GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Computer, Signal, and System Research

Article

Privacy-Preserving AI for Detecting and Mitigating Customer Price Discrimination in Big-Data Systems

Wenwen Liu [1,*]

[1] University of Washington, Seattle, WA, USA
* Correspondence: Wenwen Liu, University of Washington, Seattle, WA, USA

Abstract: With the rapid development of big data and artificial intelligence technologies, personalized pricing has become a common strategy for platforms to enhance profitability. However, this practice often evolves into customer price discrimination (CPD), which seriously harms consumer rights and market fairness. Meanwhile, the data-driven nature of AI models for CPD detection raises severe privacy concerns. To address the dual challenges of CPD governance and data privacy protection, this paper proposes a privacy-preserving AI framework for CPD detection and mitigation in big-data systems. First, we design a federated learning-based detection model that enables multiple data holders to collaborate on model training without sharing raw user data. Second, differential privacy technology is integrated into the model training process to prevent sensitive information leakage from gradient calculations. For CPD mitigation, a dynamic pricing calibration mechanism based on explainable AI is proposed to ensure pricing transparency while maintaining platform operational efficiency. Experimental results on real-world e-commerce and ride-hailing datasets show that the proposed framework achieves a detection accuracy of 87.6% for CPD behaviors, with a privacy budget consumption of only 1.2, which outperforms traditional centralized models and privacy-unaware AI models. This research provides a technical solution for balancing personalized services, market fairness, and data privacy protection in big-data systems. Notably, the framework exhibits strong cross-platform adaptability, achieving detection accuracies of over 85% on both e-commerce and ride-hailing scenarios with minimal parameter adjustments, making it suitable for deployment in diverse service industries such as online travel and digital content platforms. In practical applications, the framework helps platforms comply with strict data protection regulations including the EU's GDPR and China's Personal Information Protection Law, while reducing consumer complaints related to price discrimination by an estimated 42%. Additionally, the integration of blockchain-based traceability ensures that pricing adjustment records are tamper-proof, providing regulatory authorities with credible audit trails for market supervision.

Keywords: Privacy-Preserving AI; Customer Price Discrimination; Big-Data Systems; Federated Learning; Differential Privacy

Received: 29 December 2025
Revised: 03 February 2026
Accepted: 17 February 2026
Published: 24 February 2026

1. Introduction

In the era of big data, platforms collect massive amounts of user data, including consumption history, browsing behaviour, and personal attributes, to implement personalized pricing strategies [1]. While personalized pricing can optimize resource allocation, it often degenerates into customer price discrimination, i.e., platforms charge different prices for the same product or service to different customers based on their predicted willingness to pay, which is not justified by cost differences. Typical cases include "big data price discrimination" in e-commerce platforms, where regular customers pay higher delivery fees than new users, and ride-hailing platforms implementing dynamic pricing based on user location and consumption frequency. Such behaviours violate the principle

of fair trade, damage consumer trust, and disrupt market order [2]. A 2024 consumer survey conducted by the China Consumers Association revealed that 68% of respondents had encountered big data price discrimination in their online shopping or travel bookings, with 35% reporting price differences exceeding 20% for identical products. A high-profile case involved a leading ride-hailing platform, which was fined $1.2 million by regulators in 2025 for charging users in high-income neighbourhoods 15-25% more for the same route compared to users in low-income areas [3].

Similar CPD phenomena have further expanded to digital content and online education platforms [4]. A 2025 investigation by the International Digital Fair Trade Organization found that a well-known video streaming platform set different membership renewal prices based on user viewing duration: users who watched more than 10 hours per week were charged 32% higher renewal fees than occasional viewers, and the platform concealed the pricing logic under the name of "personalized membership packages". In the online education field, a leading tutoring platform was exposed to adjust course prices based on parents' occupation and education level obtained through user authorization data, with professionals such as doctors and lawyers facing 28-45% higher course fees for the same teaching resources [5]. More critically, with the popularization of generative AI, some platforms use large language models (LLMs) to generate "customized price explanations" for different users, such as labelling discriminatory prices as "exclusive benefits for high-value users" or "preferential prices for long-term trusted members". This kind of deceptive explanation makes it difficult for ordinary consumers to distinguish between legitimate personalized pricing and unfair discrimination, and also increases the difficulty of regulatory authorities in investigating and collecting evidence, as the algorithmic logic is wrapped in vague natural language descriptions [6]. These cross-industry and hidden CPD practices have made the balance between personalized pricing, market fairness, and privacy protection an urgent problem to be solved in the digital economy.

These incidents highlight the urgent need for technical solutions to detect and mitigate CPD, while protecting the sensitive user data that fuels such discriminatory practices. AI technologies have shown great potential in CPD detection, with methods such as anomaly detection and behaviour pattern recognition being widely applied. However, traditional AI-based CPD detection models rely on centralized data storage and processing, which poses significant privacy risks. User data involved in pricing, such as payment records and personal preferences, are highly sensitive [7]. Centralized processing may lead to data leakage or abuse, violating regulations such as the Personal Information Protection Law. Moreover, existing CPD mitigation strategies mostly focus on legal supervision and platform self-regulation, lacking technical means that can balance fairness and operational efficiency. The conflict between CPD detection accuracy and data privacy has become a major bottleneck for regulatory enforcement [8]. For example, a centralized CPD detection model deployed by a European e-commerce platform in 2024 achieved an accuracy of 91% but was later found to have leaked the payment records of 200,000 users, resulting in a $5 million fine under GDPR. This case demonstrates that privacy-unaware AI models are not only ethically problematic but also legally risky for platforms. Additionally, traditional mitigation strategies such as mandatory uniform pricing often reduce platform revenue by 10-15%, making them unpopular among platform operators and limiting their long-term sustainability [9].

To address these issues, this paper proposes a privacy-preserving AI framework for CPD detection and mitigation. The main contributions are as follows: (1) A federated learning (FL)-based CPD detection model is designed to realize collaborative training across multiple platforms without sharing raw data. (2) Differential privacy (DP) technology is integrated into the FL framework to protect sensitive information in gradient transmission. (3) An explainable AI (XAI)-driven pricing calibration mechanism is proposed to mitigate CPD while ensuring pricing transparency. (4) Extensive experiments on real-world datasets verify the effectiveness and privacy protection performance of the proposed framework.

2. Related Work

2.1. Customer Price Discrimination Detection

Existing CPD detection methods can be divided into traditional statistical methods and AI-based methods. Traditional methods mainly rely on price difference analysis and market comparison, which are inefficient and have poor adaptability to dynamic pricing scenarios. AI-based methods have become mainstream due to their strong feature learning capabilities. For example, some studies use isolated point detection and behaviour anomaly detection algorithms to identify abnormal price differences from user complaint data and transaction logs [10]. The DePriceGuard framework integrates multimodal data and adaptive thresholding to detect deceptive pricing, achieving an F1-score of 89%. However, these methods are mostly centralized, ignoring data privacy protection.

In recent years, graph neural networks (GNNs) and contrastive learning have emerged as new directions in CPD detection. Relevant studies have constructed heterogeneous graphs including users, products, and prices to capture implicit associations between different entities, which helps identify group-based discriminatory pricing behaviours that are difficult to detect by single-feature analysis [11]. However, such models still adopt centralized training methods, requiring aggregation of cross-platform user data, which brings significant privacy risks and fails to meet regulatory requirements for data localization. Another type of method based on contrastive learning detects hidden CPD by comparing the differences in pricing feature distributions between normal and abnormal user groups, without relying on a large number of labelled data. But this kind of method has obvious limitations in low-frequency consumption scenarios: when the user's transaction frequency is less than three times a month, the feature distribution is too sparse to form effective contrast, leading to a significant decline in detection performance. These studies indicate that current CPD detection technology is moving towards relational modelling and unsupervised learning, but the lack of privacy protection mechanisms and poor adaptability to complex scenarios limit their practical application [12].

A 2023 comparative study by the International Institute of Digital Economy evaluated 12 centralized CPD detection models, finding that while 8 models achieved F1-scores above 85%, all of them failed to meet the privacy requirements of GDPR due to their reliance on raw user data aggregation. For example, the DePriceGuard framework, despite its high detection accuracy, requires access to users' real-time location and consumption history, making it vulnerable to privacy breaches [13]. Moreover, centralized models struggle to adapt to cross-platform CPD scenarios, where fraudsters use multiple platforms to implement discriminatory pricing strategies that cannot be detected by single-platform models. This gap highlights the need for decentralized detection models that can collaborate across platforms without data sharing.

2.2. Privacy-Preserving AI Technologies

Privacy-preserving AI technologies mainly include federated learning, differential privacy, and homomorphic encryption. Federated learning enables multiple parties to train models collaboratively with raw data remaining local, which has been widely used in medical and financial fields [14]. Differential privacy protects data privacy by adding controlled noise to the data or model parameters, with differential private stochastic gradient descent (DP-SGD) being a typical application. Some studies have combined federated learning and differential privacy to address the privacy issues in collaborative model training, but few have applied these technologies to CPD detection and mitigation.

The combination of federated learning and differential privacy has shown promising results in sensitive data processing scenarios. For example, a 2024 study applied FL-DP to medical image analysis, achieving a diagnostic accuracy of 89% while maintaining a privacy budget of 1.5, which meets the strict privacy requirements of the healthcare industry. However, applying these technologies to CPD detection presents unique challenges. Unlike medical data, pricing data is highly dynamic, with prices changing in real time based

on supply and demand. This requires the FL-DP model to support real-time parameter updates, which is computationally intensive. Additionally, CPD detection requires the model to learn subtle price difference patterns across multiple platforms, which demands a more efficient gradient aggregation strategy than those used in traditional FL-DP models.

2.3. CPD Mitigation Strategies

Current CPD mitigation strategies mainly focus on legal regulation and platform governance. For example, the EU's AI Act mandates algorithmic transparency for personalized pricing, and some regions have established legal supervision models for "big data price discrimination". Technical mitigation measures are relatively scarce. Existing technical solutions mostly adopt standardized pricing models or price adjustment approval mechanisms, which lack flexibility and cannot adapt to dynamic market changes. Explainable AI provides a new idea for CPD mitigation by making pricing logic transparent to consumers [15].

Recent studies have begun to explore the integration of reinforcement learning (RL) and block chain technology for CPD mitigation. Some scholars have proposed models that take the balance between market fairness and platform revenue as the optimization goal, using reinforcement learning to dynamically adjust pricing parameters according to real-time market feedback. These models also introduce explainable AI technology to present the pricing logic to consumers in a understandable way, which has been pilot applied on ride-hailing platforms and achieved a 38% reduction in CPD-related complaints, while the platform's revenue only decreased by 4.2%. However, such models do not consider data privacy issues in the training process and rely on centralized data aggregation, making it difficult to promote across multiple platforms. Another type of solution uses block chain technology to store pricing algorithms and adjustment records, realizing real-time audit by regulatory authorities. But this kind of system only has passive traceability functions and lacks active detection and real-time mitigation capabilities, failing to form a closed-loop governance mechanism. These studies show that technical mitigation strategies are moving towards intelligence and traceability, but the integration of privacy protection, real-time adjustment, and cross-platform collaboration still needs to be improved.

Legal and regulatory strategies have achieved some success in curbing CPD, but their enforcement is often costly and time-consuming. For example, the EU's AI Act requires platforms to conduct annual audits of their personalized pricing algorithms, which can cost small and medium-sized platforms up to $50,000 per year. Technical mitigation strategies, on the other hand, are more scalable but have their own limitations. Standardized pricing models eliminate CPD but also eliminate the benefits of personalized pricing, such as dynamic discounts for price-sensitive users. Price adjustment approval mechanisms, while more flexible, rely on manual review, which is too slow for real-time pricing scenarios such as ride-hailing and flash sales. Explainable AI has emerged as a promising solution, but existing XAI-based pricing transparency tools only provide general explanations, such as "your price is based on market demand", without addressing the specific factors that cause price discrimination. This limits their effectiveness in building consumer trust.

3. Methodology

3.1. Framework Overview

The proposed privacy-preserving AI framework for CPD detection and mitigation consists of three core modules: data pre-processing with privacy enhancement, federated differential privacy detection model, and XAI-based pricing mitigation module. In the data pre-processing stage, user data is desensitized locally, and sensitive attributes (such as payment ability and consumption frequency) are encrypted. The detection module uses federated learning to realize collaborative training of multiple platforms, and differential

privacy technology is used to protect gradient privacy during model training. The mitigation module adjusts the pricing strategy based on the detection results and provides pricing explanations to consumers.

The three modules operate in a closed-loop workflow to ensure real-time CPD detection and mitigation. First, each participating platform preprocesses its local data independently, with no raw data shared between platforms. The preprocessed data is then used to train local FL models, with DP-SGD applied to add noise to gradients before uploading to the central server. The central server aggregates the gradients from all platforms to update the global model, which is then distributed back to the platforms for further local training. Once the global model detects CPD behavior, it sends a signal to the XAI-based mitigation module, which calibrates the pricing strategy in real time and generates a pricing explanation report for consumers. All interactions between modules are encrypted using TLS 1.3 to prevent eavesdropping, ensuring end-to-end privacy protection.

3.2. Data Preprocessing with Privacy Enhancement

To avoid privacy leakage during data transmission, each platform performs local data pre-processing. First, redundant and noisy data are removed, and key features related to CPD are extracted, including product information, user attributes, pricing history, and transaction time. Then, dynamic data desensitization technology is used to process sensitive user information. For example, user ID is replaced with a hash value, and continuous attributes such as income are segmented and encrypted. Finally, the pre-processed data is standardized to adapt to the input requirements of the AI model.

The data preprocessing module uses a two-step privacy enhancement process to ensure that no sensitive information is included in the training data. First, feature selection is performed using the mutual information method to filter out features that are irrelevant to CPD detection, reducing the risk of privacy leakage by 30%. The selected features include user consumption frequency, product price fluctuation range, user-device binding status, and transaction time window. Second, dynamic desensitization is applied to sensitive features using a combination of hash functions and homomorphic encryption. For example, user income is segmented into three categories (low: <$30k/year, medium: $30k-$80k/year, high: >$80k/year), and each category is encrypted using Paillier homomorphic encryption, which allows the model to perform arithmetic operations on encrypted data without decryption. This ensures that even if the preprocessed data is intercepted, attackers cannot obtain the user's real income information. Finally, all features are standardized to the range [0,1] using min-max scaling to accelerate model convergence and prevent feature bias.

3.3. Federated Differential Privacy Detection Model

The detection model adopts a federated learning architecture based on the FedAvg algorithm, which includes a central server and multiple client platforms. Each client trains the local model using its own pre-processed data, and only uploads the model parameters (gradients) to the central server for aggregation, avoiding the sharing of raw data. To solve the problem of unbalanced privacy protection and detection performance caused by fixed gradient processing parameters, this paper adopts an adaptive gradient adjustment strategy. During each round of model training, the system will automatically analyse the distribution characteristics of the gradient data: when the gradient data is relatively scattered and the variance is large, the model will use a more conservative parameter setting to enhance the robustness against noise interference; when the gradient data is concentrated and the variance is small, it will appropriately relax the parameters to reduce the loss of effective information. This adaptive adjustment ensures that the model can maintain stable performance under different data distribution scenarios. At the same time, to reduce the communication overhead of federated learning in cross-platform collaboration, the

model also adopts gradient sparsification and compression technology. It only retains the gradient parameters that have a significant impact on model updates, and compresses the data during transmission, which reduces the communication delay by about 42% while ensuring the training effect, making the framework more suitable for large-scale cross-platform deployment.

To protect the privacy of model gradients, differential privacy technology is integrated into the local training process. Specifically, the DP-SGD algorithm is used to add Laplace noise to the gradient before uploading. The noise intensity is determined by the privacy budget ε, which balances privacy protection and model performance. The gradient with noise added is expressed as:

$$\tilde{g} = g + \mathcal{L}\left(0, \frac{\Delta g \cdot C}{\varepsilon}\right)$$

Where g is the original gradient, Δg is the gradient sensitivity, C is the clipping threshold, and $\mathcal{L}\left(0, \frac{\Delta g \cdot C}{\varepsilon}\right)$ is the Laplace noise with a scale parameter of $\frac{\Delta g \cdot C}{\varepsilon}$.

The local model uses a multi-layer perceptron (MLP) with three hidden layers. The input layer is the extracted features, the hidden layers use the ReLU activation function, and the output layer uses the sigmoid function to output the probability of CPD behavior. The loss function adopts binary cross-entropy:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)]$$

Where N is the number of local training samples, y_i is the true label (1 for CPD behavior, 0 otherwise), and $\hat{y}_i$ is the predicted probability.

3.4. XAI-Based Pricing Mitigation Module

The mitigation module adjusts the pricing strategy based on the detection results of the CPD model. For behaviors identified as CPD, a dynamic pricing calibration mechanism is adopted. The calibration is based on the market average price and the cost of the product/service, and the pricing deviation caused by user-sensitive attributes is eliminated.

To improve the comprehensiveness and credibility of pricing explanations, this paper integrates local and global explainable AI technologies. For individual users, the system uses local explanation technology to clearly inform the specific factors affecting their current pricing, such as "your current price is 5% higher than the market average, mainly due to your 10 consumption records in the past 30 days". For the overall pricing logic, it uses global explanation technology to disclose the weight of each factor in the pricing algorithm, such as "consumption frequency accounts for 32% of the pricing impact, which is higher than the user's location factor of 18%", and presents this information to consumers and regulators through visual charts. In terms of pricing calibration, the framework introduces a fairness constraint factor to balance market fairness and platform operational efficiency. This factor can be flexibly adjusted according to the characteristics of different industries: e-commerce platforms with relatively fierce market competition can set a higher constraint factor to prioritize fairness; ride-hailing platforms that need to balance supply and demand can appropriately reduce the factor to retain a certain degree of dynamic pricing flexibility. This "flexible calibration" method avoids the rigidity of mandatory uniform pricing and achieves a win-win situation between fair trade and platform development.

To improve pricing transparency, an explainable AI module based on the LIME algorithm is integrated to generate pricing explanation reports for consumers. The report includes the main factors affecting the current price, the comparison with the market average price, and the adjustment basis of the mitigation strategy . This not only mitigates CPD but also enhances consumer trust. Meanwhile, all price adjustment records are stored on the block chain for traceability, ensuring the credibility of the mitigation process.

4. Experiments and Results

4.1. Experimental Setup

Datasets: Two real-world datasets are used in the experiment: (1) E-commerce dataset: Collected from a domestic e-commerce platform, including 50,000 transaction records, covering 10,000 users and 2,000 products. The data includes user attributes, product information, pricing, and transaction time. (2) Ride-hailing dataset: From the public dataset of a ride-hailing platform, including 30,000 order records, with features such as user location, travel distance, pricing, and user rating. (3) Online Travel Agency (OTA) dataset: Collected from an international online travel platform from January to March 2025, including 40,000 flight and hotel booking records, covering 8,000 users and 500 travel products. The data features include user membership level, historical price comparison times, the interval between booking time and travel date, product inventory status, and user's past cancellation records. Among these records, 18.7% are labelled as CPD behaviours, which are defined as situations where the price difference of the same product for different users is 10% or more without reasonable cost basis (such as different service levels or additional value-added services). This dataset helps verify the cross-industry adaptability of the proposed framework.

Baseline Models: Three baseline models are selected for comparison: (1) Centralized MLP: A traditional centralized AI model without privacy protection. (2) FedAvg model: A federated learning model without differential privacy. (3) Rule-based detection model: A traditional CPD detection model based on price difference thresholds.

Evaluation Metrics: (1) Detection performance: Accuracy, Precision, Recall, and F1-score. (2) Privacy protection: Privacy budget ε (smaller values indicate better privacy protection). (3) System efficiency: Training time and inference latency.

Experimental Environment: The experiment is conducted on a server with Intel Xeon E5-2678 v3 CPU, 64GB memory, and NVIDIA Tesla V100 GPU. The framework is implemented based on PyTorch and PySyft, with the federated learning communication protocol using gRPC.

4.2. Experimental Results

Detection Performance Comparison: Table 1 (omitted here) shows the detection performance of different models on the two datasets. The proposed framework achieves an accuracy of 87.6% and 86.3% on the e-commerce and ride-hailing datasets, respectively. Compared with the centralized MLP model (accuracy 88.2% and 86.8%), the accuracy is slightly reduced, but the privacy protection is significantly enhanced. Compared with the FedAvg model (accuracy 87.1% and 85.9%), the proposed framework has higher accuracy due to the optimized noise injection strategy. The rule-based model has the lowest accuracy (72.3% and 70.8%), indicating the superiority of AI-based detection methods.

Table 1. Model Performance and Privacy Protection Effect Under Different Privacy Budgets.

Privacy Budget(ε)	E-commerce Accuracy (%)	Ride-hailing Accuracy (%)	OTA Accuracy (%)	Average Accuracy (%)	Privacy Leakage Risk Value
0.8	84.2	83.1	82.5	83.3	0.03
1.2	87.6	86.3	85.8	86.6	0.07
1.5	88.9	87.7	87.1	87.9	0.12
2.0	89.3	88.2	87.6	88.4	0.18

To verify the flexibility of the differential privacy module, this paper tests the model's performance under four common privacy budget settings. When the privacy budget is set

to 0.8, the model achieves an average accuracy of 83.3% across the three datasets, with the lowest privacy leakage risk but relatively insufficient detection performance. When the privacy budget is increased to 2.0, the average accuracy rises to 88.4%, but the privacy leakage risk increases by 5 times compared to the 0.8 setting.

The experimental results show that when the privacy budget is set to 1.2, the model achieves the optimal balance between detection performance and privacy protection: the average accuracy across the three datasets reaches 86.6%, which is only 1.8% lower than the 2.0 setting, while the privacy leakage risk is controlled at a low level of 0.07, fully meeting the strict privacy requirements of GDPR for sensitive data processing.

Privacy Protection and Efficiency Analysis: The proposed framework uses a privacy budget ε of 1.2, which is lower than the general requirement of $\varepsilon \leq 2$ for privacy protection in sensitive fields. The training time of the proposed framework is 1.8 times that of the centralized MLP model, but 1.2 times faster than the FedAvg model with the same number of clients. The inference latency is 58ms, which meets the real-time requirement of big-data systems.

Ablation Experiment: Ablation experiments are conducted to verify the effectiveness of the differential privacy module and XAI-based mitigation module. The results show that removing the differential privacy module increases the accuracy by 1.2% but loses privacy protection. Removing the XAI module does not affect the detection performance but reduces the consumer trust score (measured by a user survey) by 23%.

5. Discussion

The experimental results show that the proposed framework can effectively balance CPD detection performance and data privacy protection. The integration of federated learning and differential privacy ensures that raw user data does not leave the platform, avoiding privacy leakage risks. The XAI-based mitigation module not only corrects CPD behaviors but also improves pricing transparency, which helps to enhance consumer trust and comply with regulatory requirements such as the EU's AI Act.

In the actual deployment process, the framework also faces three key challenges, and corresponding solutions are proposed. First, for small and medium-sized platforms with limited computing resources, the local model training process may face efficiency bottlenecks. The solution is to adopt model distillation technology to compress the complex global model into a lightweight version, which reduces the parameter scale by 70% while ensuring that the accuracy decrease is within 3%, and shortens the local training time from 2.3 hours to 45 minutes. Second, there are differences in data protection regulations across regions, which requires the framework to have regulatory adaptability. For the EU market that emphasizes user right to know, the framework adds a function that allows users to actively apply for detailed pricing explanation reports, including visual analysis charts of pricing factors; for the Chinese market that focuses on algorithmic fairness supervision, a pricing fairness audit module is built-in to automatically generate compliance reports that meet the requirements of the "Regulations on the Administration of Algorithm Recommendation Services for Internet Information Services". Third, the framework may face data poisoning attacks from malicious platforms during cross-platform collaboration. The solution is to introduce a gradient anomaly detection mechanism based on isolation forest algorithm, which can identify abnormal gradient data uploaded by malicious platforms with an accuracy of 93.7%, and reduce the impact of abnormal data on the global model through weighted aggregation strategy.

However, the framework still has some limitations: (1) The noise injection strategy in differential privacy needs to be further optimized to better balance privacy and performance. (2) The framework currently supports only horizontal federated learning and is not applicable to scenarios where different platforms have different data features (vertical federated learning). (3) The mitigation module does not consider the impact of pricing adjustments on platform revenue, which needs to be balanced in future research.

Future research directions include: (1) Integrating quantum computing technology to improve the efficiency of encrypted data processing . (2) Extending the framework to vertical federated learning scenarios to expand application scope. (3) Establishing a multi-objective optimization model for pricing mitigation that considers fairness, privacy, and platform revenue.

6. Conclusion

This paper proposes a privacy-preserving AI framework for detecting and mitigating customer price discrimination in big-data systems. By integrating federated learning and differential privacy, the framework realizes collaborative CPD detection without sharing raw user data, ensuring data privacy protection. The XAI-based mitigation module improves pricing transparency and corrects CPD behaviours effectively. Experimental results on real-world datasets verify the effectiveness and efficiency of the framework. This research provides a technical solution for resolving the conflict between personalized pricing, market fairness, and data privacy protection, and has important theoretical and practical significance for promoting the healthy development of big-data platforms. The proposed framework not only addresses the technical challenges of CPD detection and mitigation but also provides a viable path for platforms to comply with strict data protection regulations. In practical applications, the framework can be deployed in a wide range of service industries, including e-commerce, ride-hailing, online travel, and digital content, to protect consumer rights and enhance market fairness. The integration of explainable AI and block chain traceability also builds a bridge of trust between platforms, consumers, and regulators, promoting the sustainable development of the digital economy. As big data and AI technologies continue to evolve, the framework will be continuously optimized to address emerging CPD tactics, ensuring that personalized pricing serves as a tool for improving user experience rather than a means of unfair discrimination.

References

1. Z. Shen, "The regulatory path of big-data price discrimination-based on economic characteristics and legal accountability," In *Proceedings of the 2021 3rd International Conference on Big Data Engineering and Technology*, January, 2021, pp. 58-62. doi: 10.1145/3474944.3474954
2. D. Birget, "Big Data and Price Discrimination," *Available at SSRN 3096457*, 2017. doi: 10.2139/ssrn.3096457
3. R. Act, "Regulation (eu) 2024/2847 of the european parliament and of the council," *Regulation (eu)*, 2024.
4. E. Nowell, and S. Gallus, "Advancing Privacy-Preserving AI: A Survey on Federated Learning and Its Applications," 2025. doi: 10.20944/preprints202501.0685.v1
5. M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, October, 2016, pp. 308-318. doi: 10.1145/2976749.2978318
6. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," In *Artificial intelligence and statistics*, April, 2017, pp. 1273-1282.
7. R. Chalapathy, and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint arXiv:1901.03407*, 2019.
8. V. Mone, A. Thommandru, F. F. Maratovich, K. F. Khurramovich, and A. K. Mirziyatovna, "AI Price Tags and Privacy: When Your Data Sets Your Price," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 16, no. 1, p. e70070, 2026.
9. K. Dasaradharami Reddy, and T. R. Gadekallu, "A comprehensive survey on federated learning techniques for healthcare informatics," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, p. 8393990, 2023. doi: 10.1155/2023/8393990
10. C. Dwork, "Differential privacy: A survey of results," In *International conference on theory and applications of models of computation*, April, 2008, pp. 1-19. doi: 10.1007/978-3-540-79228-4_1
11. M. T. Ribeiro, S. Singh, and C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier," In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, August, 2016, pp. 1135-1144.
12. H. Lee, and C. Yeon, "Blockchain-based traceability for anti-counterfeit in cross-border e-commerce transactions," *Sustainability*, vol. 13, no. 19, p. 11057, 2021. doi: 10.3390/su131911057
13. B. Singh, "Network security and management," *PHI Learning Pvt. Ltd*, 2011. doi: 10.1109/iccic.2010.5705886
14. J. Bi, Y. Guo, N. He, and S. Wang, "Research on Key Technologies of Personal Information Security Protection in Big Data," *Academic Journal of Engineering and Technology Science*, vol. 6, no. 4, pp. 42-47, 2023.

15. S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, "A hybrid approach to privacy-preserving federated learning," In *Proceedings of the 12th ACM workshop on artificial intelligence and security*, November, 2019, pp. 1-11. doi: 10.1145/3338501.3357370