INTERNATIONAL JOURNAL OF INTELLIGENT DATA AND MACHIN LEARNING (IJIDML-3087-4262)

eISSN: 3087-4264

Volume. 03, Issue. 02, pp. 05-14, February 2026"

Ai In Dispute Management: Automating Resolution and Reducing False Claims in E-Commerce

Inna Simonova
Point Break Capital, Associate (Business Consulting)
Riverside, USA

Article received: 28/12/2025, Article Revised: 07/01/2026, Article Accepted: 19/01/2026, Article Published: 04/02/2026

ABSTRACT

Against the backdrop of accelerating digitalization of the global economy and the expansion of electronic commerce, dispute management is transforming into one of the key factors of business financial sustainability. The most problematic manifestation is friendly fraud: in 2025, it accounts for up to 61% of all transactions for which disputes are initiated. Under these conditions, traditional approaches to claims handling, based on manual processing and a set of heuristic rules, demonstrate limited effectiveness due to the scalability of attacks and the increasing complexity of behavioral scenarios of abuse. This paper describes the specific features of applying modern artificial intelligence algorithms to automate dispute resolution, including graph neural networks (GNN) and large language models (LLM) integrated into a RAG architecture. The empirical basis is formed through the analysis of academic publications and industry reports, supplemented by an in-depth examination of the Social Discovery Group case, which makes it possible to substantiate the practical viability of hybrid AI solutions. The results obtained indicate that the implementation of automation can reduce dispute processing time from several hours to seconds, increase the win-rate coefficient in the digital goods segment to 72%, and simultaneously significantly reduce operational costs. Thus, the study is embedded in the context of the development of FinTech risk-management methodology and forms an architectural approach to designing autonomous systems oriented toward revenue protection.

Keywords: dispute management, e-commerce fraud, friendly fraud, chargebacks, graph neural networks, LLM, RAG.

Introduction

The post-pandemic period was accompanied by a sustained shift in consumer practices toward online channels, which became one of the key drivers of accelerated growth in electronic commerce.

Simultaneously with the expansion of digital markets, the intensity of financial abuses also increased, acquiring characteristics of systemic risk. According to current estimates, global losses from fraud in e-commerce demonstrate a pronounced upward trajectory: it is expected that by 2029 they will reach 107 billion US dollars, which is equivalent to an increase of 141% relative to the levels of 2024 [1]. For 2025, damage on the order of 48 billion dollars is forecast, that is, approximately 16% higher than the values of the previous year [1, 3].

For retailers and digital platforms, such dynamics are directly converted into erosion of operating margin. The average reduction in global revenue of e-commerce companies due to fraud is estimated at 2.9% [2]. At the same time, the actual cost of fraud is not exhausted by the loss of a good or service. A substantial share of aggregate damage is formed by chargeback fees, penalty sanctions from payment systems, non-recoverable logistics expenses, and costs for dispute support. Empirical studies

show that one dollar of direct losses as a result of fraud in fact transforms for the retailer into 3.35–4.61 dollars of total costs [4].

In 2024–2025, a noticeable redistribution of the threat profile occurred, which complicated traditional counteraction models. If earlier scenarios associated with compromise of payment data and misuse of stolen cards dominated, then in the current risk structure the leading role was assumed by friendly fraud. This phenomenon describes a situation in which a legitimate cardholder initiates a purchase and receives a good or service, and then launches a return procedure through the issuing bank, claiming non-authorization of the transaction or non-receipt of the order.

Statistical observations indicate the dominant nature of this vector: up to 61% of all disputes in 2025 are classified as friendly fraud [1], and 72% of merchants recorded an increase in such cases in 2024 [4]. An additional indicator is the intensification of return abuses, which increased by 48% in 2024, and the expansion of practices of manipulating loyalty programs, the growth of which is indicated by 57% of merchants [1]. The behavioral nature of the phenomenon is characterized by high heterogeneity: in some cases, initiation of a chargeback is caused by cognitive errors and misunderstanding of the bank statement, in others it represents a deliberate strategy of cyber shoplifting based on exploitation of procedural vulnerabilities in dispute mechanisms. Generational differences further increase variability of motives: representatives of Generation Z, according to observations, account for up to 60% of chargebacks associated with regret over an impulsive purchase, whereas millennials are 30% more likely to dispute charges under subscription models [6]. Social networks function as a risk amplifier: 27% of consumers encounter content that normalizes practices of free receipt of goods through manipulation of disputes [6].

The established configuration of threats reveals the limits of the effectiveness of traditional dispute management. The procedure for contesting a chargeback represents a formalized process with high transactional complexity: it requires collection and verification of an evidentiary base (event logs, delivery tracking numbers, correspondence, confirmations of digital delivery), its alignment with the requirements of specific Visa and Mastercard reason codes, as well as compliance with strict regulatory deadlines for submission of materials. Under conditions of manual processing, one dispute requires from 2 to 5 hours of work by a qualified specialist [8]. With the estimated cost of processing one incident in the range of 35–100 dollars [7], disputing transactions with a low ticket often turns out to be economically irrational, which in practice forms a favorable environment for repeated abuses. An additional problem is the lack of scalability: seasonal surges in activity (for example, during sale periods) are capable of overloading risk departments and leading to the accumulation of queues. The time lag for dispute resolution often constitutes 30–150 days, which increases the duration of freezing of working capital and worsens the financial liquidity of the business [9, 10].

Within the framework of the study, a methodological approach is formed for automating dispute management by means of hybrid artificial intelligence systems that combine models of various classes. The concept is based on the use of graph neural networks (GNN) for preventive identification of anomalous interrelations and patterns of first-party misuse, as well as natural language processing (NLP) solutions for automated formation of the evidentiary base and preparation of materials required for representment. It is assumed that integration of such components is capable of simultaneously reducing operating costs and increasing the share of won disputes (win rate) by accelerating decision-making, standardizing evidence, and increasing its relevance to the requirements of payment regulations. The research logic is built from a description of the applied methodology and the source base to an analysis of the taxonomy of disputes and the regulatory frameworks of payment systems, after which the architecture of the proposed AI solutions (GNN and RAG) is considered, results are synthesized on the material of the Social Discovery Group case, and ethical and regulatory limitations of implementing such systems are discussed.

The purpose of the work consists in the development and substantiation of an architectural approach to automation of dispute management in e-commerce based on hybrid AI (GNN + LLM/RAG), oriented toward acceleration of representment and reduction of the share of unfounded disputes while simultaneously increasing the win rate and reducing operating costs.

Scientific novelty is reduced to the assumption that an integral two-loop model preventive graph detection + post-dispute generation of an evidentiary base, which links the taxonomy of payment reason codes and the standards of compelling evidence (including CE3.0) with specific classes of AI methods (GNN for network dependencies, RAG for normatively correct generation of materials), thereby transferring dispute management

from a heuristic craft into a reproducible computational procedure.

Practical **significance** consists in the fact that the results define an applied design for implementation of AI orchestration of disputes (ranking by probability of success, automatic assembly of a digital trace, RAG response templates, time-bar control and XAI audit), allowing reduction of the processing cycle from hours to seconds, increasing the share of won cases in digital goods to ~72%, and decreasing losses from false claims and administrative burden.

Materials and Methods

The study is based on the principles of a systematic literature review and the case study method. This approach makes it possible to combine the theoretical developments of the academic community with practical industry data.

The source selection process included the following stages:

– Identification: Search for publications in the databases Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and arXiv for the period 2020–2025. The following search queries were used: AI in dispute resolution, machine learning fraud detection, automated chargeback management, GNN in finance.

– Screening: From the initially selected publications, works that lacked an empirical base or duplicated results were excluded. Priority was given to articles published in Q1/Q2 journals and in proceedings of A*-level conferences (NeurIPS, AAAI, WWW).

– Inclusion: The final pool included 20 key sources covering technical aspects (neural network architectures), legal aspects (payment system regulations), and statistical data.

– Analysis of industry reports: To verify academic hypotheses, reports of leading analytical agencies and payment associations (Visa, Merchant Risk Council, Gartner, Chargebacks911) for 2024–2025 were analyzed.

The literature analysis revealed substantial growth of interest in the topic. Until 2022 most works focused on classical ML methods (Random Forest, SVM), then since 2023 a shift toward Deep Learning and Generative AI has been observed.

IEEE and ACM: Provided the technical foundation on GNN architectures (detectGNN) and NLP methods for legal analysis.

Scopus/WoS: Provided an overview of systemic approaches to automation in e-commerce and fashion retail.

Industry reports (Visa, MRC): Became a source of critically important statistics on friendly fraud trends and operational metrics.

Special attention is given to works describing the application of hybrid ensemble models and methods of interpretable AI (XAI), since transparency of decisions is a mandatory requirement in the financial domain.

Results and Discussion

The development of effective dispute automation requires a precise understanding of the internal logic of the chargeback as a multilateral and iterative process in which the cardholder, the issuing bank, the payment network, the acquiring bank, and the merchant interact sequentially. Such a circuit is fundamentally not reducible to a direct chain of actions: the exchange of messages and documents occurs through regulated stages, and the decision is formed under conditions of distributed responsibility and information asymmetry among participants. The regulatory framework and operational rules of 2024 fix strict time constraints at key stages: initiation of a dispute is permitted within 120 days from the transaction date, and in certain cases the maximum period is extended to 540 days; after receipt of the notification the merchant obtains a limited window to prepare a response, most often 10–30 days, while some pre-arbitration procedures require a response within 10 days; completion of the full cycle at the level of arbitration procedures may take approximately 30–75 days [10]. In such a configuration, time becomes a critical resource: delays in manual compilation of the evidentiary base increase the probability of missing procedural deadlines, which leads to an automatic outcome unfavorable to the merchant regardless of the actual legitimacy of the transaction [11, 12].

The classification of disputes in the payment infrastructure relies on a standardized taxonomy of reason codes that determines not only the formal qualification of the dispute ground but also the requirements for the composition and form of supporting materials. Each code corresponds to a specific list of

compelling evidence, where the significance lies not so much in the volume and heterogeneity of documents as in their procedural relevance to the specific ground of the dispute and compliance with the formal expectations of the network. Thus, Reason Codes in effect function as a regulatory template for the evidentiary strategy: deviation from the required structure, incompleteness, or irrelevance of the submitted data reduces the probability of successful representation even in the presence of factual arguments.

Within Table 1, the key Visa and Mastercard reason codes and the specificity of evidence will be presented.

Table 1. Key Visa and Mastercard reason codes and the specificity of evidence (compiled by the author based on [10, 11, 12]).

Dispute Category	Visa Code	Mastercard Code	Essence of the Claim	Required Evidence (Evidence)	Complexity of Automation
Fraud (Fraud)	10.4 (Card-Absent Environment)	4837 (No Cardholder Authorization)	The cardholder asserts that they did not make the purchase	IP address, Device ID, AVS/CVV match, history of previous purchases, linkage to a social media account	High (requires analysis of relationship graphs)
Consumer Dispute (Goods)	13.3 (Not as Described)	4853 (Defective/Not as Described)	The goods are defective, counterfeit, or do not match the description	Product description on the website, photos of the packaging, correspondence with the customer, return policy	Medium (NLP-based semantic analysis)
Service / Non-Receipt	13.1 (Merchandise Not Received)	4855 (Goods/Services Not Provided)	The goods were not delivered	Tracking number (POD), recipient signature, access logs to the digital service	Low (structured data)
Process Error	12.6 (Duplicate Processing)	4834 (Point of Interaction Error)	Duplicate charge	Transaction logs, confirmation of a single charge	Low

Most labor-intensive for evidentiary support is considered to be the group of disputes Visa 10.4 / MC 4837, because it is precisely in this class that a substantial share of friendly fraud is concentrated: the transaction initially passes with authorization from the cardholder; however, subsequently the fact of sanctioning the operation is denied. In response to the scaling of such scenarios, Visa implemented the Compelling Evidence 3.0 (CE3.0) initiative, which provides the merchant with a mechanism to preclude the dispute in the presence of a verifiable historical relationship between the buyer and the point of sale. The key requirement is the demonstration of at least two prior undisputed transactions associated with the same device and/or IP

address, which transfers the dispute from the plane of declarative assertions to the plane of verifiable behavioral correlations [16]. This regulatory framework forms a pronounced potential for automation: an algorithmic module must, within minimal timeframes, extract relevant elements of the user history, match them against the CE3.0 criteria, and generate an evidentiary package in a strictly specified template [14, 16].

The economic logic of the dispute process renders a strategy of ignoring challenges unsustainable, because the final financial losses are determined not only by the direct debit, but also by the structure of the industry probability of success, the commission burden, and administrative expenses. Win rate indicators under manual management vary substantially across segments. In the sphere of digital goods, the share of won disputes can reach 72,56%, which is usually explained by the high provability of the fact of service consumption or access to content due to digital traces (login logs, service usage events), whereas for physical deliveries the risk of alternative interpretations remains, including theft after delivery (porch piracy) [5]. In the categories of apparel and mass retail, the win rate is noticeably lower and is estimated at approximately 35,81%, because the subjectivity of claims associated with expectations regarding product characteristics often increases the likelihood of a decision in favor of the cardholder [5]. At the level of an aggregated benchmark, the average share of won disputes across industries is about 45%; however, the net recovery rate decreases to 18% due to commissions and administrative costs accompanying each stage of the process [4]. The implementation of automated tools modifies the indicated economics: in industrial practice, a reduction in the level of false positives and an increase in the effectiveness of contestation are noted when advanced solutions are used, which is associated with stricter evidence relevance and a reduction of time losses for preparing responses [16, 18].

To cover the full scope of dispute management, a two-level intelligent architecture is required that combines preventive and reactive mechanisms. At the preventive level, the focus shifts to identifying probable friendly fraud scenarios before the completion of the operation or at the moment of payment processing, whereas at the reactive level the key task becomes the automated preparation and submission of materials for chargebacks after receipt of a dispute notification [19, 20].

In the preventive contour, the limitations of traditional machine learning models manifest in that methods such as Random Forest or Gradient Boosting often interpret a transaction as a relatively autonomous observation, whereas fraudulent activity in real ecosystems is network in nature and is reproduced through repeated use of devices, IP addresses, and behavioral patterns within chains of synthetic identities. For representing and analyzing such interrelations, graph neural networks (GNN) are applicable, including architectures of the detectGNN class [13]. In this formulation, transactional data are described by a heterogeneous graph G=(V,E), where the set of nodes VV includes users, cards, merchants, devices, and IP addresses, and the edges EE encode the fact of a transaction, matches of environment attributes (for example, a device), and other forms of connectivity. The computational logic of a GNN is based on aggregating information from neighboring nodes with subsequent updating of vector representations, which makes it possible to detect hidden dependencies: for example, a transaction of an account without explicit risk indicators can be qualified as suspicious when a connection is discovered with a device previously associated with confirmed fraud, even if other indicators remain within acceptable limits [21, 22].

A substantial extension of modern solutions, including detect GNN, is temporal encoding, which is necessary to account for event dynamics and to recognize short-term bursts of activity typical for card testing or other explosive scenarios [13]. In comparative tests, the superiority of GNN over classical approaches is reported in terms of the metrics accuracy (97,5% versus 93,2%) and recall (94,2% versus 89,7%); at the same time, high recall has priority value due to the asymmetry of losses: missing a fraudulent operation usually costs more than additional verification of a legitimate transaction [13].

At the post-dispute stage, the task shifts from classification to the formation of procedurally correct text and attachments: it is required to prepare a reasoned response, accompanying it with relevant evidence corresponding to a specific reason code. In this contour, large language models (LLM) play a central role, applied in conjunction with Retrieval-Augmented Generation (RAG), because the direct use of a generative model without an external grounding increases the risk of hallucinations, including incorrect references to non-existent provisions of regulations. The RAG architecture reduces this risk by splitting into two stages: first, retrieval is performed—extracting relevant materials from an external knowledge base (current Visa/Mastercard rules, event logs for a specific

transaction, correspondence with support), then generation—generating text based on the found context [23, 24]. Typical logic can be illustrated by the scenario of a dispute under Visa 13.3 (Not as Described): at the retrieval stage, the product listing in the state at the time of purchase, delivery information (for example, the tracking number), and the client's dialogues with support are extracted; at the processing stage, an NLP module is able to analyze the sentiment and substantive markers of the correspondence, identifying internal contradictions in the client's statements [26]. After that, the LLM forms a structured letter that correlates the stated dispute argument with documented facts, and links the conclusions to relevant norms of the payment regulations, ensuring legally robust argumentation without going beyond the boundaries of the confirmed context [24].

Table 2 contains the results of a comparison of machine learning methods for dispute detection and resolution.

Table 2. Comparative characteristics of machine learning methods for dispute detection and resolution (compiled by the author based on [13-18]).

Method	Application	Advantages	Limitations
Random Forest / XGBoost	Transaction classification (Pre-transaction)	High interpretability, training speed, performance with tabular data.	Poor capture of complex network dependencies and sequential patterns.
Graph Neural Networks (GNN)	Detection of network fraud, linked accounts	Detection of hidden relationships, high accuracy on imbalanced data.	High computational complexity, complexity of deployment (Graph Infrastructure).
NLP (Transformer-based)	Analysis of complaint texts, support chats, response generation	Context understanding, automation of routine tasks, multilingual capability.	Risk of hallucinations (without RAG), sensitivity to data quality.
Hybrid Ensemble Models	Combination of methods (Voting Classifiers)	Robustness, reduction of error variance, capability for data balancing (IHT-LR).	Complexity of maintenance and hyperparameter tuning.

The global technology company Social Discovery Group (SDG) manages a portfolio comprising more than 70 brands in the Social Discovery segment, encompassing dating services, social games, and entertainment digital products. The commercial model of SDG is predominantly based on the sale of digital goods and subscriptions (SaaS/Subscription model), which forms a specific risk profile: in such ecosystems, friendly fraud manifests particularly frequently, because the consumer can activate a subscription, actually use the functionality, and then initiate a dispute, motivating it by a forgotten cancellation or by denying the fact of enrollment. For the subscription context, such arguments create a typical first-party misuse scenario, in which the transaction is initially authorized but retrospectively interpreted as unauthorized or unwanted [25, 27].

A pressure factor is the high share of microtransactions: with an average ticket of about 10–20 dollars, manual dispute handling often proves economically negative, because fixed administrative costs for collecting and preparing evidence exceed the potential effect of a

successful contestation. An additional methodological complexity is associated with the absence of physical delivery: traditional evidence typical for goods retail (recipient signature, courier service confirmation) is not applicable in the digital segment, which requires reliance on other sources of confirmation—access events, login logs, device parameters, and service-usage telemetry. Finally, the operational scale of SDG—activity in 150 countries—complicates dispute management due to the heterogeneity of local payment practices, differences in regulatory expectations, and the variability of dispute processing procedures on the side of acquirers and issuers.

A key element is the application of predictive analytics for ranking incoming disputes by the probability of a successful outcome. This formulation makes it possible to automatically exclude cases with low recovery potential (for example, in the presence of a confirmed processing error) and to concentrate computational and operational resources on categories typical for friendly fraud. As a result, for digital goods an average win rate at the level of 72–75% is achievable, which exceeds the averaged industry benchmarks of about 45–50% for aggregate categories [5]. Of decisive importance here is not the volume of argumentation, but the speed and completeness of aggregating the digital footprint as Compelling Evidence: systematized data on login IP addresses, session duration, the sequence of actions in the application, and other telemetry markers provide a reproducible evidentiary base that is difficult to refute when authorization is denied.

A second measurable effect of automation manifests in the reduction of reaction time. Automated API interaction with payment gateways in combination with template-based generation of responses based on RAG (RAG templates) transfers representment preparation from a manual assembly mode (30–60 minutes per case) to a mode of practically instantaneous compilation—up to several seconds per dispute [15]. Such compression of the response cycle not only reduces the unit processing cost, but also structurally eliminates the risk of procedural loss due to missing deadlines (time-bar), which for high-throughput subscription services has a direct financial expression.

A third result is a reduction in the share of erroneous triggers in the preventive contour. The combination of graph approaches (GNN) and behavioral analysis increases the accuracy of distinguishing legitimate users from abusive subjects, which reduces the probability of customer insult—blocks or restrictions imposed on bona fide clients. For the subscription model, this parameter is critical, because errors toward excessive strictness undermine LTV (Lifetime Value) through increased churn and degradation of trust in the service, whereas more correct segmentation makes it possible to retain valuable users while simultaneously strengthening protection against first-party misuse [16].

The transition from manual procedures to automated dispute management, as a rule, is accompanied by a pronounced increase in economic return and forms a high return on investment (ROI) due to the reduction of transaction costs and an increase in the share of retained revenue. In the manual contour, the total cost of processing a single dispute is composed of mandatory fees and internal costs, and also includes a component of lost benefit arising from the reallocation of limited specialist resources to low-productivity operations.

Under automation, the main variable component reflecting the dependence of costs on staff time (Time×Rate) functionally tends toward zero, because evidence collection, criteria verification, and representment formation are transferred to a mode of machine orchestration. Industry estimates indicate that specialized automated platforms are capable of delivering ROI up to 5300% by recovering revenue that was previously recorded as unconditional losses [28]. For an organization of the scale of SDG, such dynamics are interpreted as an annual preservation of millions of dollars, because even a moderate increase in win rate and a reduction in time-bar losses under a high flow of microtransactions leads to a significant cumulative effect.

The systemic deployment of AI in dispute management simultaneously generates a complex of new challenges associated with changes in the threat profile and with institutional requirements for the transparency of algorithmic decisions. A transition is observed to a mode of adversarial AI, in which the use of intelligent defensive means stimulates symmetric technological escalation by malicious actors. Generative models begin to be used to construct synthetic identities capable of passing basic KYC procedures [1], to create deepfakes for the purpose of bypassing biometric authentication, and an increase in such incidents by 28% in 2024 is recorded [1], as well as to automate the preparation of appeals to issuing banks, where texts are formed so as to look as plausible and emotionally saturated as possible, which increases the effectiveness of social engineering when initiating disputes [29]. This evolution means that defensive

contours must function as adaptive systems: regular retraining, monitoring of concept drift, and the introduction of generated-content detectors capable of separating organic user messages from synthetically formed complaints and confirmations are required.

In parallel, the black-box problem and the need for explainability (XAI) become more acute. Financial regulation, including the regulatory regimes of the EU and the USA, embeds the requirement of interpretability into the practice of making decisions affecting access to payment services and the processing of disputed operations. If an algorithmic system declines a transaction or forms a legally significant response, it is required to be able to reconstruct the causal logic of the decision taken and to identify the features that exerted decisive influence. The use of deep learning class models increases the risks of opacity; therefore, the implementation of Explainable AI methods such as SHAP (SHapley Additive exPlanations) and LIME becomes critically important, enabling assessment of the contribution of individual factors and visualization of the structure of influences in the final classification or ranking [17].

An additional level of uncertainty is set by the regulatory dynamics of the payment infrastructure. The evolution of network policies and changes in procedural standards, including the implementation of Visa Compelling Evidence 3.0 in 2023/2024, require high adaptability from merchant solutions, because statically specified rules (hard-coded rules) rapidly lose relevance. In this context, the RAG architecture considered within the research framework has an applied advantage: adjustment to new requirements can be implemented through updating the knowledge base and reindexing sources without resorting to retraining the core model, which reduces the cost of changes and decreases the risk of quality degradation under regulatory updates.

Conclusion

The conducted study, synthesizing the results of academic data analysis and applied industry developments, makes it possible to formulate a clear conclusion: the automation of dispute management based on artificial intelligence methods in 2025 acquires the status of a strategically inevitable direction of development for e-commerce. The sustained complication of threats and the growth of operational workload take dispute management beyond the boundaries of an auxiliary function and transform it into a critical revenue-protection contour, where manual practices cease to correspond to the pace and scale of digital markets.

Friendly fraud has become established in the role of the dominant source of losses, which requires a revision of the counteraction logic and a departure from a predominantly reactive model oriented toward point blockings and ex post reactions, in favor of an analytical paradigm based on reconstructing behavioral patterns and verifying historical relationships. Under first-party misuse conditions, the key resource of provability is formed not by individual transaction attributes, but by the context of interactions: device repetitiveness, stable indicators of digital service consumption, and the coherence of the event trail, which makes behavioral analytics the methodological core of modern protection.

The technological advantage of graph approaches manifests in the ability to model fraudulent activity as a network phenomenon rather than as a set of independent operations. Architectures of the detect GNN class demonstrate a qualitative increase in detection accuracy, including an increase in Accuracy by 4,3% relative to traditional ML models, through identifying latent dependencies and transitive links between subjects of schemes and infrastructural artifacts (devices, IP, behavioral chains). Such an improvement in quality has applied value not only as a metric, but also as a means of reducing asymmetric losses that arise when fraudulent operations are missed.

In the post-dispute contour, the automation of representment acquires critical importance, because it is precisely here that the decision on refunding funds is formed and the dispute outcome is determined. The use of NLP and generative models in the RAG architecture transfers the preparation of legally significant responses from a craft process to a scalable software procedure: retrieval of relevant documents and event data, formation of an evidentiary package, and generation of a structured rebuttal become reproducible and quality-controlled. A practical indicator of effectiveness is the growth of the win rate to 72% in the digital goods segment, where the digital footprint provides high evidentiary density and reduces the space for bad-faith interpretations.

The economic effect of implementing hybrid AI systems is expressed in the transformation of the role of risk management. Using the example of scenarios applicable to the Social Discovery Group case, the integration of predictive models, graph analytics, and automated

evidence formation changes the balance of costs and benefits: a function traditionally perceived as an inevitable operational expense acquires the properties of a revenue-preservation center due to accelerated processing, reduced losses from time-bar, and increased effectiveness of contestation in the most expensive categories of first-party misuse.

Further competitiveness in this area will be determined by the ability to build shared intelligence ecosystems, implying the exchange of signals and fraud patterns between market participants while preserving privacy and complying with regulatory constraints. As a promising direction, approaches relying on Federated Learning are distinguished, enabling training of models on distributed data without centralization of sensitive information and thereby reducing barriers to inter-organizational cooperation. The escalation of the confrontation between the fraud community and defensive contours retains the character of a technological race in which artificial intelligence becomes the determining instrument of both attack and defense.

References

1. 70+ eCommerce Fraud Statistics [2025]: Trends, Data, & Facts | Cropink. Retrieved from: https://cropink.com/ecommerce-fraud-statistics (date accessed: October 12, 2025).

2. Ecommerce Fraud Statistics: Key Trends & Insights for 2025 | ClickPost. Retrieved from: https://www.clickpost.ai/blog/ecommerce-fraud-statistics (date accessed: October 13, 2025).

3. 23+ eCommerce Fraud Statistics (2024) | Exploding Topics. Retrieved from: https://explodingtopics.com/blog/ecommerce-fraud-stats (date accessed: October 14, 2025).

4. Chargeback Stats: All the Key Dispute Data Points for 2025 | Chargebacks911. Retrieved from: https://chargebacks911.com/chargeback-stats/ (date accessed: October 15, 2025).

5. 23+ Chargeback Statistics Every Merchant Should Know for 2025 | Chargeback.io. Retrieved from: https://www.chargeback.io/blog/chargeback-statistics (date accessed: October 16, 2025).

6. The Ultimate Chargeback Statistics 2025: Trends, Costs, and Solutions | Chargeflow. Retrieved from: https://www.chargeflow.io/blog/chargeback-statistics-trends-costs-solutions (date accessed: October 17, 2025).

7. 2025 Global eCommerce Payments and Fraud Report | Merchant Risk Council. Retrieved from: https://merchantriskcouncil.org/learning/mrc-exclusive-reports/global-payments-and-fraud-report (date accessed: October 18, 2025).

8. Chargeback Dispute Statistics for Merchants | Clearly Payments. Retrieved from: https://www.clearlypayments.com/blog/chargeback-dispute-statistics-for-merchants/ (date accessed: October 19, 2025).

9. Automated vs. Manual Chargeback Management | Disputifier. Retrieved from: https://www.disputifier.com/post/automated-vs-manual-chargeback-management (date accessed: October 20, 2025).

10. How Long Do Chargebacks Take? Real Timelines by Network and What You Can Control | Disputifier. Retrieved from: https://www.disputifier.com/post/how-long-do-chargebacks-take-real-timelines-by-network (date accessed: October 21, 2025).

11. How Long Does a Chargeback Take? | Chargeback.io. Retrieved from: https://www.chargeback.io/blog/how-long-does-a-chargeback-take (date accessed: October 22, 2025).

12. Goti, A., Querejeta-Lomas, L., Almeida, A., Gaviria de la Puerta, J., & López-de-Ipiña, D. (2023). Artificial intelligence in business-to-customer fashion retail: A literature review. Mathematics, 11(13), 2943. https://doi.org/10.3390/math11132943

13. Sultana, I., Maheen, S. M., Kshetri, N., & Zim, M. N. F. (2025). detectGNN: Harnessing graph neural networks for enhanced fraud detection in credit card transactions (arXiv preprint arXiv:2503.22681). https://doi.org/10.48550/arXiv.2503.22681

14. Ariai, F., Mackenzie, J., & Demartini, G. (2025). Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges. ACM Computing Surveys, 58(6), 1–37. https://doi.org/10.1145/3777009

15. Automated Chargeback Management: Increasing Win Rates with Machine Learning | ResearchGate. Retrieved from: https://www.researchgate.net/publication/387491165_Automated_Chargeback_Management_Increasing_Win_Rates_with_Machine_Learning (date accessed: October 23, 2025).

16. Saha, B., Rani, N., & Shukla, S. K. (2025). Generative AI in financial institution: A global survey of opportunities, threats, and regulation (arXiv preprint arXiv:2504.21574). https://doi.org/10.48550/arXiv.2504.21574

17. Hosseini Chagahi, M., Delfan, N., Mohammadi Dashtaki, S., Moshiri, B., & Piran, M. J. (2025). Explainable AI for fraud detection: An attention-based ensemble of CNNs, GNNs, and a confidence-driven gating mechanism (arXiv preprint arXiv:2410.09069). https://doi.org/10.48550/arXiv.2410.09069

18. Talukder, M. A., Hossen, R., Uddin, M. A., Uddin, M. N., & Acharjee, U. K. (2024). Securing transactions: A hybrid dependable ensemble machine learning model using IHT-LR and grid search. Cybersecurity, 7, Article 32. https://doi.org/10.1186/s42400-024-00221-z

19. Dispute Management Guidelines for Visa Merchants (June 2024) (PDF) | Visa. Retrieved from: https://usa.visa.com/dam/VCOM/global/support-legal/documents/merchants-dispute-management-guidelines.pdf (date accessed: October 25, 2025).

20. Chargeback Reason Codes: The Ultimate Guide for 2025 | Chargebacks911. Retrieved from: https://chargebacks911.com/chargeback-reason-codes/ (date accessed: November 1, 2025).

21. Mastercard Chargeback Reason Codes (2024 Guide for Merchants) | Chargeflow. Retrieved from: https://www.chargeflow.io/blog/mastercard-chargeback-reason-codes (date accessed: November 2, 2025).

22. Chargeback Trends by Industry: 2024 Data and Analysis | Chargeflow. Retrieved from: https://www.chargeflow.io/blog/chargeback-trends-by-industry-2024-data-and-analysis (date accessed: November 3, 2025).

23. Siam, A. M., Bhowmik, P., & Uddin, M. P. (2025). Hybrid feature selection framework for enhanced credit card fraud detection using machine learning models. PLOS ONE, 20(7), e0326975. https://doi.org/10.1371/journal.pone.0326975

24. Antal, M., & Buza, K. (2025). Evaluating open-source LLMs in RAG systems: A benchmark on diploma theses abstracts using Ragas. Acta Universitatis Sapientiae, Informatica, 17, Article 5. https://doi.org/10.1007/s44427-025-00006-3

25. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks (arXiv preprint arXiv:2005.11401). https://doi.org/10.48550/arXiv.2005.11401

26. NLP-Based Automation in Customer Support and Case Management | inLIBRARY. Retrieved from: https://inlibrary.uz/index.php/ijns/article/download/108445/110067 (date accessed: November 4, 2025).

27. Hidden Costs of Manual Chargeback Management in 2025 | Zenskar. Retrieved from: https://www.zenskar.com/blog/chargeback-management (date accessed: November 5, 2025).

28. Answered: What Is The ROI of Chargeflow? | Chargeflow. Retrieved from: https://www.chargeflow.io/blog/whats-the-roi-of-chargeflow (date accessed: December 10, 2025).

29. Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. Patterns, 5(5), 100988. https://doi.org/10.1016/j.patter.2024.100988