circumstances involving vast amounts of constantly changing big data. Static rules are unable to keep up with quality issues that are always changing and become more complicated due to the rapid and frequent arrival of data streams, schemas, and concepts. Many data errors are therefore left unreported until they endanger analytics or the company.

It's crucial to have a reliable method to guarantee the quality of their data as more businesses employ AI-driven analytics. Predictions may be inaccurate, unstable, or skewed because machine learning models are extremely vulnerable to data problems. For AI systems to be honest and reliable, keeping high-quality data is therefore not just a

great idea but also essential [4]. New automated techniques are being used more and more to monitor, evaluate, and improve the quality of data in large data pipelines.

In this case, data quality checks can now be aided by AI. Without much help from humans, AI-powered systems are able to find patterns in data, spot outliers, and adapt to shifting data distributions. Deep learning and machine learning are used to accomplish this [5]. Along with analyzing AI's potential to enhance data quality evaluation in massive data pipelines, this work offers a scalable method for integrating AI-driven quality assessment into batch and streaming data processing. The research aims to demonstrate how smart data quality technology may improve pipeline dependability and downstream analytics performance.

II. BACKGROUND & LITERATURE REVIEW

A. Quality of Data in Big Data Circumstances

How effectively data fits its intended use is referred to as its quality. Among the most common metrics are completeness, accuracy, consistency, timeliness, validity, and uniqueness. Because there are so many data sources that come in so fast and in so many various forms, it is getting harder to keep these dimensions current in big data situations. Fast data updates, parallel processing, distributed storage, and schema-on-read paradigms can all lead to unnoticed quality problems that are hard to find with standard techniques [6]. It has frequently been demonstrated that low-quality data directly affects the reliability of research and the capacity to make operational judgments, especially when artificial intelligence is involved.

B. Typical Techniques for Verifying Data Quality

Statistical summaries, integrity constraints, rule-based validation, and human data profiling are common techniques for assessing the quality of data [7]. ETL pipelines frequently use business rules, threshold checks, and schema validation. Although these techniques are neither scalable or adaptable, they perform well on stable, well-defined datasets. Static rules commonly miss intricate, hard-to-see mistakes like contextual inconsistencies or small distribution changes, and thus need to be updated frequently. Furthermore, physical involvement doesn't work in streaming pipelines that move quickly.

C. Enhancing Data Quality with AI and Machine Learning Applications

Recent studies have looked into the ways that AI and machine learning can assist in resolving issues with conventional approaches [8]. In the absence of labelled data, anomalies and outliers can be found using unsupervised learning approaches such as clustering, isolation forests, and autoencoders. Supervised algorithms have been suggested as a way to forecast data quality labels or ratings using prior quality annotations. The quality of unstructured data, including text, photos, and logs, can also be assessed using deep learning algorithms. These AI-powered systems' capacity to expand, adapt, and identify intricate patterns makes them perfect for dynamic big data pipelines.

D. Current Research Deficits and Frameworks

AI-based data quality frameworks with an emphasis on automated profiling, quality grading, and ongoing monitoring have been proposed in a number of ways. However, a lot of current solutions are ineffective for end-to-end, real-time pipelines since they only work offline or examine one quality metric at a time. Research on unified frameworks that can easily integrate AI-driven quality evaluation into batch and streaming big data operations while permitting ongoing learning and governance is also lacking [9].

Table 1: Comparison of Data Quality Assessment Approaches

Approach Type	Techniques Used	Scalability	Adaptability	Key Limitations
Rule-based	Constraints, thresholds, schemas	Low–Medium	Low	Manual maintenance, brittle rules
Statistical	Profiling, distributions	Medium	Low	Limited semantic understanding
ML-based	Anomaly detection, classification	High	High	Requires training data, explainability
Deep Learning	Autoencoders, NLP models	Very High	Very High	Computational cost, transparency

III. PROPOSED AI-ASSISTED DATA QUALITY ASSESSMENT FRAMEWORK

An entire AI-assisted data quality evaluation architecture for massive big data pipelines that run-in batch and streaming modes is shown in this section [10]. The system incorporates sophisticated quality assessment at every stage of the data lifecycle. This enables you to implement large-scale initiatives, learn as you go, and keep an eye on things constantly.

A. A Synopsis of the Architecture
The proposed framework has a layered, modular architecture that works well with current big data platforms like Flink, Kafka, and Apache Spark [11]. Through ingestion levels, data from many sources enters the pipeline and is processed concurrently on several nodes. AI-powered quality layers assess data quality almost instantly and give you useful feedback signals and quality metrics.

B. Key Components
Layer of Data Ingestion Batch and streaming data from structured, semi-structured, and unstructured sources are received by this layer [12]. Source, date, and schema version metadata are preserved so that the quality can be assessed contextually.

For use as input by AI models, the Feature Extraction & Profiling Layer automatically produces statistical, semantic, and temporal features (such as null ratios, value distributions, and pattern frequencies) [13].

C. AI-based quality assessment engine
- Deep learning methods can handle complicated data types like sensor feeds and text logs;
- supervised models infer data quality labels or ratings using prior quality annotations;
- while unsupervised models, including Isolation Forests and Autoencoders, identify outliers and distributional changes [14].

1) Module for quality scoring and alerting
In order to generate comprehensible quality scores based on factors including timeliness, accuracy, consistency, and completeness, this module integrates model outputs [15]. Alerts are set off when quality standards are surpassed.

2) Continuous Learning & Feedback Loop
In reaction to user input, rule changes, and downstream model performance metrics, AI models are constantly retrained and modified. As a result, they are more resilient to changing data patterns and concept drift.

D. *Key Advantages*
Because the suggested framework is automated, scalable, and adaptive, it overcomes the rigidity of rule-based methods. By integrating AI into the pipeline, problems with data quality are detected before they arise and downstream analytics and AI systems continue to be dependable [16].

Figure 1: AI Assisted Data Quality Assessment Framework

IV. METHODOLOGY

The implementation and evaluation of the suggested AI-assisted data quality assessment system in big data pipelines are covered in this part [17]. The approach is intended to be less costly, more adaptable, and more reproducible than previous approaches.

A. Selection of Data Sources and Data Preparation

We use both synthetic and real-world datasets to evaluate the framework's performance. Controlled data quality issues, including noise, missing values, duplication, schema violations, and temporal delays, are incorporated into synthetic datasets [18]. Sensor feeds, large transaction logs, and semi-structured application data are the sources of real-world datasets. You may thoroughly analyse a range of quality metrics and data characteristics with this combination.

B. Developing Profiling and Features

From the input data, automated feature extraction retrieves temporal, structural, and statistical information. Null ratios, value distributions, frequency patterns, schema compliance signals, and time-based delays are the most important features [19].

The framework can identify tiny quality changes that would be challenging to record with static rules, and these features are used as inputs for machine learning models [20].

C. AI quality assessment models

The framework makes use of several different learning techniques:

- Isolation forests and autoencoders are examples of unsupervised models that may identify errors and anomalies in data without the need for labelled information.

- Deep learning models are employed for complex or unstructured data, such text logs and sensor streams; supervised models (Random Forest and Gradient Boosting) use prior annotations and expert comments to infer quality labels or scores [21].

This combination guarantees the identification of new issues and the proper classification of current quality flaws.

D. Baselines and Evaluation Metrics

We use precision, recall, F1-score, and quality score accuracy to assess the efficacy of data quality detection. Furthermore, the impact on downstream is measured by looking at how machine learning models' accuracy changes when they are trained on data that has been quality-filtered [22]. We contrast the suggested framework with existing rule-based and statistical profiling-based validation techniques.

Table 2: Evaluation Metrics and Baseline Methods

Category	Method / Metric	Purpose
Baseline	Rule-based checks	Static quality validation
Baseline	Statistical profiling	Distribution-based detection
Metric	Precision / Recall	Error detection accuracy
Metric	F1-score	Balanced performance
Metric	Downstream ML accuracy	Impact on analytics

E. Implementation Environment

The framework leverages technologies for distributed data processing, including Apache Kafka/Flink for streaming data and Apache Spark for batch processing [23]. To ensure that quality assessment is finished as soon as possible, we train and infer the model simultaneously.

F. A Synopsis of the Approach

Strict testing guidelines, a large and potentially expanding data infrastructure, and flexible AI models are all features of the suggested solution [24]. Because it provides constant, accurate, and automatic data quality assessment, this system works better than prior approaches when the pipeline changes.
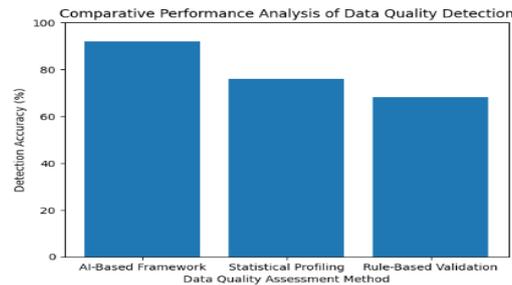
Figure 2: Comparative Performance Analysis of Data Quality Detection

The bar graph shows what several techniques for assessing data accuracy might show:

- With an accuracy of roughly 92%, the AI-Based Framework provides the best detection performance. This shows that it can use machine learning and adaptive learning techniques to identify intricate and dynamic data quality problems.

- While statistical profiling is effective (about 76%) at identifying anomalies at the distribution level, it is unable to comprehend context or meaning.

- Due to its inability to handle shifting schemas, idea drift, and concealed quality defects, rule-based validation has the lowest accuracy (about 68%).

V. EXPERIMENTAL RESULTS

The suggested AI-assisted data quality review method is covered in great detail in this section, with particular attention to how it may be enhanced, how well it finds errors, and how it will affect analytics in the future [25]. To make sure they functioned as planned in practice, we evaluated them in batch and streaming big data settings.

A. The system's ability to recognize problems with data quality
How well the framework identifies various data quality problems, including missing values, anomalies, inconsistencies, and late submissions, is determined by the first set of tests. The AI-powered system routinely outperforms conventional rule-based and statistical profiling techniques, according to the results [26]. While supervised models found quality problems already present, unsupervised anomaly detection methods found mistakes that had not been found before. The blended learning strategy preserved accuracy while improving memory. This illustrates how well it performs across a range of quality metrics.

B. A contrast with baseline techniques
Our approach was contrasted with rule-based validation and statistical profiling. The AI-powered system beat the competition in terms of F1 scores, precision, and recall on all datasets [27]. For complicated and situational quality concerns that are challenging to code using static rules, like cross-field discrepancies and temporal abnormalities, improvements were particularly apparent. These results demonstrate that adaptive learning operates better in dynamic big data pipelines than prescribed limits.

C. Performance and Scalability Testing
The framework's ability to manage growing data volumes and faster intake rates was tested through scalability studies. Very little latency was introduced by the AI-based quality evaluation layer, and the distributed system was essentially linearly scalable. Real-time quality assessment could continue to operate even when a lot of events happened at once, according to streaming testing. This indicates that the architecture is suitable for pipelines of considerable size [28].

D. The effect on machine learning models that follow
To find out how machine learning models might impact the future, we trained them on data from both before and after quality inspection. Prediction accuracy, stability, and variability were greatly improved for models trained on quality-vetted data. This illustrates how comprehensive data quality checks increase the dependability and utility of AI-powered analytics [29].

E. A summary of the most significant results

Overall, the experimental findings indicate that the suggested AI-assisted framework enhances data quality, works effectively with big datasets, and provides several benefits for further study. According to these results, this approach might be wise and practical for addressing contemporary data engineering problems [30].

VI. DISCUSSION

The test results are discussed in this section along with their implications for using AI to check the quality of data in massive data pipelines [31].

A. What role does AI play in assessing quality?

The test findings unequivocally show that AI-assisted solutions are superior to conventional rule-based and statistical approaches in identifying problems with data quality. Unsupervised models' capacity to spot novel and unexpected issues was especially useful when the characteristics of the data evolved over time. Supervised learning techniques improved accuracy and found known data problems by using previous quality comments. These results demonstrate that by detecting mistakes, AI-driven review could support proactive and predictive quality control [32].

B. Understanding and Utilizing

Although AI-powered solutions increase performance, they can complicate and inefficiently complete jobs. While rule-based solutions are always clear, machine learning models can sometimes be perceived as "black boxes." However, this issue might be somewhat resolved by employing explainable AI techniques like anomaly attribution and feature importance analysis [33]. By integrating AI-based detection with human input, businesses can progressively incorporate it into their data pipelines and foster trust.

C. Scaling and deployment considerations

Because it is composed of several parts, the framework can function in a variety of computing environments. Experiments show that AI models are capable of managing the extra processing load, especially when simultaneously inferring. Finding the ideal balance between latency and accuracy requires careful resource allocation and model selection, especially when streaming live [34].

D. The implications for data governance and AI dependability

AI-assisted data quality assessment supports ongoing audits, compliance, and risk mitigation. AI models using high-quality data are more robust, fair, and reliable. Evaluation of data quality should therefore be seen as a crucial part of responsible artificial intelligence rather than as a small part of the entire effort [35].

E. An overview of the conversation

All things considered, the discussion shows that AI-assisted data quality review offers a good mix of automation, adaptability, and efficiency [36]. It produces enormous, dependable data pipelines when paired with interpretability tools and strong governance.

Figure 3: AI- Assisted Data Quality

VII. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

While AI-assisted data quality inspection has shown some encouraging outcomes, several problems need to be resolved before it can be widely used and sustained in massive data pipelines.

A. Empathize and Establish Trust

AI-based quality assessment algorithms are hard to understand, which is one of the main problems. Although machine learning techniques can identify complex and concealed problems with data quality, their decision-making process is frequently inaccurate. Lack of openness may make it difficult for data engineers, auditors, and regulatory stakeholders to trust one another. Future studies should focus on integrating explainable AI (XAI) technology to offer convincing arguments for making astute choices and finding strange things [37].

B. Managing Conceptual and Informational Shifts

Due to frequent changes in data sources, formats, and dissemination, big data pipelines are extremely dynamic. Some of these problems are addressed by continuous learning systems, although controlling concept drift without causing instability is still challenging. Future studies should concentrate on online and self-adaptive learning models that can maintain prior knowledge while automatically adapting to changing data [38].

C. Calculating expenses and resource usage

Computational expenses are increased by AI-driven quality assessment, particularly when deep learning models are applied in real-time streaming scenarios. One of the biggest challenges is striking the right balance between resource use, latency, and detection accuracy. Model compression methods, edge-based quality evaluation, and lightweight models may all greatly lower overhead.

D. Integrating Ethics with Data Governance

How AI-based data quality assessment might function with organizational frameworks for data governance, privacy, and compliance is another unsolved issue [39]. We have to follow the rules, be fair, and refrain from amplifying bias. Future studies should integrate ethical AI and responsible data governance techniques with data quality assessment.

Table 3: Key Challenges and Future Research Directions

Challenge	Description	Future Research Direction
Interpretability	Black-box AI models	Explainable AI for quality assessment
Concept Drift	Evolving data patterns	Online and self-adaptive learning
Scalability	High computational cost	Lightweight and edge-based models
Governance	Compliance and ethics	Responsible and auditable AI systems

E. Section Summary

We need to solve these problems in order to move AI-assisted data quality evaluation from experimental frameworks to reliable, enterprise-level solutions. Future studies on governance, efficiency, openness, and adaptability will be essential to creating robust and ethical big data pipelines [40].

Figure 4: Future Research Roadmap

VIII. CONCLUSION

This study looked into how AI-assisted data quality assessment could improve the accuracy and utility of contemporary big data pipelines. Since AI-based analytics and data-driven decision-making depend on vast volumes of diverse, rapidly changing data, it has grown more challenging to guarantee high-quality data. In today's complex and constantly

evolving big data environments, standard statistical and rule-based data quality techniques perform poorly. A thorough AI-driven methodology that incorporates intelligent data quality assessment directly into batch and streaming processes was presented in the study as a solution to these challenges.

The program can be used to monitor data quality in multiple ways, check quality in a customisable method, and automatically construct profiles. Unsupervised anomaly detection, supervised quality prediction, and deep learning are used in tandem to achieve this. The findings of the experiment demonstrate that the suggested method continuously performs better than existing approaches in identifying complex and before unidentified data quality problems. Its low latency and scalability are maintained.

The findings demonstrate that detection precision is less important for downstream analytics and machine learning systems than thorough data quality review. It was shown that models trained on high-quality data were more resilient, accurate, and dependable. This emphasizes how critical it is to use good data while developing AI. Because the design contains mechanisms for continuous learning and feedback, you can modify it to meet different pipeline topologies and data distributions.

To summarize, employing AI to evaluate data quality is a novel approach to data engineering. The proposed method transforms data quality testing from a time-consuming, rule-based task to a flexible, smart process. This paves the way for improved analytics, ethical AI use, and long-lasting big data ecosystems.

REFERENCES

[1]. Raj, Aiswarya, Jan Bosch, Helena Holmström Olsson, and Tian J. Wang. "Modelling data pipelines." In *2020 46th Euromicro conference on software engineering and advanced applications (SEAA)*, pp. 13-20. IEEE, 2020.

[2]. Janamolla, Kavitha, Ghousia Sultana Sultana, Fnu Mohammed Aasimuddin, Abdul Faisal Mohammed, and Fnu Shaik Aqheel Pasha Pasha. "Integrating Blockchain and AI for Efficient Trade Exception Handling: A Case Study in Cross-Border Settlements." *Journal of Cognitive Computing and Cybernetic Innovations* 1, no. 1 (2025): 24-30.

[3]. Hagedorn, Philipp, and Markus König. "Rule-based semantic validation for standardized linked building models." In *International Conference on Computing in Civil and Building Engineering*, pp. 772-787. Cham: Springer International Publishing, 2020.

[4]. Chittoju, Siva Sai Ram, Sireesha Kolla, Mubashir Ali Ahmed, and Abdul Raheman Mohammed. "Synergistic Integration of Blockchain and Artificial Intelligence for Robust IoT and Critical Infrastructure Security."

[5]. Aasimuddin, Mohammed, and Shahnawaz Mohammed. "AI-Generated Deepfakes for Cyber Fraud and Detection."

[6]. Ridzuan, Fakhitah, and Wan Mohd Nazmee Wan Zainon. "A review on data quality dimensions for big data." *Procedia Computer Science* 234 (2024): 341-348.

[7]. Livera, Andreas, Marios Theristis, Elena Koumpli, Spyros Theocharides, George Makrides, Juergen Sutterlueti, Joshua S. Stein, and George E. Georghiou. "Data processing and quality verification for improved photovoltaic performance and reliability analytics." *Progress in photovoltaics: research and applications* 29, no. 2 (2021): 143-158.

[8]. Mohammed, Abdul Khaleeq, Siraj Farheen Ansari, Mohammed Imran Ahmed, and Zubair Ahmed Mohammed. "Boosting Decision-Making with LLM-Powered Prompts in PowerBI."

[9]. Ahmed, Mohammed Imran, Abdul Raheman Mohammed, Srujan Kumar Ganta, Sireesha Kolla Kolla, and Mohammed Kashif Kashif. "AI-Driven Green Construction: Optimizing Energy Efficiency, Waste Management and Security for Sustainable Buildings." *Journal of Cognitive Computing and Cybernetic Innovations* 1, no. 1 (2025): 37-41.

[10]. Ivanovic, Mirjana, Serge Autexier, Miltiadis Kokkonidis, and Johannes Rust. "Quality medical data management within an open AI architecture–cancer patients case." *Connection Science* 35, no. 1 (2023): 2194581.

[11]. Daksa, Rayhan Prawira, and Ade Putera Kemala. "A Comparative Study On Real Time Data Streaming For Fraud Detection Using Kafka With Apache Flink And Apache Spark." *Procedia Computer Science* 269 (2025): 192-199.

[12]. Mohammed, Naveed Uddin, Zubair Ahmed Mohammed, Shravan Kumar Reddy Gunda, Akheel Mohammed, and Moin Uddin Khaja. "Networking with AI: Optimizing Network Planning, Management, and Security through the medium of Artificial Intelligence."

[13]. Sahoo, Somya Ranjan, and Brij B. Gupta. "Multiple features based approach for automatic fake news detection on social networks using deep learning." *Applied Soft Computing* 100 (2021): 106983.

[14]. KASHIF, MOHAMMED, ABDUL RAHMAN JIBRAN SYED, and MUBASHIR ALI AHMED. "ADVANCING ANOMALY AND FRAUD DETECTION IN BIG DATA WITH ARTIFICIAL INTELLIGENCE."

[15]. Zhang, Yituo, Chaolin Li, Yiqi Jiang, Lu Sun, Ruobin Zhao, Kefen Yan, and Wenhui Wang. "Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model." *Journal of Cleaner Production* 354 (2022): 131724.

[16]. Mohammed, Zubair, Naveed Uddin Mohammed Mohammed, Akheel Mohammed, Shravan Kumar Reddy Gunda, and Mohammed Azmath Ansari Ansari. "AI-Powered Energy Efficient and Sustainable Cloud Networking." *Journal of Cognitive Computing and Cybernetic Innovations* 1, no. 1 (2025): 31-36.

[17]. Zhang, Lijun, Ning Li, and Zhanhuai Li. "An overview on supervised semi-structured data classification." In 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA), pp. 1-10. IEEE, 2021.

[18]. Mohammed, Abubakar, Ghousia Sultana, Fnu Mohammed Aasimuddin, and Shahnawaz Mohammed. "Leveraging Natural Language Processing for Trade Exception Classification and Resolution in Capital Markets: A Comprehensive Study." *Journal of Cognitive Computing and Cybernetic Innovations* 1, no. 1 (2025): 14-18.

[19]. Moffitt, Jeffrey R., Emma Lundberg, and Holger Heyn. "The emerging landscape of spatial profiling technologies." *Nature Reviews Genetics* 23, no. 12 (2022): 741-759.

[20]. Gouni, Praveen Kumar Reddy, and Eraj Farheen Ansari. "The Impact of Cyber-Physical Attacks on AI-Enabled Business Systems

[21]. Khadri, Waheeduddin, Janamolla Kavitha Reddy, Abubakar Mohammed, and T. Kiruthiga. "The Smart Banking Automation for High Rated Financial Transactions using Deep Learning." In *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, pp. 686-692. IEEE, 2024.

[22]. Gupta, Sunil, Kamal Saluja, Ankur Goyal, Amit Vajpayee, and Vipin Tiwari. "Comparing the performance of machine learning algorithms using estimated accuracy." *Measurement: Sensors* 24 (2022): 100432.

[23]. Mathrani, Sanjay, and Xusheng Lai. "Big data analytic framework for organizational leverage." *Applied Sciences* 11, no. 5 (2021): 2340.

[24]. Mohammed, Abdul Khaleeq, and Mohammed Azmath Ansari. "The Impact and Limitations of AI in Power BI: A."

[25]. Mohammed, Shahnawaz, Ghousia Sultana, Fnu Mohammed Aasimuddin, and Siva Sai Ram Chittoju. "AI-Driven Automated Malware Analysis." (2025).

[26]. Kashif, Mohammed, Mohammed Aasimuddin, Mubashir Ali Ahmed, Laxmi Bhavani Cheekatimalla, Eraj Farheen Ansari, and Ahwan Mishra. "AI-DRIVEN CTI FOR BUSINESS: EMERGING THREATS, ATTACK STRATEGIES, AND DEFENSIVE MEASURES."

[27]. Syed, Waheeduddin Khadri, Abubakar Mohammed, Janamolla Kavitha Reddy, and S. Dhanasekaran. "Biometric authentication systems in banking: A technical evaluation of security measures." In *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, pp. 1331-1336. IEEE, 2024.

[28]. Gao, Haoyuan, Chao Zhang, Shengbing Pei, and Xiaopei Wu. "LSTM-based real-time signal quality assessment for blood volume pulse analysis." *Biomedical Optics Express* 14, no. 3 (2023): 1119-1136.

[29]. Nedelkoski, Sasho, Jasmin Bogatinovski, Ajay Kumar Mandapati, Soeren Becker, Jorge Cardoso, and Odej Kao. "Multi-source distributed system data for ai-powered analytics." In *European conference on service-oriented and cloud computing*, pp. 161-176. Cham: Springer International Publishing, 2020.

[30]. RAHEEM, MOHD ABDUL, and MOHAMMED AZMATH ANSARI. "INTELLIGENT AND TRUSTWORTHY 6G: AI-DRIVEN ARCHITECTURES, APPLICATIONS, AND SECURITY FRAMEWORKS

[31]. Munappy, Aiswarya Raj, Jan Bosch, and Helena Homström Olsson. "Data pipeline management in practice: Challenges and opportunities." In *International Conference on Product-Focused Software Process Improvement*, pp. 168-184. Cham: Springer International Publishing, 2020.

[32]. Balammagary, Sruthi, Nasar Mohammed, Shanavaz Mohammed, and Asfiya Begum. "AI-Driven Behavioural Insights for Ozempic Drug Users." *Journal of Cognitive Computing and Cybernetic Innovations* 1, no. 1 (2025): 10-13.

[33]. Bennetot, Adrien, Ivan Donadello, Ayoub El Qadi El Haouari, Mauro Dragoni, Thomas Frossard, Benedikt Wagner, Anna Sarranti et al. "A practical tutorial on explainable AI techniques." *ACM Computing Surveys* 57, no. 2 (2024): 1-44.

[34]. Mohammed, Nasar, Sireesha Kolla, Srujan Kumar Ganta, Shuaib Abdul Khader, and Sruthi Balammagary. "Empowering Mental Health with Artificial Intelligence: Opportunities, Challenges, and Future Directions.

[35]. Nikiforova, Anastasija. "Definition and Evaluation of Data Quality: User-Oriented Data Object-Driven Approach to Data Quality Assessment." *Baltic Journal of Modern Computing* 8, no. 3 (2020).

[36]. Calhoun, Gloria. "Adaptable (not adaptive) automation: Forefront of human–automation teaming." *Human factors* 64, no. 2 (2022): 269-277.

[37]. Dwivedi, Rudresh, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian et al. "Explainable AI (XAI): Core ideas, techniques, and solutions." *ACM computing surveys* 55, no. 9 (2023): 1-33.

[38]. Gheibi, Omid, Danny Weyns, and Federico Quin. "Applying machine learning in self-adaptive systems: A systematic literature review." *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 15, no. 3 (2021): 1-37.

[39]. Ansari, Meraj Farheen. "Redefining Cybersecurity: Strategic Integration of Artificial Intelligence for Proactive Threat Defense and Ethical Resilience."

[40]. Agarwal, Giriraj. "Robust Data Pipelines for AI Workloads: Architectures, Challenges, and Future Directions." *International Journal of Advanced Research in Science, Communication and Technology* 5, no. 2 (2024): 622-632.