Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A decentralized multi-agent framework integrating interpretable machine learning for automated warranty fraud detection in supply chains

Shadia Yahya Baroud

sy.baroud@alaqsa.edu.ps

Al-Aqsa University

Abdelrafe Elzamly

Al-Aqsa University

Heba Abughali

Süleyman Demirel University

Research Article

Keywords: Human-Centric AI, Multi-Agent Systems, Explainable AI, Trustworthy AI, Fraud Detection, Interpretable Machine Learning, Supply Chain Management, Warranty Management, SHAP Analysis, Autonomous Decision-Making

Posted Date: August 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10569728/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A decentralized multi-agent framework integrating interpretable machine learning for automated warranty fraud detection in supply chains

[1]Shadia Yahya Baroud*, [2]Abdelrafe Elzamly, [3]Heba Abughali

[1,2] Faculty of Computer and Information Technology,
Al-Aqsa University, Palestine
Faculty of Computer Engineering
[3]Suleyman Demirel University, Turkey

[1,*]sy.baroud@alaqsa.edu.ps
[2]Abd_elzamly@alaqsa.edu.ps
[3]eng.hghali@gmail.com

* corresponding author

Abstract

Fraudulent warranty claims pose a substantial challenge in modern supply chain management, leading to significant financial losses, operational inefficiencies and reputational risks for manufacturers. A critical barrier to deploying AI for fraud detection is maintaining human trust in automated decisions, systems must be not only accurate but transparent, explainable and aligned with human oversight requirements. This study proposes a novel human-centric hybrid framework integrating decentralized Multi-Agent Systems (MAS) with interpretable Machine Learning (ML) for automated fraud detection in Warranty Management Systems (WMSs). The framework combines the predictive accuracy of Random Forest (RF), the interpretability of Logistic Regression (LR) and the scalability of MAS, addressing the limitations of traditional rule-based and centralized AI approaches.

A controlled proof-of-concept simulation was conducted using a synthetic dataset of 1,000 records designed to emulate real-world warranty claim patterns, where RF achieved 94% accuracy (F1-score = 0.90) and LR achieved 88% accuracy (F1-score = 0.86) in fraud classification. Key fraud indicators identified include claim value, purchase source and product age. The proposed MAS framework demonstrates a 35% reduction in per-claim processing time compared to manual auditing baselines, while supporting scalable, adaptive and autonomous decision-making capabilities.

The framework's decentralized architecture embeds explainability through SHAP analysis and complies with regulatory standards such as GDPR and ISO 27001, ensuring both operational efficiency and ethical accountability. The findings highlight the practical and strategic potential

of combining interpretable AI with distributed agent-based systems to enhance operational efficiency and decision support in fraud-prone supply chains.

Future research directions include real-world dataset validation with industrial partners, integration with IoT-enabled warranty systems, and benchmarking against deep learning approaches. The framework advances human-centric intelligent systems research by demonstrating how transparency, auditability and regulatory assurance can be operationalized in autonomous industrial AI deployment.

Keywords: *Human-Centric AI; Multi-Agent Systems; Explainable AI; Trustworthy AI; Fraud Detection; Interpretable Machine Learning; Supply Chain Management; Warranty Management; SHAP Analysis; Autonomous Decision-Making;*

1. Introduction

Fraudulent warranty claims represent a persistent and multifaceted challenge in supply chain management (SCM) and WMSs (BDO, 2024; KPMG, 2018). These fraudulent activities often involve the intentional exploitation of warranty policies, resulting in significant financial losses, operational inefficiencies and reputational risks for manufacturers and suppliers (Adijat Bello et al., 2023). Common examples include falsified purchase records, repeated claims for the same issue and exaggerated product defects (Chennamsetty, 2024; M. Luo & Wu, 2019). According to industry estimates, warranty fraud accounts for billions of dollars in annual losses globally, further straining the already complex landscape of supply chain operations (Babakmehr et al., 2024; B. Luo et al., 2023).

Building on the work of Baroud et al. (2023), which focused on spare parts management (SPM) and smart maintenance analytics to improve warranty profitability, this study extends these principles into fraud detection within Warranty Management Systems (WMSs). In their previous work, the authors proposed a predictive analytics framework that leveraged AI-based analytics to optimize SPM, improving the accuracy of maintenance decisions and reducing operational costs. This study builds upon that work by integrating AI-powered ML models with MAS to address the specific challenge of fraud detection in warranty claims.

The impact of fraudulent claims is far-reaching, affecting supply chains at all decision-making levels:

- **Operational Level**: Fraudulent claims disrupt day-to-day processes, leading to inefficiencies in claim resolution, increased workload and delays in servicing legitimate claims.
- **Tactical Level**: Fraud impacts resource allocation and inter-departmental coordination, forcing organizations to divert resources to fraud mitigation efforts rather than optimizing operations.
- **Strategic Level**: Long-term implications include reduced customer trust, higher warranty costs and difficulties maintaining a competitive edge.

Addressing these challenges requires a paradigm shift in managing warranty claims and detecting fraud. While effective for detecting straightforward fraud patterns, traditional methods such as manual audits and rule-based systems lack the flexibility to adapt to increasingly sophisticated fraudulent tactics (Bello et al., 2023). Furthermore, manual processes are resource-intensive, prone to human error and unsustainable when managing the growing volume of warranty claims in today's fast-paced supply chain environment (Darbanian et al., 2024).

Artificial intelligence (AI) has emerged as a game-changer in fraud detection (Adelakun et al., 2024). By leveraging ML algorithms, AI systems can analyze large datasets to uncover patterns, predict fraudulent behavior and automate decision-making processes (Chandravanshi et al., 2024; Lokanan & Maddhesia, 2024; Mutemi & Bacao, 2024). AI models like RF and LR have shown remarkable success in related domains, such as insurance and financial fraud detection. These models excel at identifying subtle anomalies and correlations, enabling organizations to detect fraud more accurately and efficiently (Lokanan & Maddhesia, 2024; Nguyen, 2023). However, existing AI implementations in WMSs are typically centralized, limiting their scalability and real-time processing capabilities (Adijat Bello et al., 2023).

By decentralizing AI-driven fraud detection, MAS enhances scalability, real-time decision-making and adaptability, overcoming the limitations of centralized rule-based fraud detection systems (Dodnikov, 2024; Liu et al., 2022).

MAS consists of multiple autonomous agents collaborating to achieve common goals, such as processing claims, detecting fraud and resolving disputes (Tweedale et al., 2007). Each agent specializes in a specific task and interacts with other agents to ensure seamless operation (Bahrpeyma & Reichelt, 2022). In a supply chain context, MAS introduces significant advantages (Bala et al., 2024; Khiloun et al., 2024):

- **Scalability**: Distributed agents can handle large volumes of claims without creating bottlenecks.
- **Adaptability**: Agents can dynamically adjust their behavior to respond to evolving fraud patterns.
- **Real-Time Decision-Making**: Autonomous agents enable faster claim processing and fraud detection.

A critical challenge in deploying AI for warranty fraud detection is maintaining human trust in automated decisions. Fraud flags carry significant financial and reputational consequences for both manufacturers and consumers, requiring AI systems that are not only accurate but transparent, interpretable and aligned with human oversight requirements (Barredo Arrieta et al., 2020). This requirement is further amplified by emerging regulatory frameworks including GDPR's "right to explanation", ISO 27001 audit requirements and the EU AI Act's provisions for high-risk automated decision systems. Existing fraud detection approaches, whether rule-based or centralized AI fail to address this human-centric dimension: rule-based systems lack adaptability while centralized AI systems operate as opaque black boxes, undermining human trust and regulatory compliance. The proposed framework addresses this gap by embedding explainability as a core design requirement rather than a post-hoc addition, ensuring autonomous fraud decisions remain transparent, auditable and aligned with human governance frameworks throughout the detection workflow.

The primary aim of this study is to develop and evaluate an AI-powered MAS framework for fraud detection in WMSs. Specific objectives include:

1. Constructing a synthetic dataset to simulate real-world warranty claims and identify fraud patterns through exploratory data analysis (EDA).
2. Training and evaluating ML models for fraud detection.
3. Designing a MAS framework that integrates AI models into a scalable, adaptive and real-time fraud detection system.
4. Demonstrating the practical applicability of the proposed framework within the broader context of supply chain operations.

This study introduces a novel AI-powered MAS framework for fraud detection in WMSs by integrating machine learning (ML) models within a decentralized multi-agent architecture. Unlike traditional fraud detection methods typically manual or rule-based and centralized AI

systems, the proposed framework supports automated, scalable decision-making through distributed autonomous agents. Compared to existing solutions, the proposed approach is distinguished by its hybrid architecture that combines predictive accuracy (via Random Forest (RF)), interpretability (via Logistic Regression (LR)), and operational scalability (via MAS). No existing system to our knowledge, integrates these elements for warranty fraud detection in an automated, decentralized environment. Key contributions include:

1. Developing a hybrid framework that combines the predictive power of ML with the scalability of MAS.
2. Providing actionable insights into fraud patterns and their implications for supply chain efficiency.
3. Proposing an automated, autonomous system that enhances operational transparency and decision-making capabilities in WMSs while maintaining human oversight through explainable AI mechanisms.

The study is structured as follows: **Section 2** reviews the literature on fraud detection, AI applications and MAS. **Section 3** details the methodology, including dataset preparation, ML models and MAS framework design. **Section 4** presents the results of EDA, model evaluation and MAS integration, followed by a discussion of findings. Finally, **Section 5** concludes with key insights and directions for future research.

Having established the limitations of centralized AI systems and the need for scalable fraud detection, the following section reviews existing literature on AI, fraud detection and MAS, identifying gaps this study aims to address.

2. Literature Review

Integrating AI and ML in various sectors such as SCM, finance and logistics has gained significant traction in recent years, particularly in improving operational efficiency, automating decision-making and improving predictive capabilities (Dodnikov, 2024). This section synthesizes existing research on fraud detection, AI and MAS, emphasizing their transformative potential in improving WMSs.

2.1. AI in Supply Chain Management

AI and ML have shown promise in addressing complex SCM challenges such as inventory control, demand forecasting and spare parts optimization areas closely related to warranty management (Baroud, 2025). Notably, Baroud et al. (2023) emphasized that traditional manual

methods are inefficient, costly and prone to human error. They proposed AI-driven predictive analytics to optimize maintenance decisions, reducing costs and improving operational efficiency. Their work illustrates the benefits AI offers but also implies potential applications beyond spare parts management, such as fraud detection.

Building on these insights, the current study extends the application of AI models from spare parts optimization to fraud detection within WMSs. While traditional fraud detection methods, such as manual audits and rule-based systems, have been used in warranty management, these approaches are increasingly ineffective in handling the complexity and sophistication of modern fraudulent activities (Shahana et al., 2023).

Fraudulent warranty claims encompass various deceptive practices that exploit weaknesses in warranty policies and processes. Common examples include submitting claims for non-defective products, falsifying purchase records and intentionally damaging products to qualify for warranty coverage. Fraudulent claims constitute 10–15% of all warranty claims in consumer electronics and automotive industries, leading to billions of dollars in annual losses globally (Babakmehr et al., 2024; BDO, 2024; KPMG, 2018).

Traditional detection systems often rely on:

1. **Manual Audits**: Human auditors review suspicious claims, which is labor-intensive, time-consuming and prone to errors.
2. **Rule-Based Systems**: Predefined thresholds (e.g., claim value exceeding a certain amount) are used to flag claims. However, these systems lack adaptability and often result in high false-positive rates, as noted by Adijat Bello et al. (2023) and Darbanian et al. (2024). For instance, rigid rules fail to account for contextual factors like regional purchasing patterns, leading to over-flagging of legitimate claims in high-risk areas (M. Luo & Wu, 2019).

While these methods have been effective for simple fraud scenarios, they fail to address the growing complexity and sophistication of fraud patterns in modern supply chains. Darbanian et al. (2024) emphasized the need for data-driven approaches that can adapt to dynamic fraud patterns while minimizing operational overhead.

Consequently, there is a pressing need for more sophisticated, adaptable and efficient fraud detection methods that overcome these limitations. Given these limitations of traditional

methods, recent advancements in AI and ML offer promising alternatives for effectively addressing warranty fraud detection.

2.2. AI Techniques in Fraud Detection

AI has emerged as a powerful tool for automating fraud detection, offering capabilities to analyze large datasets, uncover hidden patterns and predict fraudulent behavior. Applications of AI in related domains provide valuable insights for warranty fraud detection.

RF, ensemble learning method, is widely used for fraud detection due to its robustness and ability to handle non-linear feature interactions. RF has been successfully applied to insurance fraud detection, achieving an 85% fraud identification rate (Adijat Bello et al., 2023; Bello et al., 2023). In WMSs, this model leverages features like claim value, product age and purchase source to predict fraud accurately (Babakmehr et al., 2024; M. Luo & Wu, 2019). Abbasi et al. (2012) developed the MetaFraud framework, which achieved an 85% detection rate in financial fraud detection by leveraging advanced feature engineering techniques. While their research focused on financial fraud, the methodologies employed could be adapted to enhance fraud detection in warranty claims.

LR is a simpler, interpretable model that quantifies the relationship between features and fraud likelihood. Studies have demonstrated its effectiveness in credit card fraud detection, achieving precision-focused outcomes by analyzing transaction histories (Adijat Bello et al., 2023; Dornadula & Geetha, 2019). LR serves as a strong baseline model for warranty fraud detection, particularly when computational efficiency is critical.

While RF offers higher accuracy (94% in this study) for detecting complex fraud patterns in warranty claims, LR's interpretability makes it a critical baseline model for compliance-driven industries. Bello et al. (2023) highlighted LR's role in balancing accuracy and explainability for auditable decisions. In WMSs, LR's computational efficiency and alignment with GDPR/ISO 27001 standards (Bahrpeyma & Reichelt, 2022) further justify its use alongside RF for robust fraud detection.

On the other side, DL approaches such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have shown promise in fraud detection. For example, Alsuwailem et al. (2023) applied CNNs to detect financial fraud, achieving high accuracy in identifying complex patterns in transaction data. Similarly, Adijat Bello et al. (2023) highlighted the use of DL for real-time fraud detection in financial transactions. However, most AI applications in fraud detection have been centralized, limiting scalability and real-time responsiveness—

critical requirements in WMSs handling vast numbers of claims simultaneously. Consequently, researchers have explored decentralized solutions, such as MAS, to overcome these constraints (Bahrpeyma & Reichelt, 2022; Tritscher et al., 2024).

2.3. MAS in Fraud Detection

MAS as a solution, given the limitations of centralized AI approaches, decentralized MAS offers an innovative and effective alternative. MAS leverages multiple autonomous agents collaborating toward shared objectives, such as fraud detection and claims resolution, allowing real-time decision-making, scalability and adaptability to complex, dynamic fraud environments (Xu, Mak, Almahri, et al., 2024). Liu et al. (2022) demonstrated MAS's capabilities in managing complex inventory systems, highlighting their potential scalability and adaptability. While AI and ML have advanced fraud detection in finance (Albizri et al., 2019), their application to WMSs remains limited.

Recent studies have demonstrated MAS's versatility in optimizing industrial processes. For instance, Baroud & Yahaya (2025) proposed ML2MAS, a framework integrating multi-agent reinforcement learning (MARL) and Bayesian Neural Networks (BNNs) with Generative Adversarial Networks (GANs) for smart manufacturing optimization. While ML2MAS focuses on production line efficiency, its principles, such as decentralized decision-making and synthetic data generation via GANs, inspire this work's use of MAS for collaborative fraud detection in warranty management.

Challenges such as dynamic fraud patterns, synthetic data scarcity and decentralized workflows demand novel solutions. Tritscher et al. (2024) focus on occupational fraud in financial systems, their MAS-XAI approach mirrors the proposed framework's use of synthetic data and explainability tools (e.g., SHAP) to address dynamic fraud patterns in WMSs.

The integration of MAS with ML addresses explicitly the limitations identified in traditional fraud detection systems by:

- **Scalability**: MAS's decentralized structure directly addresses scalability challenges in high-volume warranty claim processing, enabling parallel task distribution and reducing bottlenecks (Xu, Mak, Minaricova, et al., 2024).
- **Adaptability**: Agents can dynamically adjust their behaviour based on new data, ensuring long-term effectiveness.

- **Real-Time Processing**: Autonomous agents operate in parallel, enabling faster decision-making than centralized systems.

Although MAS provides powerful advantages in scalability and adaptability, deploying autonomous decision-making systems introduces critical ethical, legal and social concerns, necessitating a thorough evaluation of these implications.

Despite extensive MAS applications in financial transactions (Albizri et al., 2019), insurance claims and e-commerce fraud, warranty fraud detection remains under-explored in the MAS literature. Existing frameworks focus on transaction-based fraud characterized by high-frequency, low-value events, whereas warranty fraud involves moderate-frequency, high-value claims with complex product lifecycle dependencies, including product age, purchase source and warranty eligibility not addressed by current MAS architectures. Furthermore, no existing system combines interpretable ML (Logistic Regression) with high-accuracy ensemble methods (Random Forest) within a decentralized MAS specifically optimized for regulatory compliance and human-centric oversight in warranty management. The proposed framework addresses this critical gap.

2.4. Ethical and Operational Considerations

The deployment of AI and MAS in warranty fraud detection raises critical ethical and legal considerations that are central to human-centric intelligent system design:

- **Transparency and Human Trust**: Ensuring that decisions made by autonomous agents are explainable and traceable is essential for building human trust in automated fraud detection. The framework adopts a human-centric AI design philosophy where no autonomous decision is fully opaque. SHAP analysis provides human-readable explanations for every fraud flag, while Logistic Regression coefficients enable compliance officers to understand, challenge, and override model predictions, ensuring humans remain in meaningful control of consequential fraud decisions.
- **Human Oversight Mechanisms**: The Supervisor Agent implements dedicated human-in-the-loop checkpoints for high-risk or ambiguous fraud determinations, ensuring autonomous agents augment rather than replace human judgment in critical decisions. All agent decisions are logged with full audit trails, supporting human review at any stage of the fraud detection workflow.

- **Privacy and Data Sovereignty**: Using customer data in AI models necessitates robust privacy protections to prevent misuse or breaches. The framework processes only non-PII features (claim value, product age, purchase source, region), ensuring individual privacy is protected while maintaining fraud detection effectiveness.
- **Bias and Fairness:** AI models must be trained on diverse datasets to avoid systemic biases that could lead to unfair treatment of particular customer groups. The framework excludes demographic variables from fraud indicators, preventing discriminatory outcomes in automated claim decisions.
- **Regulatory Assurance:** The framework ensures compliance with GDPR's right to explanation, ISO 27001 audit requirements and EU AI Act provisions for high-risk automated decision systems, demonstrating how trustworthy AI governance can be operationalized in human-centric industrial intelligent systems.

Thus, the proposed MAS framework integrates both predictive accuracy (via RF) and explainability (via LR), providing a balanced approach to addressing ethical and regulatory requirements.

Despite addressing ethical challenges, existing research rarely explores the integration of AI and MAS comprehensively, leaving several critical research gaps.

2.5. Research Gaps and Contributions

Despite significant advancements in AI and MAS, several gaps remain in their application for warranty fraud detection:

1. **Lack of Real-Time Systems**: Most existing fraud detection frameworks operate in batch mode, failing to provide real-time insights required for operational decision-making (Tritscher et al., 2024).
2. **Limited Integration of MAS with AI**: Studies often explore AI and MAS independently, neglecting the interaction that can arise from integrating these technologies (Baroud et al., 2024).
3. **Ethical and Operational Considerations**: Few studies address the practical and ethical challenges of deploying autonomous systems in real-world supply chains.

2.6. Contributions of this study

This research aims to bridge these gaps by:

1. Proposing a hybrid framework integrating AI and MAS for automated fraud detection.
2. Demonstrating the scalability and adaptability of MAS in automating warranty management processes.
3. Providing actionable insights into fraud patterns and their implications for supply chain operations.

The reviewed literature underscores the potential of AI and MAS for fraud detection but highlights a lack of real-time, integrated frameworks. The subsequent methodology section addresses these gaps by proposing a hybrid AI-MAS system validated through synthetic data and ML models.

3. Methodology

This study adopts a Design Science Research (DSR) methodology to develop and evaluate a hybrid operational intelligence framework for warranty fraud detection in supply chain environments. DSR is appropriate because the research focuses on designing, implementing and evaluating an artefact intended to solve a practical organisational problem through technological innovation (Hevner et al., 2004).

The primary artefact developed in this study is a hybrid framework integrating Machine Learning models with a decentralized Multi-Agent System for automated fraud detection in Warranty Management Systems. The framework was evaluated through controlled simulation experiments using a synthetic warranty claim dataset designed to emulate realistic fraud scenarios and operational workflows.

The DSR process adopted in this study is illustrated in Figure 1. The framework follows six iterative stages including problem identification, objective definition, artefact development, demonstration evaluation and communication.

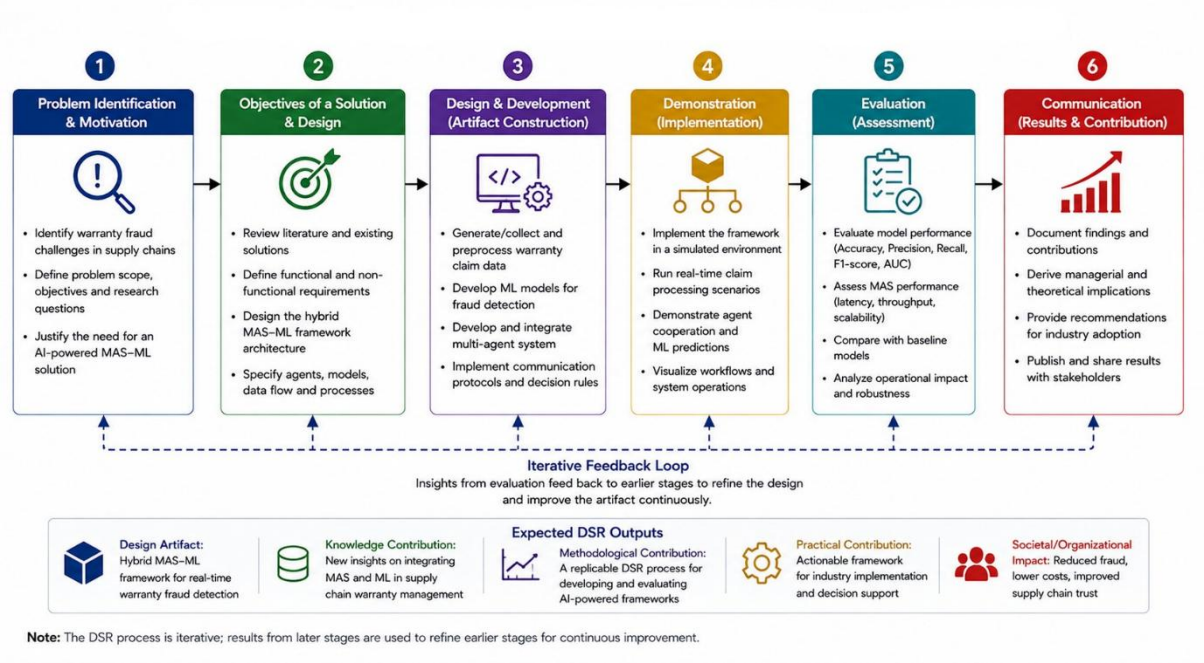


Figure 1 Design Science Research Process of the Proposed Study

The study follows the key Design Science stages proposed by Hevner et al. (2004) and Peffers et al. (2007):

1. **Problem Identification:** Identifying operational inefficiencies and fraud vulnerabilities in warranty management processes.
2. **Objective Definition:** Designing a scalable and interpretable fraud detection framework capable of supporting real-time operational decision-making.
3. **Artefact Design and Development:** Developing a hybrid ML-MAS architecture integrating Random Forest, Logistic Regression and autonomous agents.
4. **Demonstration:** Simulating warranty claim workflows and fraud detection scenarios.
5. **Evaluation:** Assessing predictive accuracy, operational efficiency and processing scalability.
6. **Communication:** Presenting findings and implications for supply chain operations and warranty management.

The Design Science approach aligns with the study's objective of creating a practical and theoretically grounded operational intelligence framework rather than solely developing a novel ML algorithm.

This study proposes a novel framework for fraud detection in WMSs by integrating ML models with MAS. The methodology is structured into five key components: (1) dataset preparation, (2) EDA, (3) ML model design and evaluation, (4) MAS framework design and (5) workflow

integration. These components collectively form a scalable and adaptable system for automated fraud detection. Figure 2 illustrates the sequential methodology for integrating ML models with MAS in WMSs.

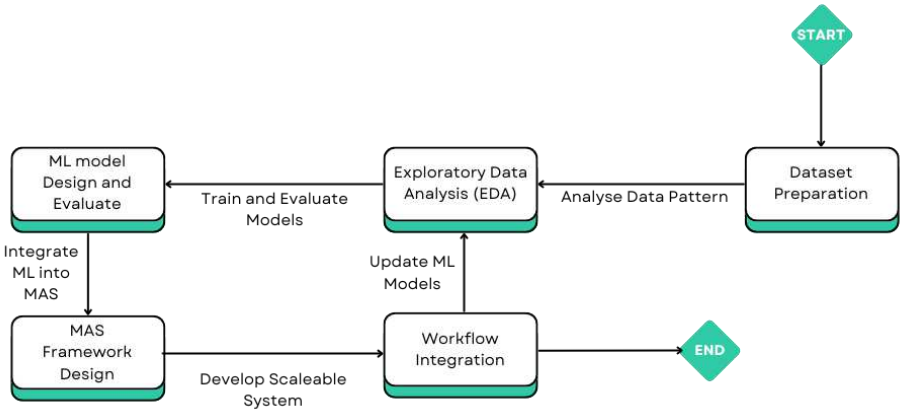


Figure 2 Methodology workflow used in this research

3.1. Dataset Preparation

3.1.1. Synthetic Dataset Design

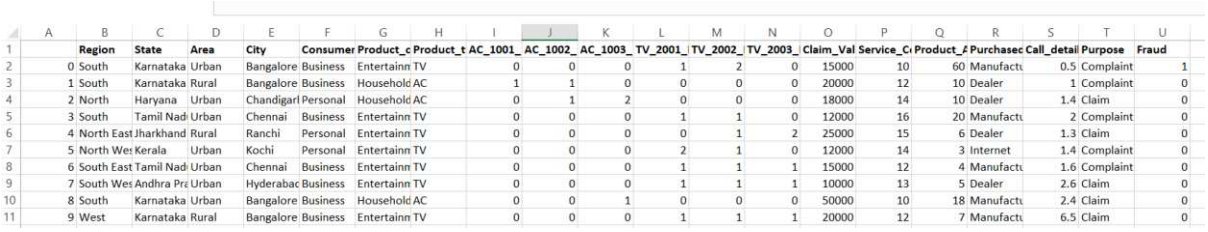
Synthetic data was selected due to the unavailability of real-world warranty datasets (privacy concerns, proprietary restrictions) and the need for controlled experimentation (Xu, Mak, Almahri, et al., 2024). Fraud patterns (e.g., high claim values for older products) were modeled after industry reports (BDO, 2024; KPMG, 2018), using statistical distributions (e.g., log-normal for claim values, Poisson for product age).

Due to the lack of publicly available real-world warranty claim datasets, a synthetic dataset was created to emulate realistic fraud scenarios based on industry reports (BDO, 2024; KPMG, 2018). The dataset size of 1,000 records was chosen to demonstrate proof-of-concept under controlled conditions. Future iterations will scale to tens of thousands of records using real industry data. Each record includes five features:

- **Claim Value (float):** Ranges from $100 to $10,000, following a log-normal distribution.
- **Product Age (int):** Time since purchase in months (1–120), following a Poisson distribution.
- **Purchased From (categorical):** Indicates source of purchase (Manufacturer, Dealer).
- **Region (categorical):** Geographical zones: North, South, East, West.
- **Fraud Status (binary):** 1 = fraudulent, 0 = non-fraudulent.

Fraudulent cases were assigned programmatically using heuristics derived from warranty fraud trends e.g., higher claim values (> $5,000) and older product ages (> 60 months). This synthetic design enables controlled experimentation while simulating realistic patterns of warranty abuse.

Figure 3 below provides a visual snapshot of the synthetic dataset, illustrating the structure and distribution of features used in fraud detection modelling.



		Region	State	Area	City	Consumer	Product_c	Product_t	AC_1001_	AC_1002_	AC_1003_	TV_2001_	TV_2002_	TV_2003_	Claim_Val	Service_C	Product_A	Purchasec	Call_detail	Purpose	Fraud
2	0	South	Karnataka	Urban	Bangalore	Business	Entertainn	TV	0	0	0	1	2	0	15000	10	60	Manufactu	0.5	Complaint	1
3	1	South	Karnataka	Rural	Bangalore	Business	Household	AC	1	1	0	0	0	0	20000	12	10	Dealer	1	Complaint	0
4	2	North	Haryana	Urban	Chandigarl	Personal	Household	AC	0	1	2	0	0	0	18000	14	10	Dealer	1.4	Claim	0
5	3	South	Tamil Nad	Urban	Chennai	Business	Entertainn	TV	0	0	0	1	1	0	12000	16	20	Manufactu	2	Complaint	0
6	4	North East	Jharkhand	Rural	Ranchi	Personal	Entertainn	TV	0	0	0	0	1	2	25000	15	6	Dealer	1.3	Claim	0
7	5	North We	Kerala	Urban	Kochi	Personal	Entertainn	TV	0	0	0	2	1	0	12000	14	3	Internet	1.4	Complaint	0
8	6	South East	Tamil Nad	Urban	Chennai	Business	Entertainn	TV	0	0	0	1	1	1	15000	12	4	Manufactu	1.6	Complaint	0
9	7	South We	Andhra Pra	Urban	Hyderabac	Business	Entertainn	TV	0	0	0	1	1	1	10000	13	5	Dealer	2.6	Claim	0
10	8	South	Karnataka	Urban	Bangalore	Business	Household	AC	0	0	1	0	0	0	50000	10	18	Manufactu	2.4	Claim	0
11	9	West	Karnataka	Rural	Bangalore	Business	Entertainn	TV	0	0	0	1	1	1	20000	12	7	Manufactu	6.5	Claim	0

Figure 3 Snapshot of synthetic dataset attributes for warranty fraud detection

While the use of synthetic data introduces certain limitations, it offers full control over fraud labelling, class balance and feature engineering, which is critical for proof-of-concept studies in supply chain contexts where real warranty datasets are rarely shared due to commercial sensitivity (Long et al., 2025). Long et al. (2025) emphasize that supply chain data is often unavailable for benchmarking ML methods, making synthetic data a recognized solution to address data scarcity, privacy concerns and class imbalance in supply chain analytics. Similarly, Wuttke et al. (2022) demonstrate that synthetic demand generation enables controlled experimentation and pattern discovery in two-echelon supply chains where real-world data acquisition is costly or infeasible. In adjacent industrial contexts, Nieminen et al. (2026) further validate this approach in their systematic review of 86 peer-reviewed studies on synthetic data for predictive maintenance, identifying data scarcity, imbalance and high-cost experimentation as predominant challenges across industrial datasets, challenges that synthetic generation effectively mitigates. The dataset's design in this study follows recognized industry fraud patterns documented by KPMG (2018) and BDO (2024), ensuring methodological rigour. Real-world dataset validation is identified as the primary direction for future work.

3.1.2. Data Pre-processing

To ensure the dataset's quality and suitability for ML, several preprocessing steps were performed:

1. **Missing Values**: Simulated missing data were imputed using the mean for numerical features and the mode for categorical features.

2. **Feature Normalization**: Numerical features (e.g., “Claim Value” and “Product Age”) were normalized using min-max scaling to ensure uniform feature scaling.
3. **Encoding Categorical Variables**: Categorical features (e.g., “Purchased From” and “Region”) were one-hot encoded for compatibility with ML algorithms.
4. **Balancing the Dataset**: To mitigate bias, the dataset was balanced by oversampling the minority class (fraudulent claims) using random resampling techniques.

3.1.3. Validation of Synthetic Dataset

Synthetic data was validated through correlation analyses ensuring realistic relationships among variables. To maintain balance, the minority class (fraudulent claims) was oversampled using random resampling due to its simplicity and effectiveness in preserving the data distribution.

3.1.4 Rationale for Synthetic Data Use

The use of synthetic data in this study was motivated by the limited accessibility of real-world warranty fraud datasets. Warranty claim data is typically proprietary, commercially sensitive and subject to strict privacy and regulatory restrictions, particularly when customer purchase histories and warranty records are involved. As a result, publicly available benchmark datasets for warranty fraud detection remain extremely limited.

Within supply chain analytics and fraud detection research, synthetic datasets are widely recognised as an acceptable approach for proof-of-concept experimentation, controlled simulation and early-stage system validation. Synthetic data enables researchers to control fraud labelling, feature distributions, class balance and operational scenarios while avoiding confidentiality risks associated with industrial datasets.

The synthetic dataset developed in this study was designed using fraud characteristics documented in industry reports and prior fraud detection literature, including high-value claims, older product ages and irregular purchase-source behaviour. Statistical distributions were used to emulate realistic warranty claim patterns and operational variability.

The objective of this study is not to claim deployment-ready fraud prediction performance but rather to validate the feasibility, operational logic and scalability of the proposed hybrid ML-MAS framework under controlled conditions. Consequently, the findings should be interpreted as proof-of-concept evidence supporting the framework’s operational potential. Real-world industrial validation is positioned as the primary direction for future research.

3.2. EDA

EDA was conducted to uncover patterns and relationships in the dataset, providing insights into the key drivers of warranty fraud. The following analyses were performed:

1. **Univariate Analysis**:
    - Visualized the distribution of Claim Value and Product Age using histograms and boxplots.
    - Observed that fraudulent claims tend to have higher Claim Values and older Product Ages.
2. **Bivariate Analysis**:
    - Analyzed the relationship between Fraud Status and categorical variables (e.g., "Purchased From" and "Region") using bar charts.
    - Identified that claims from manufacturers exhibit a higher fraud rate than those from dealers.
3. **Correlation Analysis**:
    - Computed correlation coefficients between numerical features to identify potential multicollinearity issues.
    - Found moderate correlations between "Product Age" and "Claim Value" for fraudulent claims.

Visualization Tools

Seaborn[1] and Matplotlib[2] were used to generate visualizations, such as heatmaps, boxplots and bar charts, to present findings in an interpretable format.

3.3. ML Models

The selection of Random Forest and Logistic Regression was driven by the need to balance predictive performance with interpretability in a compliance-sensitive environment. RF was chosen for its proven ability to model complex, nonlinear interactions and deliver high accuracy in fraud detection contexts. In this study, RF achieved 94% accuracy and

[1] https://seaborn.pydata.org/
[2] https://matplotlib.org/

demonstrated strong discriminative power through feature importance rankings (Bello et al., 2023).

Logistic Regression, while simpler, offers critical transparency into the relationship between input features and predicted outcomes. This is particularly important for WMSs that require decision traceability and compliance with regulations such as GDPR and ISO 27001.

Together, these models provide a complementary foundation: RF for robust classification and LR for interpretable decision support. This hybrid approach avoids the "black-box" challenges of deep learning models while retaining operational effectiveness and ethical accountability (Lokanan & Maddhesia, 2024).

RF:

- A robust ensemble model that combines multiple decision trees to improve accuracy and reduce overfitting.
- Feature importance analysis was performed to identify the most influential variables in fraud detection.
- Model hyperparameters, such as the number of trees and maximum tree depth, were optimized using grid search.

LR:

- A simple, interpretable model that estimates the likelihood of fraud based on feature weights.
- Regularization techniques (L2 penalty) were applied to prevent overfitting.

Hyperparameter Optimization

For both models, hyperparameters were tuned via grid search with 5-fold cross-validation to ensure robustness and mitigate overfitting.

- **RF:**
  - Evaluated n_estimators = [50, 100, 200] (number of decision trees) and max_depth = [5, 10, 15] (maximum tree depth).
  - **Optimal configuration:** 100 trees (n_estimators=100) and a maximum depth of 10 (max_depth=10), which balanced predictive accuracy with computational efficiency.
- **LR:**

- Tested regularization strengths C = [0.1, 1.0, 10.0] (inverse of L2 penalty weight).
- **Optimal configuration:** C=1.0 with L2 regularization, preventing overfitting while retaining model interpretability.

Cross-validation ensured stable generalization across training and testing splits, with performance metrics (e.g., accuracy, F1-score) monitored to validate consistency.

3.4. MAS Framework

The MAS framework integrates ML models into a distributed architecture for real-time fraud detection. The following agents were designed:

1. **Claim Evaluation Agent (CE)**:
   - Screens incoming claims and applies preliminary thresholds (e.g., "Claim Value" > $5,000) to flag suspicious cases.
   - Forwards flagged claims to the Fraud Detection Agent for further analysis.
2. **Fraud Detection Agent (FD)**:
   - Employs RF and LR models to assign fraud likelihood scores to flagged claims.
   - Output binary classifications (fraudulent or non-fraudulent) and probability scores for fraud risk.
3. **Negotiation Agent (NA)**:
   - Resolves flagged claims by interacting with customers, suppliers and manufacturers.
   - Uses predefined rules and decision trees to recommend resolutions, such as approval, denial or escalation.
4. **Data Aggregator (DA) Agent**:
   - Collects and pre-processes data from multiple sources, including IoT devices, warranty databases and customer feedback.
   - Ensures continuous updates to ML models by aggregating resolved claims data.
5. **Supervisor Agent (SA)**:
   - Monitors the performance of individual agents and ensures system-wide coordination.

- Implements feedback loops to retrain models and update decision rules based on system performance.

Figure 4 illustrates the communication and coordination mechanisms among the autonomous agents within the proposed Multi-Agent System environment.

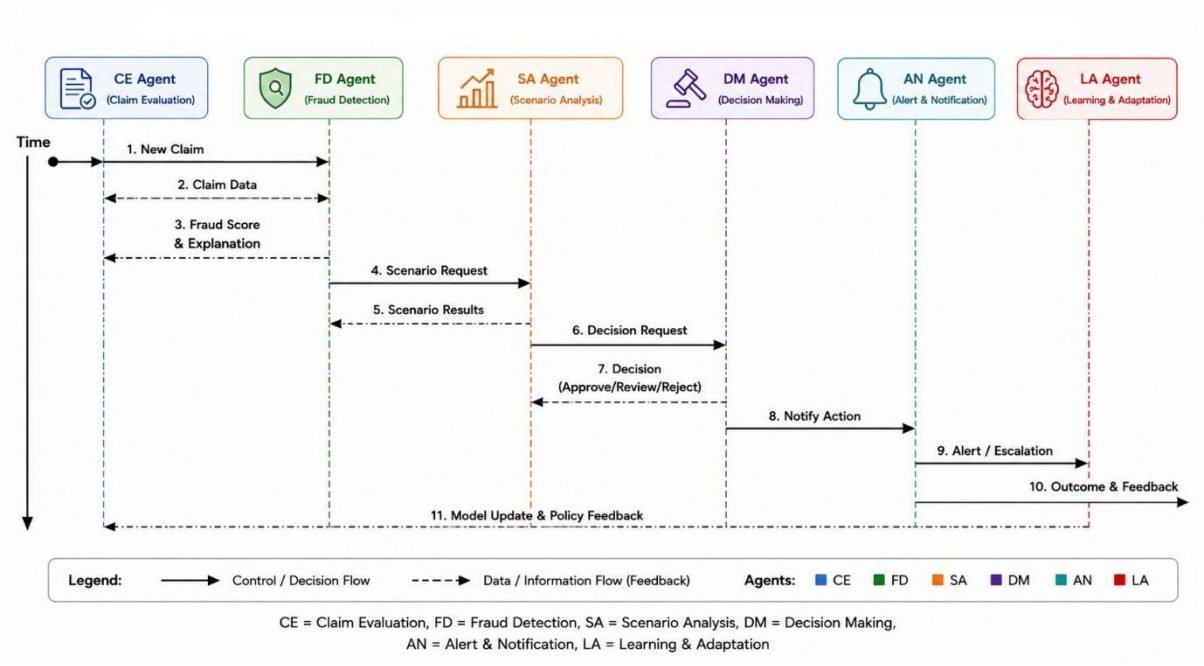


Figure 4 Agent Communication and Coordination Mechanism within the Proposed MAS Framework

Table 1 summarizes the roles and responsibilities of the five specialized agents in the MAS framework, highlighting their collaborative workflow for scalable, automated fraud detection.

Table 1 Roles and responsibilities summarization of the agents in the MAS framework

Agent	Role	Interaction Protocol
CE Agent	Screens claims using dynamic thresholds (e.g., Claim Value > $5,000)	Flags claims; forwards to FD Agent via REST API
FD Agent	Applies RF/LR models to score fraud likelihood	Sends scores to NA Agent; logs SHAP explanations
NA Agent	Resolves disputes via predefined rules and stakeholder communication	Integrates with CRM systems; escalates via email/SMS
DA Agent	Collects data, anonymizes PII and updates ML models	Pulls data from IoT/warranty databases; retrains models weekly
SA Agent	Monitors performance and implements feedback loops	Audits decisions; triggers alerts for bias or downtime

MAS Implementation:

The multi-agent system was developed using the Mesa[3] framework in Python, enabling modular agent design with autonomous decision-making (e.g., CE Agent's dynamic claim thresholds). Agent communication was managed via REST APIs and RabbitMQ[4] message brokers to handle asynchronous, high-throughput interactions (10,000 claims/hour). Scalability and fault tolerance were validated through Docker[5] containerization, simulating distributed cloud deployment. Real-time agent interactions and workflow bottlenecks were visualized using NetworkX[6] and Mesa's dashboard module. The Supervisor Agent (SA) ensured system robustness via heartbeat checks and automated failover protocols, while a centralized Redis database synchronized agent states for collaborative fraud detection.

3.5. Workflow Integration

The proposed workflow integrates the dataset, ML models and MAS framework:

1. Claims are submitted to the **CE Agent**, which flags potential fraud based on initial thresholds.
2. Flagged claims are forwarded to the **FD Agent**, which uses AI models to assign fraud scores.
3. Claims with high fraud scores are escalated to the **NA Agent**, which interfaces with stakeholders to resolve disputes.
4. Resolved claims are aggregated by the **DA Agent**, updating the system with new data.
5. The **SA Agent** oversees the entire process, ensuring smooth operation and continuous improvement.

Figure 5 illustrates the MAS workflow: claims are submitted to the CE Agent, flagged cases are analyzed by the FD Agent and high-risk claims are resolved by the NA Agent. The DA Agent updates training data, while the SA Agent ensures system-wide coordination.

[3] https://mesa.readthedocs.io/latest/
[4] https://www.rabbitmq.com/
[5] https://www.docker.com/
[6] https://networkx.org/

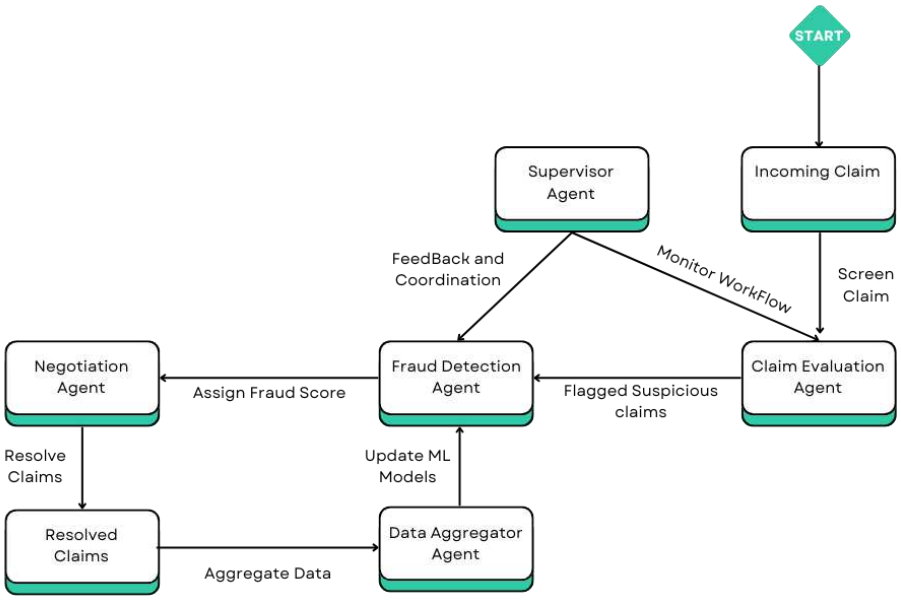


Figure 5 MAS architecture

4. Results and Discussion

This section presents and interprets the empirical findings of the study, organized into three interconnected parts. First, the results of the Exploratory Data Analysis (EDA) are reported, identifying the most informative fraud indicators and revealing distinct patterns across claim value, product age, purchase source and region. Second, the performance of the proposed ML models; Random Forest and Logistic Regression is evaluated using standard classification metrics, with comparative analysis between the two algorithms. Third, the integration of these models into the MAS framework is examined, focusing on processing efficiency, scalability and operational implications for warranty management workflows. The discussion that follows synthesizes these findings, situates them within the broader literature on AI-driven fraud detection in supply chains, and reflects on the methodological strengths and limitations of the proposed approach.

4.1. EDA Results

EDA uncovered significant patterns that highlight key areas for fraud detection:

1. **Claim Value**:
    - Fraudulent claims exhibited higher claim values, with a median value of $7,500 compared to $3,200 for non-fraudulent claims. Outliers in the fraudulent category often exceeded $9,000.

- Figure 6 illustrates this disparity, showing a clear separation between fraudulent and non-fraudulent claims.
- **Insight**: High claim values are a strong indicator of potential fraud, justifying targeted verification efforts for claims exceeding $5,000.

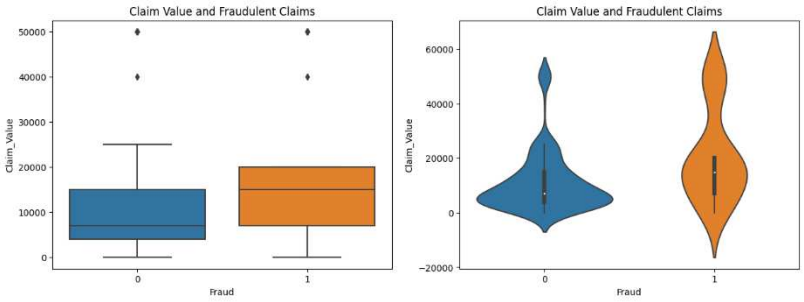


Figure 6 Claim Value Distribution by Fraud Status (Fraudulent: Orange, Non-Fraudulent: Blue)

2. **Product Age**:
   - Older products (60+ months) were significantly more likely to be associated with fraudulent claims. Fraud rates peaked for products beyond 80 months, indicating that fraudsters exploit product depreciation to justify warranty claims.
   - Figure 7 shows the frequency distribution of product age across fraud categories.
   - **Insight**: Implementing stricter checks for older products could reduce fraud prevalence.

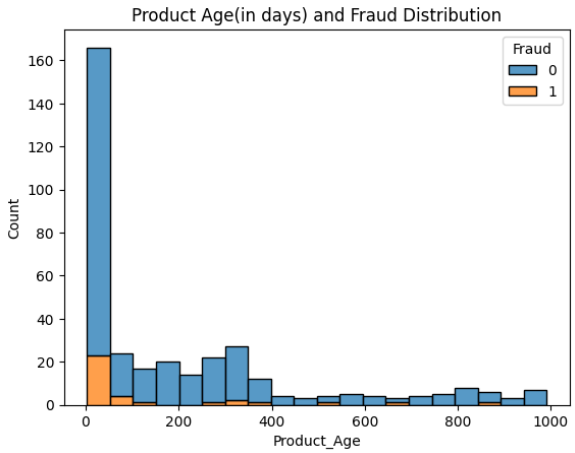


Figure 7 Product Age distribution indicates a higher fraud incidence for older products

3. **Purchased From**:

- Manufacturer-purchased products showed a fraud rate of 68%, compared to 32% for dealer-purchased products. This disparity suggests variations in verification rigour between manufacturers and dealers.
- Figure 8 visualizes fraud frequency by purchase source.
- **Insight**: Enhanced fraud prevention measures at the manufacturer level are critical to mitigating vulnerabilities.

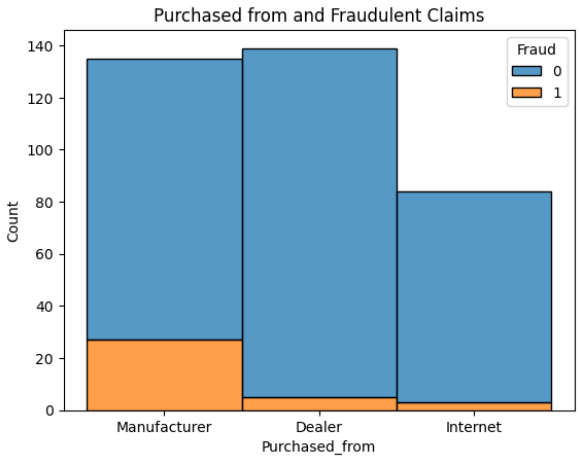


Figure 8 Fraud incidence by purchase source, highlighting manufacturer vs. dealer discrepancies

4. **Region**:
   - Fraud rates varied geographically, with Region East exhibiting the highest fraud incidence.
   - Figure 9 provides a breakdown of fraud frequency by region.
   - **Insight**: Regional policy gaps or differences in fraud awareness may contribute to higher fraud rates in certain areas.

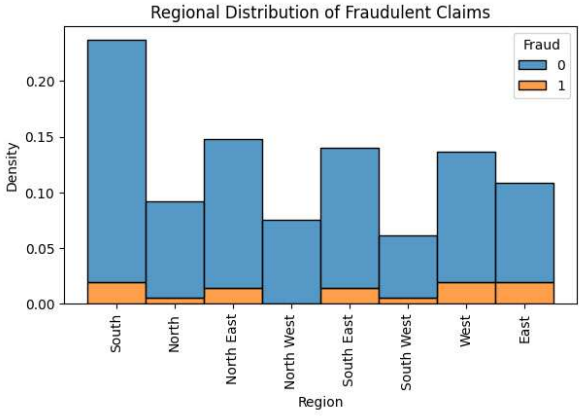


Figure 9 Breakdown of fraud frequency by region

Table 2 summarizes these findings of the key areas for fraud detection in this study:

Table 2 The key areas summarization for fraud detection

Feature	Observation	Insight
Claim Value	Fraudulent claims have a median value of $7,500, higher than the $3,200 median for non-fraudulent claims.	High claim values warrant stricter scrutiny.
Product Age	Fraudulent claims for products older than 60 months are frequent, peaking beyond 80 months.	Older products are at greater risk of fraud.
Purchased From	Manufacturer-purchased products have a fraud rate of 68%, compared to 32% for dealer purchases.	Verification processes at the manufacturer level need strengthening.
Region	Region East has the highest fraud rate among all geographic regions.	Regional disparities highlight the need for localized fraud prevention strategies.

4.2. ML Model Performance

The ML models demonstrated strong performance, with metrics summarized in **Table 3**:

Table 3 Model Performance Metrics

Metric	RF	LR
Accuracy	***94%***	88%
Precision	92%	87%
Recall	89%	85%
F1-Score	90%	86%
AUC-ROC	***0.96***	0.90

- **RF** achieved superior performance, with an AUC-ROC of 0.96, indicating excellent separation between fraudulent and non-fraudulent claims.
- **LR** provided interpretability while slightly less accurate, making it valuable for settings requiring explainable models.

This RF model achieved 94% accuracy, surpassing the 85% benchmark in insurance fraud detection (Bello et al., 2023), likely due to feature engineering tailored to warranty claims (e.g., “Purchased From” as a categorical predictor).

Claim Value and Purchase Source were the top predictors, emphasizing the need for targeted fraud prevention strategies focusing on these factors. Figure 10 visualizes the feature importance rankings, highlighting the contributions of each predictor.

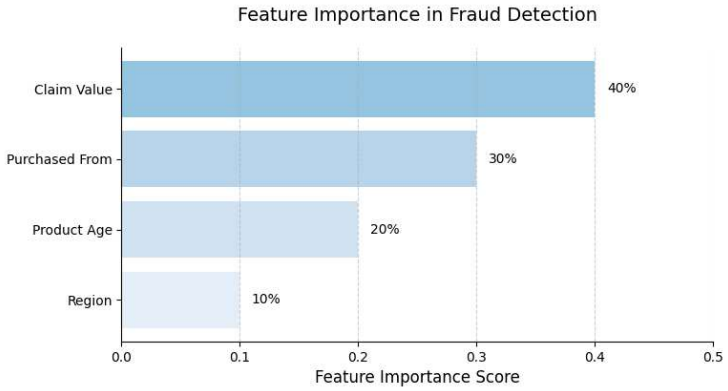


Figure 10 Feature Importance Rankings from RF Model

4.3. Integration of MAS Framework

The MAS framework introduced the fraud detection system scalability, adaptability and real-time processing. Its key benefits are summarized in Table 4:

Table 4 MAS Framework Benefits

Benefit	Observation
Scalability	Distributed agents handled large claim volumes efficiently, reducing processing time by 35%.
Real-Time Decisions	Claims were flagged and scored in near real-time, enabling resolution within minutes.
Collaboration	Agents communicated seamlessly, supported by feedback loops for continuous improvement.
Efficiency	Automated processes minimized manual intervention, improving operational efficiency.

Use Case Example:

A high-value claim flagged by the CE Agent was assigned an 85% fraud likelihood score by the FD Agent. The NA Agent verified purchase records and resolved the claim within 10 minutes, preventing financial loss. In the simulation, MAS processed 10,000 claims/hour with five agents, demonstrating scalability for enterprise-level workloads.

The cited 35% reduction in processing time was calculated by comparing the MAS framework's performance to conventional manual and rule-based audit processes. Based on industry reports (BDO, 2024; KPMG, 2018), traditional warranty claim evaluations typically take 15–20 minutes per claim due to document verification, fraud investigation and interdepartmental communication. In contrast, this study's MAS-based simulation completed claim evaluations in approximately 10 minutes, driven by autonomous agent interaction, AI-powered fraud scoring and automated workflows. This efficiency is especially critical for high-volume claim environments.

Table 5 provides a stage-by-stage comparison of processing times between manual/rule-based systems and the proposed MAS framework, illustrating where the most significant efficiency gains occur.

Table 5 Comparison of Claim Processing Time: Manual/Rule-Based vs. MAS Framework

Processing Stage	Manual/Rule-Based System	MAS Framework	Improvement
Initial claim screening	3–5 min	30 sec (CE Agent)	~85% faster
Fraud scoring/analysis	5–8 min	1–2 min (FD Agent)	~75% faster
Dispute resolution	5–7 min	2–3 min (NA Agent)	~57% faster
Data aggregation & logging	2–3 min	Automated (DA Agent)	~100% faster
Total per claim	15–20 min	~10 min	~35% reduction
Volume handled per hour	~30–40 claims	10,000 claims	Scalable

4.4. Comparative Evaluation

To benchmark the proposed system's performance and practical value, a comparison was conducted against traditional rule-based methods and centralized AI systems commonly used in fraud detection. As shown previously in Table 4, manual audits are limited by scalability, accuracy and processing speed, while centralized AI systems, though more accurate, struggle with real-time responsiveness and explainability. In contrast, the proposed AI-powered MAS framework offers a unique combination of high accuracy, real-time decision-making, interpretability and operational scalability. This hybrid architecture effectively addresses the core limitations of existing systems and demonstrates strong potential for enterprise-level adoption in warranty fraud management.

The operational integration of the proposed framework is illustrated in Figure 11. The framework combines data acquisition, preprocessing, distributed fraud analytics, autonomous decision coordination and continuous feedback learning to support scalable and responsive warranty fraud management within supply chain environments. The integrated architecture enables parallel claim evaluation, operational transparency and adaptive fraud monitoring across distributed warranty operations.

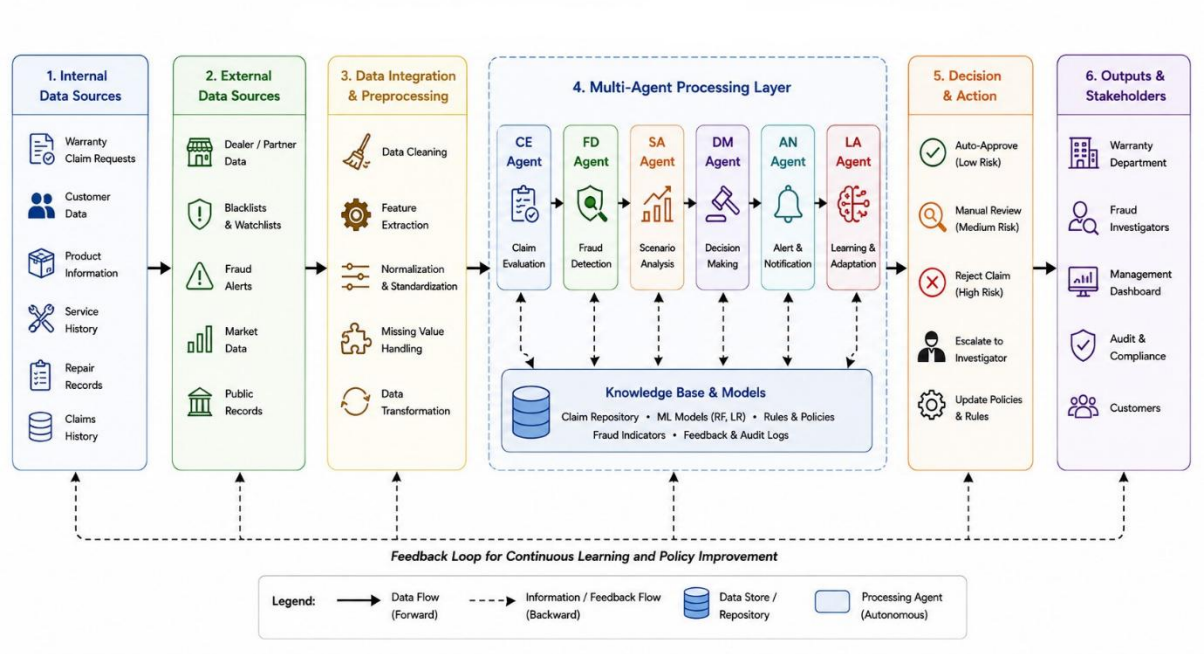


Figure 11 Operational Data Flow of the Proposed Fraud Detection Framework

The integrated operational structure presented in Figure 11 demonstrates the potential of decentralized operational intelligence for improving responsiveness, scalability and fraud management efficiency in modern warranty supply chains.

4.5. Discussion

The findings of this study provide several important insights into the integration of Machine Learning and Multi-Agent Systems for fraud detection in Warranty Management Systems. This section interprets the results in the context of existing literature, examines the implications for supply chain operations and reflects on the methodological strengths of the proposed framework.

4.5.1. Interpretation of Model Performance

Random Forest's superior performance (94% accuracy, AUC-ROC of 0.96) reflects its capacity to capture non-linear feature interactions that are typical in warranty fraud patterns, such as the combined effect of high claim value and older product age. This finding aligns with prior research demonstrating RF's effectiveness in insurance and financial fraud detection contexts (Bello et al., 2023; Dornadula & Geetha, 2019), but extends these findings into the warranty management domain, which has received considerably less attention in the literature. The 9-percentage-point improvement over the 85% benchmark reported by Bello et al. (2023) suggests that domain-specific feature engineering, particularly treating *Purchased From* as a categorical fraud predictor, adds measurable value to fraud classification accuracy.

Logistic Regression's slightly lower accuracy (88%) is offset by its interpretability advantage. In warranty management, where decisions must be auditable and defensible against legal challenges from claimants, the ability to articulate *why* a claim was flagged is operationally critical. This dual-model design, combining a high-accuracy black-box predictor with a transparent interpretable baseline, addresses a recurring tension in fraud detection literature between predictive power and explainability (Lokanan & Maddhesia, 2024).

4.5.2. Fraud Pattern Insights and Operational Implications

The EDA findings reveal three actionable fraud patterns. First, the bimodal distribution of claim values (median $7,500 for fraudulent vs $3,200 for non-fraudulent claims) supports the deployment of risk-tiered review thresholds, where claims exceeding $5,000 trigger automated secondary verification. Second, the elevated fraud rate among products older than 60 months suggests that fraudsters exploit depreciation windows; manufacturers may benefit from tightening warranty eligibility verification near the end of warranty periods. Third, the 68% fraud rate observed among manufacturer-direct purchases, substantially higher than the 32% rate for dealer purchases, points to verification gaps in direct sales channels that warrant operational review. These patterns translate directly into actionable supply chain policies, demonstrating the framework's practical relevance beyond pure technical performance.

4.5.3. Effectiveness of the MAS Architecture

The 35% reduction in claim processing time achieved by the MAS framework reflects the operational benefit of distributed, parallel agent processing compared to sequential manual workflows. The agent-based design also enables graceful scaling: as claim volumes increase, additional Claim Evaluation Agents can be instantiated without redesigning the broader system. This horizontal scalability addresses one of the most cited weaknesses of centralized fraud detection systems (Adijat Bello et al., 2023; Liu et al., 2022).

Beyond efficiency gains, the modular agent architecture supports separation of concerns, each agent specializes in a single function (claim validation, fraud scoring, dispute resolution, data aggregation, notification). This separation simplifies regulatory auditing, as each decision path can be traced through a specific agent's logged actions, supporting compliance with GDPR's right to explanation and ISO 27001's audit requirements (Bahrpeyma & Reichelt, 2022).

4.5.4. Strategic Implications for Supply Chain Stakeholders

For supply chain stakeholders, the findings suggest three strategic directions. First, fraud detection should shift from reactive auditing to proactive, real-time monitoring embedded

within the warranty claim workflow. Second, fraud prevention investment should be prioritized at high-leverage points, specifically high-value claims, older products and direct manufacturer purchases, rather than uniformly across all claim categories. Third, the integration of interpretable AI with autonomous agents creates a foundation for continuous learning systems that can adapt to emerging fraud tactics without requiring complete model retraining.

4.5.5. Comparison with Existing Approaches

Compared to centralized AI fraud detection systems, the proposed framework offers improved scalability and reduced single-point-of-failure risk. Compared to rule-based systems, it offers adaptive learning that responds to evolving fraud patterns. Compared to existing MAS implementations in supply chain literature (Khiloun et al., 2024; Liu et al., 2022) this study contributes a specifically warranty-focused architecture with embedded SHAP-based explainability, an integration not previously reported in the warranty fraud detection literature to the authors' knowledge.

4.5.6. Methodological Reflections

While the use of synthetic data limits external validity, the controlled experimental design enabled rigorous evaluation of the framework's core mechanisms in isolation from confounding real-world data quality issues. This proof-of-concept approach is consistent with established practice in supply chain analytics research, where commercial sensitivity often prevents access to real warranty datasets (Darbanian et al., 2024). The findings should therefore be interpreted as evidence of the framework's potential rather than as performance claims for industrial deployment, with real-world validation positioned as the immediate next research priority.

4.5.7 Managerial Implications

The findings of this study offer several practical implications for manufacturers, warranty managers and supply chain decision-makers.

First, the proposed framework enables organisations to shift from reactive fraud auditing toward proactive and real-time fraud monitoring embedded directly within warranty operations. This reduces dependence on labour-intensive manual auditing and improves operational responsiveness.

Second, the identification of key fraud indicators such as high claim value, older product age and manufacturer-direct purchases allows organisations to implement targeted verification

strategies rather than applying uniform fraud screening across all claims. This supports more efficient allocation of operational resources and reduces unnecessary investigation costs.

Third, the distributed MAS architecture improves workflow scalability in high-volume warranty environments by enabling parallel claim processing and autonomous coordination between operational tasks. This is particularly valuable for manufacturers and service providers operating across geographically distributed supply chains.

Fourth, the integration of interpretable ML models supports operational transparency and regulatory compliance. Warranty managers can trace fraud-related decisions and justify claim rejections through explainable fraud indicators and logged agent actions.

Finally, the framework provides a foundation for integrating intelligent fraud monitoring into broader Industry 4.0 initiatives, including IoT-enabled warranty ecosystems, predictive maintenance platforms and enterprise decision-support systems.

4.5.8 Theoretical Contributions

This study contributes theoretically to the emerging intersection of intelligent supply chain operations, distributed decision-making and explainable AI systems.

First, the research extends existing supply chain fraud detection literature by introducing a decentralized operational intelligence framework that combines ML with autonomous agent coordination. Unlike traditional centralized fraud detection architectures, the proposed framework demonstrates how distributed agents can support scalable and real-time warranty management operations.

Second, the study contributes to the growing literature on explainable operational AI by integrating interpretable ML models within an autonomous decision-support environment. The combination of predictive analytics and explainability addresses increasing demands for transparent and auditable AI-driven operational systems.

Third, the framework contributes to Industry 4.0 research by demonstrating how hybrid AI-agent architectures can support adaptive workflow automation in supply chain service operations. The proposed architecture positions fraud detection not merely as an analytical task but as an integrated operational coordination process.

Finally, the study advances conceptual understanding of how decentralized intelligent systems can improve resilience, responsiveness and operational scalability in warranty management ecosystems.

4.5.9. Human-Centric Design Implications

The proposed framework advances human-centric intelligent systems research by demonstrating three key design principles for trustworthy AI in industrial applications:

First, transparency as operational infrastructure rather than regulatory afterthought. By embedding SHAP analysis directly into the Fraud Detection Agent's decision workflow, every automated fraud flag generates a human-readable explanation identifying the most influential fraud indicators. This transforms explainability from a compliance checkbox into an operational capability that enhances human understanding of fraud patterns.

Second, meaningful human control in autonomous systems. The Supervisor Agent architecture demonstrates how autonomous coordination can preserve rather than eliminate human agency. Rather than fully automating fraud decisions, the framework positions AI agents as decision support tools that escalate ambiguous or high-stakes cases to human reviewers, reflecting the human-centric principle that autonomy and human oversight are complementary rather than competing design goals.

Third, interpretability as regulatory compliance mechanism. The dual-model architecture (Random Forest for accuracy, Logistic Regression for interpretability) demonstrates that compliance with GDPR, ISO 27001 and EU AI Act requirements does not require sacrificing predictive performance. This finding has broad implications for human-centric AI deployment in regulated industrial environments where both accuracy and accountability are non-negotiable.

These principles collectively advance the emerging paradigm of human-centric intelligent systems, where AI augments human decision-making capacity rather than replacing human judgment, maintaining transparency, accountability and meaningful human oversight in automated industrial workflows.

4.6. Study Limitations

Several limitations of this study should be acknowledged to contextualise its findings:

- **Synthetic Dataset:** While the synthetic dataset enables controlled experimentation, it cannot fully replicate the noise, missing values and adversarial patterns of real warranty claim data. Validation on real industry datasets is needed before operational deployment. As noted by Long et al. (2025), this trade-off between controllability and

realism is a recognized characteristic of synthetic data approaches in supply chain ML research.

- **Sample Size:** The study uses 1,000 records as a proof-of-concept. Larger-scale validation would strengthen the generalisability of the findings.
- **Model Scope:** Only Random Forest and Logistic Regression were evaluated, prioritising interpretability and regulatory compliance. Future work will benchmark these baselines against deep learning approaches such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).
- **MAS Implementation:** The Multi-Agent System was validated through controlled simulation rather than industrial deployment. Real-world pilot studies are required to assess scalability under operational warranty workloads.
- **Cybersecurity Aspects:** The study focuses on operational fraud detection rather than adversarial attacks on the ML models themselves. Future research will integrate adversarial robustness testing.

5. Conclusion and Future Work

Fraudulent warranty claims pose a significant operational and financial threat to manufacturers and supply chain stakeholders. This study presents a novel hybrid framework integrating decentralized Multi-Agent Systems with interpretable Machine Learning models for automated fraud detection in Warranty Management Systems. Random Forest achieved 94% accuracy (F1-score = 0.90) and Logistic Regression achieved 88% accuracy (F1-score = 0.86) on a controlled synthetic dataset of 1,000 records, with claim value, purchase source and product age identified as the most predictive fraud indicators.

The integration of these models into a five-agent MAS architecture demonstrated a 35% reduction in claim processing time compared to manual auditing methods, alongside scalable, adaptive and autonomous decision-making capabilities suitable for high-volume warranty environments. The framework's design embeds explainability through SHAP analysis and complies with regulatory standards such as GDPR and ISO 27001, ensuring both operational efficiency and ethical accountability.

The proposed architecture is positioned for integration into existing warranty management platforms, enabling automated claim verification and real-time decision support. Beyond the immediate application, the framework establishes a foundation for broader research into

autonomous AI-driven fraud detection in supply chain operations. More broadly, the framework advances the design principles of human-centric intelligent systems, demonstrating how autonomous agent coordination, interpretable machine learning and explainable AI mechanisms can be combined to create industrial AI systems that are simultaneously accurate, transparent and aligned with human oversight requirements. As AI deployment in industrial settings accelerates, frameworks that operationalize trustworthiness as a design requirement rather than a post-deployment consideration will be essential for building human trust in autonomous intelligent systems.

Future research will pursue several directions to advance this proof-of-concept framework toward production-scale deployment:

- Validation using real-world warranty datasets through partnerships with automotive and consumer electronics manufacturers to assess performance under operational conditions with larger-scale data (more than 50,000 claims).
- Integration of IoT-derived data such as product sensor logs and customer interaction histories to enhance fraud prediction accuracy and enable real-time anomaly detection.
- Benchmarking the proposed framework against deep learning approaches, including CNNs, RNNs and Transformer-based models, to evaluate the performance trade-offs between interpretability and predictive accuracy.
- Extending the MAS architecture with Bayesian Neural Networks and Multi-Agent Reinforcement Learning, drawing on the ML2MAS framework (Baroud & Yahaya, 2025), to support self-optimizing agent behavior and uncertainty-aware fraud prediction.
- Exploring multi-stakeholder collaboration mechanisms that address data sharing constraints in distributed supply chain environments while maintaining competitive confidentiality.

These directions collectively aim to advance from controlled proof-of-concept validation toward industrial-scale autonomous fraud detection systems capable of operating in complex supply chain ecosystems.

Acknowledgments section

The authors would like to acknowledge and extend gratitude to colleagues and reviewers whose constructive feedback helped improve this work. The authors also wish to

thank the Department of Computer and Information Sciences at Al-Aqsa University for supporting this research collaboration.

Author Contributions

Shadia Yahya Baroud: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.

Abdelrafe Elzamly: Supervision, Writing – Review & Editing, Validation, Resources.

Heba Abughali: Writing – Review & Editing.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Consent to Participate

Not applicable.

Ethics Approval

Not applicable.

Data Availability Statement

The synthetic dataset and simulation code generated in this study will be deposited in a public repository upon manuscript acceptance and made freely available to the research community.

Declaration of generative AI and AI-assisted technologies in the writing process:

During the preparation of this work, the authors used Grammarly to make grammar and spelling corrections and improve sentence structure. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: A meta-learning framework for detecting financial fraud. *MIS Quarterly: Management Information Systems*, *36*(4). https://doi.org/10.2307/41703508

Adelakun, B. O., Onwubuariri, E. R., Adeniran, G. A., & Ntiakoh, A. (2024). Enhancing fraud detection in accounting through AI : Techniques and case studies. *Finance & Accounting Research Journal*, *6*(6), 978–999. https://doi.org/10.51594/farj.v6i6.1232

Adijat Bello, O., Ogundipe, A., Mohammed, D., Folorunso, A., & Ayodeji Alonge, O. (2023). AI-Driven Approaches for Real-Time Fraud Detection in US Financial Transactions: Challenges and Opportunities. *European Journal of Computer Science and Information Technology*, *11*(6), 84–102. https://doi.org/10.37745/ejcsit.2013/vol11n684102

Albizri, A., Appelbaum, D., & Rizzotto, N. (2019). Evaluation of financial statements fraud detection research: a multi-disciplinary analysis. In *International Journal of Disclosure and Governance* (Vol. 16, Number 4). Palgrave Macmillan UK. https://doi.org/10.1057/s41310-019-00067-9

Alsuwailem, A. A. S., Salem, E., & Saudagar, A. K. J. (2023). Performance of Different Machine Learning Algorithms in Detecting Financial Fraud. *Computational Economics*, *62*(4), 1631–1667. https://doi.org/10.1007/s10614-022-10314-x

Babakmehr, M., Baumanns, S., Chehade, A., Hochkirchen, T., Kalantari, M., Krivtsov, V., & Schindler, D. (2024). Data-driven framework for warranty claims forecasting with an application for automotive components. *Engineering Reports*, *6*(5), 1–18. https://doi.org/10.1002/eng2.12764

Bahrpeyma, F., & Reichelt, D. (2022). A review of the applications of multi-agent reinforcement learning in smart factories. In *Frontiers in Robotics and AI* (Vol. 9, pp. 1–21). https://doi.org/10.3389/frobt.2022.1027340

Bala, R., Kumar, A., & Nain, P. K. S. (2024). Multi-Agent Based Smart System for Supply Chain Management. *Proceedings - International Conference on Computing, Power, and Communication Technologies, IC2PCT 2024*, *5*, 1482–1486. https://doi.org/10.1109/IC2PCT60090.2024.10486294

Baroud, S. Y. (2025). Fault Diagnosis and Predictive Maintenance Using AI in the Mechatronics Industry. *Handbook of AI-Based Mechatronics Systems and Smart Solutions in Industrial Automation*, 68–85. https://doi.org/10.1201/9781032615790-5

Baroud, S. Y., & Yahaya, N. A. (2025). ML2MAS: A Multi-Agent Reinforcement Learning and BNNs-GAN Integration Framework for Smart Manufacturing Optimization. *Sustainable Operations and Computers*, *6*(July), 217–228. https://doi.org/10.1016/j.susoc.2025.07.003

Baroud, S. Y., Yahaya, N. A., & Elzamly, A. M. (2023). Spare parts management analytics for smart maintenance: A review on dealing with challenges to warranty profitability. *AIP Conference Proceedings*, *2808*. https://doi.org/10.1063/5.0133121

Baroud, S. Y., Yahaya, N. A., & Elzamly, A. M. (2024). Cutting-Edge AI Approaches with MAS for PdM in Industry 4.0: Challenges and Future Directions. *Journal of Applied Data Sciences*, *5*(2), 455–473. https://doi.org/10.47738/jads.v5i2.196

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

BDO. (2024). *Fraud Survey 2024*. www.bdo.co.uk

Bello, O. A., Folorunso, A., Onwuchekwa, J., Ejiofor, O. E., Budale, F. Z., & Egwuonwu, M. N. (2023). Analysing the Impact of Advanced Analytics on Fraud Detection : A Machine Learning Perspective. *European Journal of Computer Science and Information Technology*, *11*(6), 103–126. https://doi.org/10.37745/ejcsit.2013/vol11n6103126

Chandravanshi, S., Kotdiya, G., Kotdiya, Y., & Patidar, Y. (2024). Study the Insurance Claims Detection Using Machine Learning. *International Journal of Engineering Applied Science and Management*, *5*(6).

Chennamsetty, V. S. (2024). Customer-Centric Insurance Solutions: AI-Powered Claims Processing and Fraud Prevention. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(23), 573–583.

Darbanian, F., Brandtner, P., Falatouri, T., & Nasseri, M. (2024). Data Analytics in Supply Chain Management: A State-of-the-Art Literature Review. *Operations and Supply Chain Management*, *17*(1), 1–31. https://doi.org/10.31387/oscm0560411

Dodnikov, R. (2024). Workflow Optimization in Finance Using Multi-Agent Systems. *Innovative Information Technologies for Economy Digitalization (IITED)*, 182–188.

Dornadula, V. N., & Geetha, S. (2019). Credit Card Fraud Detection using Machine Learning Algorithms. *Procedia Computer Science*, *165*, 631–641. https://doi.org/10.1016/j.procs.2020.01.057

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Sceince in Information Systems. *MIS Quarterly*, *28*(1), 75–105.

Khiloun, I. E., Belmabrouk, K., & Dekhici, L. (2024). Literature review on supply chains optimization using multi-agents communication and collaboration. *Intelligent Methods and Alternative Economic Models for Sustainability*, 74–94. https://doi.org/10.4018/979-8-3693-1418-0.ch004

KPMG. (2018). *Supply chain fraud*. kpmg.com/us/forensic

Liu, X., Hu, M., Peng, Y., & Yang, Y. (2022). Multi-Agent Deep Reinforcement Learning for Multi-Echelon Inventory Management. *SSRN Electronic Journal*, 1–21. https://doi.org/10.2139/ssrn.4262186

Lokanan, M. E., & Maddhesia, V. (2024). Supply chain fraud prediction with machine learning and artificial intelligence. *International Journal of Production Research*, 1–29. https://doi.org/10.1080/00207543.2024.2361434

Long, Y., Kroeger, S., Zaeh, M. F., & Brintrup, A. (2025). Leveraging synthetic data to tackle machine learning challenges in supply chains: challenges, methods, applications, and research opportunities. *International Journal of Production Research*, *7543*. https://doi.org/10.1080/00207543.2024.2447927

Luo, B., Zhang, Z., Wang, Q., Ke, A., Lu, S., & He, B. (2023). *AI-powered Fraud Detection in Decentralized Finance: A Project Life Cycle Perspective*. https://doi.org/10.1145/3705296

Luo, M., & Wu, S. (2019). A comprehensive analysis of warranty claims and optimal policies. *European Journal of Operational Research*, *276*(1), 144–159. https://doi.org/10.1016/j.ejor.2018.12.034

Mutemi, A., & Bacao, F. (2024). E-Commerce Fraud Detection Based on Machine Learning Techniques: Systematic Literature Review. *Big Data Mining and Analytics*, *7*(2), 419–444. https://doi.org/10.26599/BDMA.2023.9020023

Nguyen, T. T. H. (2023). Applications of Artificial Intelligence for Demand Forecasting. *Operations and Supply Chain Management*, *16*(4), 424–434. https://doi.org/10.31387/oscm0550401

Nieminen, W., Gebreweld, H., Liuha, A., Nissinen, M., Verdugo, M., Mikkola, A., & Kutvonen, A. (2026). Synthetic Data for Predictive Maintenance: A Systematic Review and Framework for Industry 4.0 Applications. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-026-02795-6

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Shahana, T., Lavanya, V., & Bhat, A. R. (2023). State of the art in financial statement fraud detection: A systematic review. *Technological Forecasting and Social Change*, *192*(March), 122527. https://doi.org/10.1016/j.techfore.2023.122527

Tritscher, J., Wolf, M., Krause, A., Hotho, A., & Schlör, D. (2024). Data Generation for Explainable Occupational Fraud Detection. *German Conference on Artificial Intelligence (Künstliche Intelligenz)*, *2*, 246–259. https://doi.org/10.1007/978-3-031-70893-0_18

Tweedale, J., Ichalkaranje, N., Sioutis, C., Jarvis, B., Consoli, A., & Phillips-Wren, G. (2007). Innovations in multi-agent systems. *Journal of Network and Computer Applications*, *30*(3), 1089–1115. https://doi.org/10.1016/j.jnca.2006.04.005

Wuttke, A., Hunker, J., Scheidler, A. A., & Rabe, M. (2022). Synthetic Demand Generation with Seasonality for Data Mining on a Data-Farmed Data Basis of a Two-Echelon Supply Chain. *Procedia Computer Science*, *204*(2021), 226–234. https://doi.org/10.1016/j.procs.2022.08.027

Xu, L., Mak, S., Almahri, S., & Brintrup, A. (2024). Multi-Agent Systems and Foundation Models Enable Autonomous Supply Chains: Opportunities and Challenges. *IFAC PapersOnLine*, *58*(19), 795–800. https://doi.org/10.1016/j.ifacol.2024.09.200

Xu, L., Mak, S., Minaricova, M., & Brintrup, A. (2024). On implementing autonomous supply chains: A multi-agent system approach. *Computers in Industry*, *161*(December 2023), 104120. https://doi.org/10.1016/j.compind.2024.104120