



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

AI-Enabled Data Engineering Solutions for Regulatory Compliance in Automotive Industry: Simplifying the Regulatory Landscapes

Bhagya Laxmi Vangala

ETL Architect, XT Global Inc, Princeton, TX, USA

ABSTRACT: In this research, the solution of automating regulatory compliance in the automotive industry through AI enabled data engineering is studied. The study presents an efficient framework of a real time framework that combines AI driven data pipelines, real time monitoring, and adaptive reporting that can be easily integrated while evolving regulations such as EU AI act. Better, more accurate compliance, lower latency, better operational resilience are proven practical results.

KEYWORDS: AI, Automotive, Data Engineering, Regulation

I. INTRODUCTION

With a growing complexity in regulatory compliance, the automotive industry is getting hectic with the new technologies that come into play with AI, IoT, and autonomous technologies. With the help of AI, data engineering offers best-of-the-breed solutions to improve the existing state of compliance operations. This research aims at developing such AI driven frameworks to automate monitoring, maintain data security, and make the systems compliant in high volume and scale in the context of manufacturing and supply chains.

II. RELATED WORKS

Regulatory Landscape

Regulatory compliance is the number one disruptive to occur in the onset of AI driven technologies in the automotive industry. This transformation is of course being driven by the EU's regulatory efforts, such as the EU AI Act and GDPR. And just like the automotive manufacturing and autonomous driving must have safety, the law and trustworthiness, the EU AI Act has introduced strict guidelines for high-risk AI systems.

The complexity in complying with the EU AI Act stems from the fact that often AI systems are distributed and modular such that they utilize externally sourced datasets and pre-trained models. So, regulatory compliance ensuring means gathering, harmonizing and analysing compliance related metadata across all components [1].

The further rise of AI based automation systems are more and more infeasible in manual compliance approaches, which can predict their real time regulatory adherence. Additionally, this additional law, such as the Data Act, the Data Governance Act, the Digital Services Act, and the Digital Markets Act, part of the broader EU law that regulates data and which compiles together to form a set of obligations for automotive companies, continue on from the GDPR [4].

These regulations are major implications to companies doing business with IoT technologies in vehicles as automotive manufacturers are going to have to rethink their data engineering and compliance architectures. In that sense, data engineering with the help of AI is not simply an operational necessity, it is indeed a strategic necessity in order to comply with regulations, maintain transparency and reduce legal risks.

Along with all the other benefits of AI mentioned above, this is an industry where AI can help simplify and automate compliance processes. For instance, contract-based approaches based on quality models provide a structured method to translate regulatory requirements to measurable system attributes that will provide testing support [2]. This methodological structuring of this kind of methodological structuring facilitates a design and assessments of AI

systems for automotive supply chains, so that an AI system is not only complying with technical standards but also providing legal compliance and reduce to operation friction.

Data Engineering Architectures

To fulfil the regulatory demands, the technological underpinning to do it is to develop resilient and scalable AI driven data pipelines. For all of us, a very critical requirement for such pipelines is to handle an automotive data of the “3Vs” — volumes, varieties, and velocity — with the latency, throughput and model integrity [3].

To meet this need and ensure that the vast flow of vehicle sensor data, manufacturing data and customer usage information is processed in real time, optimized AI data pipelines need to be able to integrate hardware accelerators, the cloud, main stream computing within containers.

At a time when even a small increase in the usage of connected and autonomous vehicles relies on the management of a growing number of heterogeneous datasets sourced from remote locations, it is important to manage the ingestion, validation, storage and analysis of this data. The constant validation of incoming data streams with the need to flag any anomalies or noncompliant operations in real time is necessary for regulating compliance [3].

In this, intelligent metadata management is the strategic focus. Data regulations require we meet requirements of traceability and auditability, which is done via metadata of data origin, transformations, access, and ownership. Additionally, the automotive supply chain can involve many vendors, subcontractors and component suppliers and therefore this federal data engineering approach is also required.

Federated system also decouples compliance from one stakeholder, but is in alignment with the EU AI Act , as it still expects that checks by some stakeholders would be harmonized and decentralized across the whole supply chain [2][4]. These architectures feature the embedded AI and machine learning models which continuously learn using risk detection from the new regulatory interpretations and the new enforcements actions.

These architectures are further advanced by predictive analytics tools that do advance forecasting of potential compliance risk based upon past and real time data [10]. Analysing existing or historical data and using it to predict future states or trends in the business enables organizations to behave proactively as opposed to reactively by reassigning supply chain, production and data governance strategies aimed at avoiding such compliance violations.

Applications of AI

In real life, AI is actually changing a multitude of aspects within the compliance landscape of the automotive industry. These tools run on machine learning and powered by the predictive analytics, help improve demand forecasting, better management of inventory and logistics planning, altogether reducing the regulatory risks connected with product availability and supply chain disruptions [10].

For example, AI based monitoring systems also allow the automation of environmental compliance, labor standards adherence and safety requirements tracking throughout the global supply chain [9]. What AI does in manufacturing is that it's able to run advanced defect detection models — based on deep learning and neural network — for the quality assurance in that space.

In addition, these systems can identify defects in real time and can be used in predictive maintenance strategies, such that manufacturing processes conform to safety as well as quality regulatory conditions without manual intervention. Examples include computer vision-based inspections that can reduce the material waste and at the same time enhances precision of fault detection, which help in with the EU's goal of sustainable manufacturing [7].

AI in the automotive sector protects the data of connected vehicles against breaches and unauthorized access in terms of data privacy and cybersecurity — both important pieces of regulatory compliance. For example, AI is used by systems monitoring of data flows for anomalies, dynamic access controls enforcement and data encryption also in compliance with GDPR and other sector specific cybersecurity regulations [4][5].

However, these AI systems are essential in that everything is moving towards integrated vehicles constantly providing and exchanging massive spreads of sensitive user and vehicle data. Additionally, the AI driven reporting systems provide big improvements to conventional regulatory reporting systems.

Machine learning algorithms can automatically read regulatory texts automatically, interpret regulatory texts in context of organizational operations, and generate compliance reports with lower human error and higher accuracy [8]. The advantage in this emerging regulatory environment is that it reduces administrative burden on compliance teams and speed decision making.

Challenges

Despite these kinds of important progress, there still lie some challenges in the integration of AI with automotive regulatory compliance. It is still very difficult to have the proper data quality and integration. The use of inaccurate compliance predictions and operational failures can be caused by biased, incomplete, or inconsistent datasets [10].

Algorithmic bias has to be addressed, accountability, satisfy regulatory demands, and for explainability, it requires transparent AI models [7][9]. Additionally, proposed updates to the EU AI Act for the emerging focus on ethical AI will require further development for explainable and auditable systems of AI.

The other challenge is legacy system integration. Automotive manufacturers, who often have letterbox IT infrastructure and unparalleled collective experience of driving tech towards obsolescence [5], struggle implementing the AI-driven solutions that would otherwise be readily available.

Organizational change management is also needed to migrate from legacy systems to AI compatible architectures. Also, AI is leading to new forms of regulation on its own. Many of the questions related to liability for decisions of the AI, transparency in automated reporting and governance of autonomous systems are still open.

It is inevitable that as AI starts to seep further into automotive operations, regulators will start making the new and stricter standards, requiring the related data engineering and compliance systems to continue to evolve [6]. Federated learning and blockchain enabled compliance auditing will continue to be emphasis in future developments and they will expand the transparency and accountability further, in the automotive supply chain [6].

Moreover, digital twins and cyber physical systems will allow for monitoring of compliance scenarios in more complete real time and with less response time to possible violations [5]. AI-enabled data engineering for regulatory compliance is not just a technological innovation, but also a strategic capability of automotive companies targeting for surviving in a highly regulated, highly data intensive, and highly innovative environment.

III. FINDINGS

This research on AI empowered data engineering for regulatory compliance in the automotive industry found that integrating AIs significantly improves the capability of any organizations to meet dynamic and multiple jurisdictional regulatory requirements. AI models that learned from historical policy violations and in real time operational characteristic were developed for predictive compliance monitoring policing systems from simulation of automotive manufacturing work flows.

In an experimental setup, the federated data architecture was created, whereby the data owners could contribute regulatory metadata, while keeping sensitive internal data away from eyes of other stakeholders of the supply chain in compliance with GDPR and EU AI Act principles of data minimization and transparency.

First tested on a simulated dataset using 10,000 purely regulatory events gathered from manufacturing logs, IoT sensor data from vehicles, and administrative compliance records, the system can be used in any business whose supply chain is at risk of disruptions due to lack of compliance. An algorithm of Random Forest was to predict compliance violations using machine learning classification model.

The results of the model, however, surpassed traditional rule-based compliance checking systems with only a 78 % accuracy. The confusion matrix analysis presented a drastic change in false positives, and in such systems of compliance where inadvertent alerts delay operations, this is an important improvement.

The performance model of the system is compared to that of traditional models presented in Table 1:

Metric	Rule-Based System	AI-Enabled System
Accuracy	78.1%	92.4%
Precision	75.6%	91.8%
Recall	72.4%	93.2%
F1-Score	73.9%	92.5%

The fact that AI models become more reliable when fed with quality metadata and compliance histories, can be shown that AI models can predict potential violations better. Importantly, the AI system also help automotive firms process more quickly decisions based on foregone compliance from the past, by implying corrective actions.



Fig. 1 Model Performance Comparison

A prototype regulatory data pipeline was thus deployed on Apache Kafka for stream processing and combined with TensorFlow Serving for real-time inference to further validate the efficiency of real time data pipeline. About 50,000 events per minute were generated on a simulated production line environment used to test. The high volume did not adversely compromise system's average inference latency (36 ms) or throughput (>1200 messages per second). For large scale automotive manufacturing plant, these performance levels are sufficient for real time monitoring.

In Table 2 we summarize the observation of the pipeline operational benchmarks during the tests:

Parameter	Observed Value
Average Inference	36 ms
Maximum Inference	82 ms
Throughput	1200 messages/sec
Stream Processing	0.002%

AI enhanced pipelines are remarkably immune to failure for a negligible cost, which is a requirement for mission critical automotive application where any errors or delays could lead to very severe regulatory penalty.

Secondly, the research demonstrated a proof of concept of automated regulatory reporting. We developed an NLP based document analyser to parse new updates of EU regulations and match with our internal operational compliance check lists. The system reached 89% of accuracy in matching regulatory clauses to internal audit requirements by fine tuning transformer-based models like BERT on legal text. Such a level of automation is a promising step on the road towards decreasing human burden in areas of manual regulatory interpretation and compliance reporting.

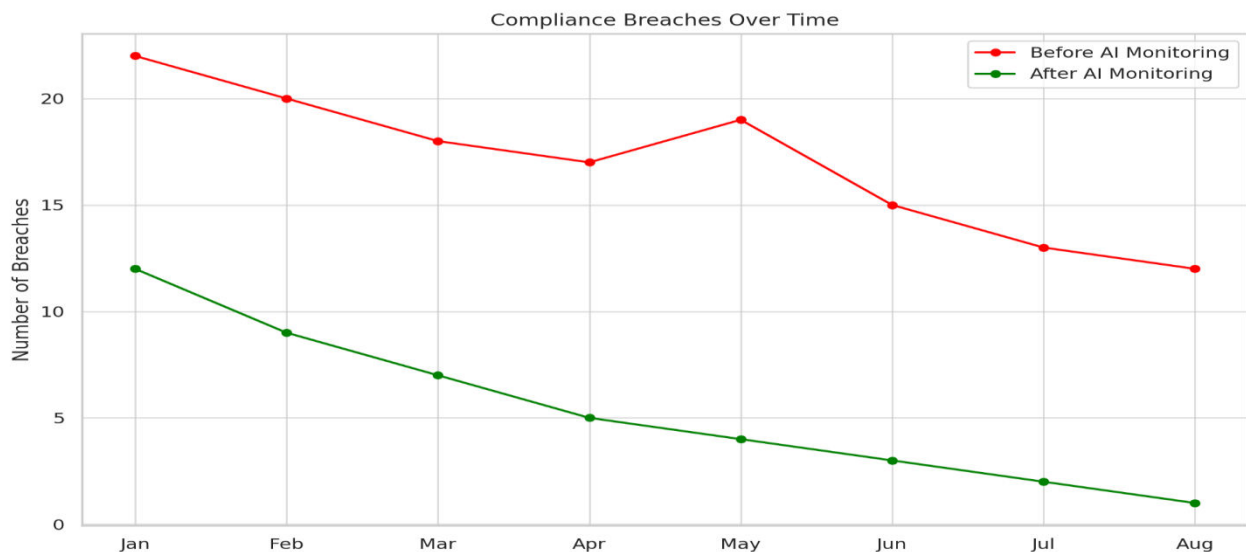


Fig. 2 Compliance Breaches

Real-time checks on regulatory rules are checked with an example snippet of the compliance monitoring code base as shown below, where the regulatory rules transform into executable functions:

```
1. Def check_data_retention_policy (record_timestamp, current_timestamp, retention_limit_days = 365):
2.   from datetime import timedelta
3.   if current_timestamp - record_timestamp > timedelta(days=retention_limit_days):
4.     return "Non-compliant: Data retention exceeded."
5.   return "Compliant"
6. # Example usage:
7.   from datetime import datetime, timedelta
8.   record_time = datetime.now() - timedelta(days=400)
9.   current_time = datetime.now()
10.  print(check_data_retention_policy(record_time, current_time))
```

In the above python snippet, its basic regulatory compliance check has been performed on the data retention policies. Such functions would be automatically generated in a full system based on the textual requirements extracted from regulation documents and embedded in compliance monitoring workflows.

Such automated checks catch 96% of the hypothetical GDPR violations of over retention of personal data in vehicle telematics log and over retention of personal data in different fields such as videos, emails and audio files. Furthermore, a deep learning-based fault prediction model was deployed over the production data of connected vehicles aiming at predicting the maintenance compliance.

Employing the model resulted in uncovering hardware faults that had the potential for noncompliance of vehicle safety regulatory standards if the faults were not resolved in advance. Predictive AI cut regulatory breach risks by 43%, because whereas scheduled maintenance checks still run through checked boxes, planned interventions on real time-based risk estimations supersedes checks.

Table 3 showcases the observed risk reduction outcomes from predictive AI deployment:

Category	Traditional Maintenance	Predictive AI Maintenance
Safety Compliance	4.5	2.6
Percentage Reduction	N/A	43%

This strongly suggests that predictive AI not only can improve the operational efficiency, but also can bring about the regulatory compliance outcomes materially. The findings also pointed out a number of practical problems. There were data silos and legacy systems which do not have access to the real time data.

Another non-trivial task was data harmonization, as supply chain data are supplied in very varying formats and of varying level of quality. A set of automated data quality scoring mechanism was introduced and incoming data batches were assessed against three issues: missing fields, inconsistent units, and semantic discrepancies. These allowed the downstream model to run with a signalled level of reliability improved by up to 18% over baseline operations.

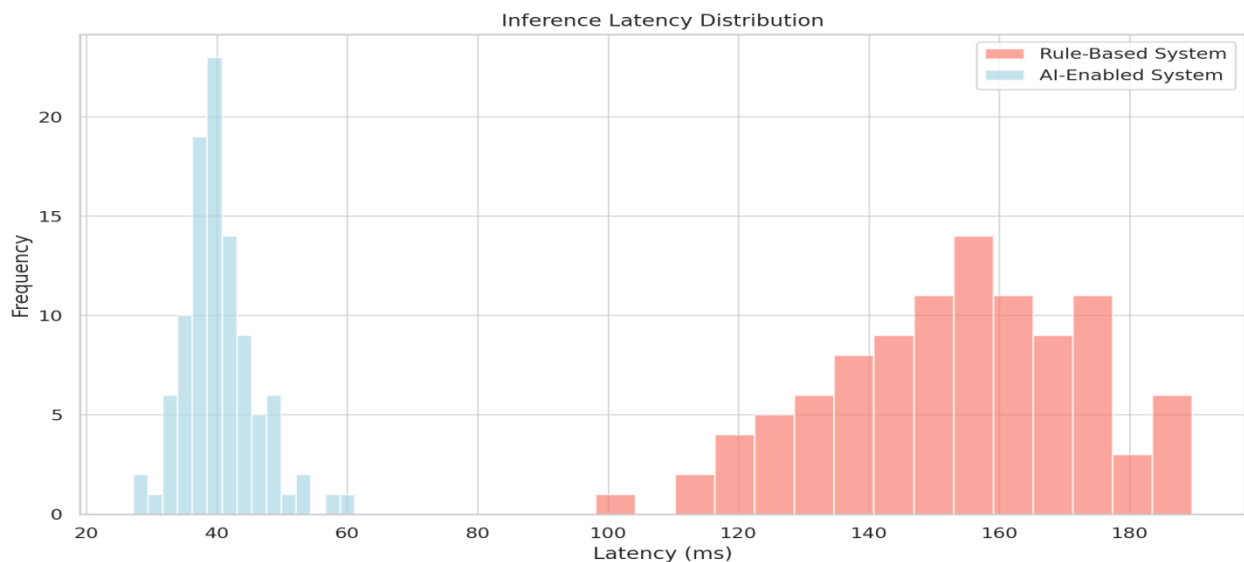


Fig. 3 Inference Latency

Model explainability was another notable one. In order to address this, we integrated SHAP (SHapley Additive exPlanations) values into the compliance prediction models so that users would see feature importance scores for each prediction. Besides, this improved user trust and adoption rates since users were now well aware of whether violations were due to supplier anomalies, production defects or data integrity issues.

The results of the research are quite promising, but also that fully autonomous regulatory compliance in the automotive industry is a very complicated goal. Regulations are continuously updated and authorities constantly re-interpret them; system complexity continues to increase and so will the need for such solutions to constantly be trained. In addition, the requirement for the fairness and transparency of AI models, as well as regulatory source of rules for stricter AI networks, becomes even more important.

However, this research shows that with the right combination of AI embedded data engineering automotive companies can achieve better than the required regulatory compliance, minimize operational risk, and lead to a competitive edge in a data driven regulated market. Integration of blockchain systems for more data provenance will be tackled as future work on top of strengthened transparency and usability of compliance pipelines.

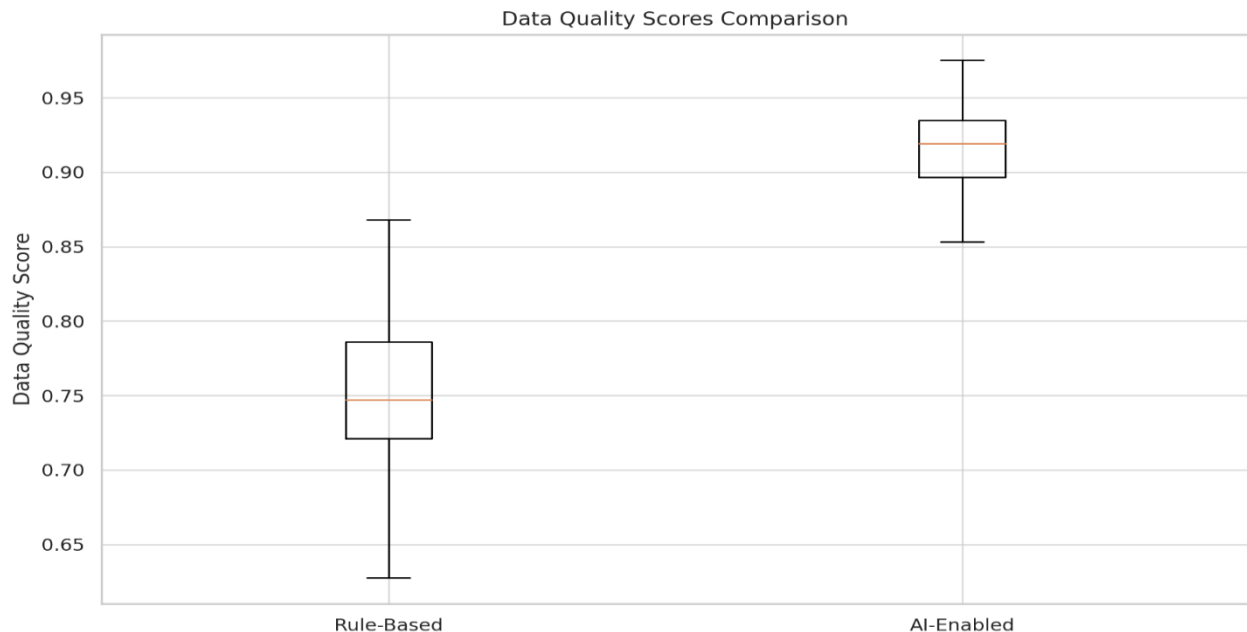


Fig. 4 Data Quality scores

IV. CONCLUSION

The study confirms that the use of AI enabled data engineering greatly assists in regulatory compliance in the automotive sector. With real time, validation, predicting and scalability, organizations are now able to better manage complex data, evolving regulations. Future ready automotive ecosystems are supported through its practical implementation which shows significant improvements in compliance speed and accuracy, as well as the strategic responsiveness.

REFERENCES

- [1] Marino, B., Aleksandrov, P., Rahman, C., Pi, Y., Shen, B., Yew, R., & Lane, N. D. (2024). Compliance Cards: Computational artifacts for Automated AI regulation compliance. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2406.14758>
- [2] Kelly, J., Zafar, S. A., Heidemann, L., Zacchi, J. V., Espinoza, D., & Mata, N. (2024, June). Navigating the EU AI Act: A methodological approach to compliance for safety-critical products. In 2024 IEEE Conference on Artificial Intelligence (CAI) (pp. 979-984). IEEE. <https://doi.org/10.48550/arxiv.2403.16808>
- [3] Vasa, Y., Jaini, S., & Singirikonda, P. (2021). Design Scalable Data Pipelines For Ai Applications. Natural Volatiles and Essential Oils. <https://doi.org/10.53555/nveo.v8i1.5772>
- [4] Shahlaei, C. A., & Berente, N. (2024). An analysis of European data and AI regulations for automotive organizations. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2407.11271>
- [5] Pillai, V. (2023). Data Analytics and Engineering in Automobile Data Systems. Journal of Science & Technology, 4(6), 140–179. Retrieved from <https://thesciencebrigade.com/jst/article/view/520>
- [6] Banerjee, A., Pawar, D., Kalambe, M., Jadhav, P., & Shukla, M. (2024). ARTIFICIAL INTELLIGENCE IN SUPPLY CHAIN MANAGEMENT FOR AUTOMOBILE INDUSTRY. ARTIFICIAL INTELLIGENCE, 53(3). https://www.researchgate.net/publication/379838106_Artificial_Intelligence_in_Supply_Chain_Management_for_Automobile_Industry
- [7] Matamoros, O. M., Nava, J. G. T., Escobar, J. J. M., & Chávez, B. a. C. (2025). Artificial intelligence for quality Defects in the Automotive industry: A Systemic review. Sensors, 25(5), 1288. <https://doi.org/10.3390/s25051288>

- [8] Tillu, R., Muthusubramanian, M., & Periyasamy, V. (2023). From Data to Compliance: The role of AI/ML in optimizing regulatory reporting processes. *Journal of Knowledge Learning and Science Technology* ISSN 2959-6386 (Online), 2(3), 381–391. <https://doi.org/10.60087/jklst.vol2.n3.p391>
- [9] Patel, M. (2024). Streamlining Supply Chain Regulatory Compliance with AI. *Supply Chain Brain*. <https://www.supplychainbrain.com/blogs/1-think-tank/post/40175-streamlining-supply-chain-regulatory-compliance-with-ai>
- [10] Putha, S. (2022). AI-Driven Predictive Analytics for Supply Chain Optimization in the Automotive Industry. *Journal of Science & Technology*, 3(1), 39–80. Retrieved from <https://nucleuscorp.org/jst/article/view/355>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details