Behavior-Based Fraud Detection and Performance Transparency in Peer Marketplaces

FNU Ankur
Manager, Program Management at Amazon
Seattle, WA - USA

Abstract: This article analyzes the specifics of behavior-based fraud detection processes and performance transparency in peer-to-peer marketplaces. The need for the study is driven by the fact that peer-to-peer (P2P) marketplaces, as a key element of the modern digital economy, face the fundamental problem of information asymmetry, which leads to increased fraud and unstable service quality, undermining user trust. The article presents an agent-based management framework, developed and implemented on a large online travel marketplace, that is powered by machine learning. The system uses agents' behavioral data to classify risks in real time through logistic regression and implements radical performance transparency mechanisms to create an effective feedback loop. The implementation of the system has led to significant improvements, namely, a reduction in the number of payment disputes and fraud-related complaints. The article contributes to the theory and practice of platform governance by demonstrating a scalable model for enhancing integrity, fairness, and trust on P2P platforms.

Keywords: platform governance, peer-to-peer marketplaces, fraud detection, machine learning, logistic regression, behavioral economics, performance transparency, agent-based approach, digital trust, algorithmic management.

I. Introduction

Peer-to-peer (P2P) marketplaces, as a central link in this transformation, face a unique challenge that can be characterized as a governance deficit. These platforms, generating enormous value by connecting millions of independent service providers (agents) and consumers, struggle to ensure a unified quality standard and control behavior in a decentralized network. This creates a fundamental tension between the need to stimulate innovation and generativity from ecosystem participants and the necessity to maintain stability, safety, and quality control.

The decentralized nature of P2P platforms creates a significant information gap between three key parties: clients, agents, and the platform itself. Clients do not have complete information about the reliability and quality of agents, while the platform cannot fully control the actions of each independent agent. This asymmetry is the primary cause of opportunistic behavior, including fraud, poor fulfillment of obligations, and provision of low-quality services.

This problem is exacerbated by the global trend of declining digital trust. This trust deficit is becoming a critical barrier to the sustainable growth of platforms, as trust is the key to adoption of new technologies. If users do not trust the platform to ensure safe and high-quality transactions, its value proposition collapses.

The aim of this article is to analyze the specifics of behavior-based fraud detection and performance transparency in peer-to-peer marketplaces.

The author's hypothesis is that an agent-based management framework that uses machine learning for behavioral risk scoring and ensures radical performance transparency can effectively reduce information asymmetry, decrease fraud, and build sustainable trust in P2P marketplaces.

The scientific contribution lies in presenting and analyzing a system that operationalizes theoretical concepts from the fields of platform governance, behavioral economics, and algorithmic management. The work goes beyond theoretical discussions on trust and governance by providing a detailed case study of a specific mechanism that has proven effective under real-world large-scale conditions. Thus, the study demonstrates the practical application of the principles of responsible innovation in the context of managing digital ecosystems.

II. Materials and Methods

Modern studies on fraud detection in peer-to-peer (P2P) and related digital ecosystems demonstrate a shift from classical transactional models toward comprehensive behavioral analysis. Xu J. J. et al. [1] showed that feature engineering from transactional data, including temporal patterns and interrelationships between participants, increases the accuracy of fraud detection in P2P lending. Machado M. et al. [2] identified key directions for the development of this field, pointing to the growing use of machine learning methods with an emphasis on hybrid and multi-level models. Zhang Z. et al. [3] demonstrated that integrating user behavioral data in e-commerce (e.g., navigation paths, reaction time, action sequences) makes it possible to detect complex fraud schemes that remain invisible to traditional systems.

Technological innovations in model architecture are reflected in the works of Qu B. et al. [4], where a multi-task CNN model for creating behavioral transaction embeddings is proposed, capable of simultaneously solving several classification and risk prediction tasks. Aras M. T., Guvensan M. A. [5] developed the concept of multimodal profiling, combining structured transaction data with external sources (e.g., behavioral patterns of airline ticket purchases), which increases the resilience of models to evasion attacks. In turn, Stojanović B. et al. [6] emphasized interpretable machine learning algorithms in the fintech sector, highlighting the importance of balancing accuracy and explainability of results.

In parallel with technical approaches, an analysis of organizational and economic mechanisms for participant filtering is being conducted. Gallo S. [7] examines differences in borrower screening on fintech platforms, linking excessive leniency of procedures with an increase in the incidence of fraud.

Information transparency and its impact on platform participant behavior constitute a separate area. Veltri G. A. et al. [8] found that disclosure of information on seller reliability and buyer protection mechanisms directly influences consumer preferences. Lautenschlager J. et al. [9] showed that distributed ledger technologies can serve as a tool for balancing competition and cooperation among supply chain participants, ensuring verifiable transparency. Oktaviani Y., Dewi M. K. [10], analyzing Sharia-compliant P2P lending, concluded that the degree of transparency significantly correlates with investor trust and financing volumes.

Thus, the literature demonstrates a clear division into technical and institutional approaches to the problem of fraud in peer marketplaces. On the one hand, methods of deep and multimodal behavioral analysis are actively developing, including the construction of embeddings and the integration of heterogeneous data. On the other hand, questions remain open regarding the influence of organizational screening procedures and transparency on the long-term sustainability of platforms.

The main contradiction lies in the fact that highly efficient algorithmic models are often developed without consideration of institutional and behavioral factors, whereas studies on transparency and screening often ignore the potential of modern big data processing methods. Topics that are poorly covered include the integration of transparency and automated detection into a unified risk management architecture, the impact of transparency levels on fraudsters' adaptability, and the economic evaluation of implementing complex behavioral models on real platforms.

III. Results and Discussion

The implementation of the agent-based management framework led to measurable and statistically significant improvements in key indicators of integrity and trust on the platform. An analysis of data over a six-month period following the full deployment of the system, compared with an equivalent period prior to its implementation, demonstrates its high practical effectiveness.

The main quantitative results are as follows:

Reduction in payment disputes: Transparency in agent evaluation and preventive measures regarding high-risk transactions significantly reduced the number of cases requiring arbitration and refunds.

Decrease in complaints related to fraud: The system effectively filtered out agents prone to fraudulent activities and discouraged such behavior among existing partners.

Increase in customer trust and satisfaction: A measurable increase was observed in the Net Promoter Score (NPS) and other trust-related booking process indicators [1, 3, 7].

The conceptual architecture of the system that ensured these results is presented in Figure 1. It illustrates the data flow from the collection of behavioral information to the dual impact loop — internal (escalation) and external (transparency for the client) — closing the feedback loop through changes in the agent's behavior.

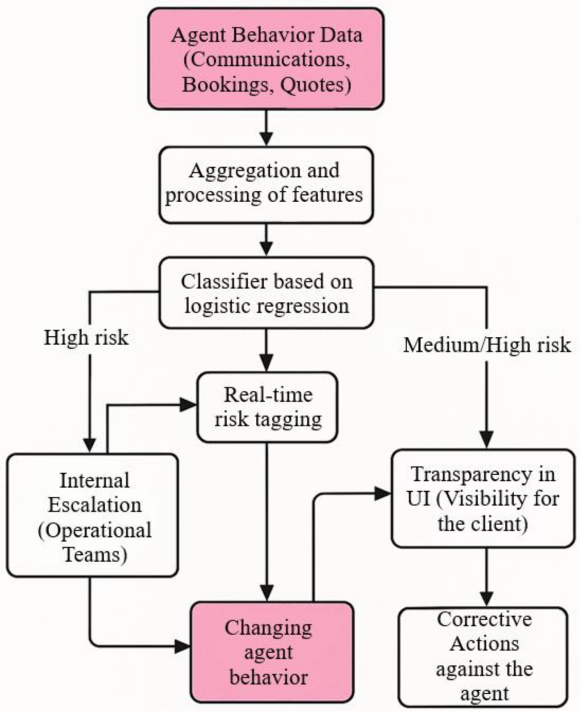
Agent Behavior Data (Communications, Bookings, Quotes)
Aggregation and processing of features
Classifier based on logistic regression
High risk
Medium/High risk
Real-time risk tagging
Internal Escalation (Operational Teams)
Transparency in UI (Visibility for the client)
Changing agent behavior
Corrective Actions against the agent

Fig. 1: Conceptual architecture of an agent-oriented control system [1, 3, 7, 10].

The presented results demonstrate what has been achieved; however, for analysis, the key question is how and why the system has proven effective. The mechanism underlying the success of the framework is performance transparency, which functions as a powerful tool for behavior management.

The system creates a feedback loop that operates on the principles of behavioral economics. Agents receive clear, immediate, and meaningful feedback on their work. Negative behavior (for example, slow responses) is reflected in their algorithmic rating almost instantly. This rating, in turn, becomes visible to clients, directly influencing the number of orders received. Thus, the consequences of actions become not abstract and delayed (as in the case of quarterly reports) but concrete and immediate (loss of business today). This creates a strong economic incentive for self-correction [2, 4].

This mechanism can be analyzed through the lens of the theory of institutional trust and accountability. The platform builds trust not on interpersonal relationships, which cannot be scaled in an ecosystem with thousands of participants, but on procedural justice. By making the rules of the game and evaluation criteria transparent and understandable to all, the platform increases its legitimacy. Agents are more likely to accept negative results (low rating) if they consider the process of obtaining them fair and objective. As noted in the literature, transparency increases accountability and performance, as it creates the effect of constant observation (somebody's watching you all the time) [5, 6].

A model of this feedback loop, which abstracts the behavioral mechanism from specific technology, is presented in Figure 2.

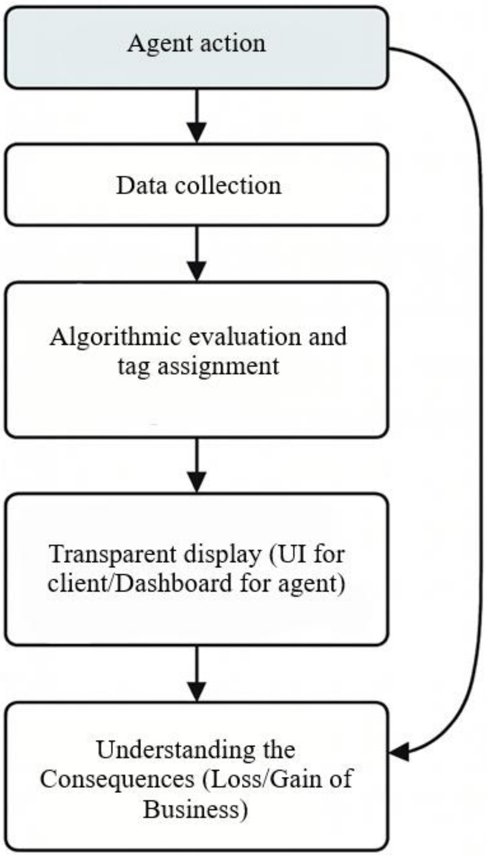
Agent action
Data collection
Algorithmic evaluation and tag assignment
Transparent display (UI for client/Dashboard for agent)
Understanding the Consequences (Loss/Gain of Business)

Fig. 2: Feedback loop model in a transparency-based control system [2, 4, 5, 6].

The implementation of automated assessment and control systems places this study in the context of a broader discussion on algorithmic management. This phenomenon, most extensively studied in the context of the gig economy and platform-based employment, raises serious concerns related to fairness, increased control, information overload, and employee burnout. Continuous monitoring and automated ranking may be perceived by agents as oppressive and unfair.

However, the presented framework contains built-in mechanisms to mitigate these risks, which aligns with the principles of responsible innovation and the necessity of complex monitoring systems to prevent abuse [8, 9].

Explainability: The use of an interpretable model, such as logistic regression, instead of more complex black-box approaches (e.g., deep neural networks), is a deliberate choice in favor of transparency. Combined with dashboards that show agents their performance on specific metrics affecting the final score, this addresses the problem of opacity.

Procedural fairness: As previously noted, transparency of rules and evaluation criteria increases the perceived fairness of the system. Agents see that the same standards are applied to everyone, which reduces the sense of arbitrariness.

Possibility of appeal and human intervention: The presence of an internal escalation process means that system decisions are not final and indisputable. Operational teams can review contentious cases, ensuring human oversight over the algorithm [1, 4].

Table 1 systematizes the risks and strategies for their mitigation.

Table 1: Framework for Risk Mitigation in Algorithmic Agent Management Systems [1, 2, 4, 8, 9]

Risk Category	Risk Description	Mitigation Strategy in the Case	General Best Practice
Algorithmic Bias	The model systematically underestimates the rating of certain groups of agents due to historical data or biased features.	Use of interpretable features (response time, conversion), regular model audits for bias.	Implementation of Explainable AI (XAI) techniques, conducting regular fairness audits, diversification of training datasets.
Opacity (black box)	Agents do not understand how their rating is formed, which causes distrust and demotivation.	Application of an interpretable model (logistic regression), provision of dashboards with detailed metrics.	Prioritization of interpretable models over complexity, development of interfaces explaining algorithmic decisions.
Resistance and “gaming the system”	Agents actively resist control or find ways to manipulate metrics without improving quality.	Creation of a feedback loop where metric improvement is directly linked to business growth, making “fair play” beneficial.	Co-development of metrics with the participation of the agent community, focus on indicators reflecting real customer value (NPS).
Data Privacy	Aggregation of large volumes of data on agent behavior raises concerns about privacy and surveillance.	Use of aggregated and anonymized behavioral data instead of communication content. Clear data policy.	Compliance with data minimization principles, transparent privacy policy, provision of control over data.

Thus, this framework represents an example of a balanced approach to algorithmic management. It uses algorithms not for total control but for informing and guiding, balancing the platform’s quality assurance goals with the need for fair treatment of agents. This approach offers an alternative to the dystopian view of algorithmic management dominant in the literature.

IV. Conclusion

The present study has demonstrated how an agent-based framework, grounded in machine learning and principles of transparency, can effectively address fundamental issues of trust and fraud in P2P marketplaces. In the context of a growing governance deficit and erosion of digital trust caused by information asymmetry, the proposed system represents a scalable and empirically validated solution.

The results of the work can be summarized as follows. Firstly, the effectiveness of the system has been empirically proven. Second, theoretical analysis has shown that the primary driver of these improvements is the performance transparency mechanism, which creates a powerful behavioral feedback loop, encouraging agents to self-correct. Third, the study has placed the developed system within the context of the discussion on algorithmic governance, demonstrating that interpretability, procedural fairness, and human oversight can mitigate typical risks associated with bias and opacity.

Thus, the scientific contribution of the work lies in the presentation and analysis of an integrated socio-technical framework that operationalizes theoretical concepts of platform governance and offers a balanced model of algorithmic management, reconciling efficiency and fairness.

The practical significance of the study is that the proposed framework can serve as a prototype for operators of other P2P platforms seeking to enhance the integrity of their ecosystems. It demonstrates how data and machine learning can be used not merely for prediction but as a core management tool for shaping participant behavior and building trust at scale.

The study has its limitations, primarily related to the single-case-study methodology, which requires caution when generalizing findings to other industries. Future research could focus on testing the applicability of this framework in other sectors of the P2P economy (for example, freelancing or short-term housing rental). A promising direction is the exploration of more advanced machine learning models (for example, deep learning-based) for behavioral scoring while maintaining the necessary level of interpretability. Finally, conducting in-depth qualitative research involving agents would make it possible to better understand their subjective experience and perception of life and work under such algorithmic governance.

References

[1]. Xu, J. J., et al. (2022). Peer-to-peer loan fraud detection: Constructing features from transaction data. MIS Quarterly, 46(3), 1777-1792.

[2]. Machado, M., et al. (2024). What do we know about fraud detection in peer-to-peer lending? A systematic literature review. A Systematic Literature Review (September 6, 2024), 1-23.

[3]. Zhang, Z., et al. (2025). Identifying e-commerce fraud through user behavior data: Observations and insights. Data Science and Engineering, 10, 24-39.

[4]. Qu, B., et al. (2024). Multi-task CNN behavioral embedding model for transaction fraud detection. In 2024 IEEE International Conference on Data Mining Workshops (ICDMW),1-7. https://doi.org/10.1109/ICDMW65004.2024.00043

[5]. Aras, M. T., & Guvensan, M. A. (2023). A Multi-Modal Profiling Fraud-Detection System for Capturing Suspicious Airline Ticket Activities. Applied Sciences, 13(24), 13121. https://doi.org/10.3390/app132413121

[6]. Stojanović, B., Božić, J., Hofer-Schmitz, K., Nahrgang, K., Weber, A., Badii, A., Sundaram, M., Jordan, E., & Runevic, J. (2021). Follow the Trail: Machine Learning for Fraud Detection in Fintech Applications. Sensors, 21(5), 1594. https://doi.org/10.3390/s21051594

[7]. Gallo, S. (2021). Fintech platforms: Lax or careful borrowers' screening?. Financial Innovation, 7, , 58.

[8]. Veltri, G. A., et al. (2023). The impact of online platform transparency of information on consumers' choices. Behavioural Public Policy, 7(1), 55–82. https://doi.org/10.1017/bpp.2020.11

[9]. Lautenschlager, J., Stramm, J., Guggenberger, T., & Urbach, N. (2025). Striking a balance: Designing a blockchain-based solution to navigate coopetition dynamics in supply chain management. Electronic Markets, 35, 70.

[10]. Oktaviani, Y., & Dewi, M. K. (2023). Is information transparency important for funders? A case study of sharia P2P lending companies in Indonesia. Journal of Accounting and Investment, 24(2), 462–486. https://doi.org/10.18196/jai.v24i2.17220