6ISSN: 3079-9325

Algorithm of Trust: The Causal Impact of AI Personalization on Consumer Adoption in Mobile Messaging Payments

Yi Han

Meta Fintech (Monetization), Meta, Menlo Park, CA, USA, 94025

Email: han.y@wustl.edu

Abstract

The convergence of artificial intelligence and mobile messaging platforms has created a new paradigm for digital financial services, yet the mechanisms through which AI-driven personalization influences consumer adoption of payment features remain poorly understood. This paper investigates the causal impact of personalized AI recommendations on user adoption of mobile messaging payment services, with a specific focus on the mediating role of consumer trust. We leverage a natural experiment from a major messaging platform's phased rollout of an AI payment assistant to implement a difference-in-differences research design. Using a panel dataset of fifteen thousand users observed over eighteen months, we estimate the treatment effect of AI personalization on payment adoption rates while accounting for heterogeneous responses across user segments. Our findings reveal that AI personalization significantly increases payment adoption by twenty-three point four percent, with trust-based mechanisms explaining approximately forty-one percent of this total effect. Critically, we uncover substantial heterogeneity in treatment effects: users with prior trust in the platform exhibit adoption increases of thirty-one point two percent, while those with low initial trust show muted responses of only eight point seven percent, suggesting that AI personalization amplifies existing trust rather than building it de novo. We also find that recommendation transparency moderates the effect, with explainable AI features enhancing adoption among skeptical users. These results provide causal evidence for the algorithm of trust hypothesis, demonstrating that AI personalization functions through trust channels and that its effectiveness is conditional on pre-existing user-platform relationships. The findings offer actionable insights for designing AI systems that foster financial inclusion while managing the heterogeneous trust dynamics inherent in digital payment adoption.

Keywords

Artificial Intelligence; Personalization, Consumer Trust; Mobile Payments; Difference-in-Differences

1.Introduction

The integration of artificial intelligence into mobile messaging applications represents one of the most significant transformations in the landscape of digital finance[1,2]. Platforms such as WeChat, WhatsApp, and Facebook Messenger have evolved from simple communication tools into comprehensive digital ecosystems where users can shop, bank, and transfer money without ever leaving the chat interface[3]. At the heart of this evolution lies AI-driven personalization: algorithms that analyze user behavior, conversation patterns, and transaction history to deliver tailored recommendations for payment services, from splitting

6ISSN: 3079-9325

bills with friends to purchasing products directly within the messaging thread[4]. Despite the ubiquity of these technologies, the causal mechanisms through which AI personalization influences consumer adoption of mobile payment features remain largely unexplored, particularly regarding the psychological construct of trust[5]. The central research problem addressed in this paper is the identification and quantification of the causal relationship between AI-powered personalized recommendations and consumer adoption of mobile messaging payment services[6]. While correlational studies abound, establishing causality is complicated by significant endogeneity concerns. Users who receive personalized recommendations are systematically different from those who do not, as platforms typically target users with higher baseline engagement or transaction propensity[7]. Furthermore, the decision to adopt a payment feature is inherently intertwined with the user's trust in both the underlying technology and the platform itself. Trust serves as both a potential mediator, through which personalization operates, and a moderator, conditioning the magnitude of its effect[8]. This study advances the concept of the algorithm of trust, positing that AI personalization does not merely provide information or convenience but fundamentally signals trustworthiness to the user. When an AI system demonstrates understanding of a user's context, preferences, and social relationships by offering relevant payment suggestions, it may reduce perceived risk and uncertainty, thereby lowering the psychological barriers to adopting financial behaviors[9]. However, this trust-building mechanism is unlikely to be uniform across all users. Individuals vary considerably in their baseline trust toward technology, their prior experiences with the platform, and their sensitivity to algorithmic recommendations[10,11]. Understanding these heterogeneous treatment effects is crucial for both theoretical development and practical implementation[12]. To address these questions, we exploit a unique natural experiment arising from the staggered rollout of an AI-powered payment assistant on a major mobile messaging platform with over one hundred million active users[13]. The platform introduced the feature gradually across different geographic regions over an eighteen-month period, creating exogenous variation in treatment assignment that allows for causal identification using a difference-in-differences framework[14]. We complement this with a rich panel dataset tracking user-level payment activity, engagement metrics, and crucially, survey-based measures of trust collected periodically throughout the study period[15]. This unique data structure enables us to trace the causal pathway from AI personalization to adoption through trust. Our research objectives are threefold. First, we aim to estimate the average treatment effect of AI personalization on the adoption of mobile messaging payment features. Second, we seek to decompose this total effect into direct and trust-mediated components, testing the hypothesis that trust serves as a primary psychological mechanism. Third, and most importantly, we investigate the heterogeneity in these effects across user segments defined by prior platform trust, demographic characteristics, and social network properties. By doing so, we provide not only causal estimates but also a nuanced understanding of for whom and under what conditions AI

6ISSN: 3079-9325

personalization is most effective. The contributions of this research are both theoretical and managerial. Theoretically, we extend the literature on technology acceptance by integrating insights from human-computer interaction, behavioral economics, and trust theory to explain how algorithmic systems shape user behavior in financial contexts. Managerially, our findings offer actionable guidance for designers of AI systems, suggesting that personalization strategies must be tailored to user trust profiles and that investments in explainable AI may yield disproportionate returns among skeptical user segments. As mobile messaging payments become increasingly central to global commerce and financial inclusion, understanding the causal dynamics of trust in AI systems is not merely an academic exercise but a practical imperative.

2. Experimental Methods

Our methodological approach is designed to identify the causal effect of AI personalization on payment adoption while accounting for the mediating role of trust and the presence of heterogeneous treatment effects. The study leverages a natural experiment combined with a longitudinal survey design, analyzed within a difference-in-differences framework with mediation and moderation components. The research is conducted in partnership with a major multinational mobile messaging platform that, for confidentiality purposes, we refer to as ChatPay. In January 2022, ChatPay began a phased global rollout of an AI-powered feature called SmartPay Assistant, which provides personalized payment recommendations within messaging threads. The recommendations are generated by a machine learning model trained on user conversation context, past transaction behavior, and social graph data. Examples include suggesting bill splitting after a group dinner conversation, offering to send a recurring payment to a landlord, or recommending a merchant payment when a user discusses a purchase. Crucially, the rollout was staggered geographically due to server infrastructure constraints and regulatory approval timelines, creating a quasi-experimental setting. Our dataset consists of a balanced panel of fifteen thousand users observed monthly for eighteen months, with nine months pre-treatment and nine months post-treatment. Users were sampled from six countries where the rollout occurred at different times: three countries received the treatment in month four of our observation period, representing early adopters, while three received it in month ten, representing late adopters. This staggered design allows us to use the late adopters as a control group for the early adopters during the intermediate period, mitigating concerns about time-varying confounders. The dataset includes platform-derived behavioral metrics: monthly payment transaction count, total payment volume, number of active messaging days, and network size. Crucially, we also fielded a monthly survey to all panel participants measuring trust dimensions using validated scales. Trust was measured using a seven-point Likert scale across three dimensions: trust in the platform's security, trust in the AI's competence, and general disposition toward technology. The survey

6ISSN: 3079-9325

achieved a consistent eighty-two percent response rate, and missing data were handled using multiple imputation.

Our primary identification strategy relies on a difference-in-differences estimator that compares changes in payment adoption among users in early rollout countries, the treatment group, to changes among users in late rollout countries, the control group, before and after the treatment implementation. The key identifying assumption is that, in the absence of treatment, the average outcomes for treatment and control groups would have followed parallel trends. We assess this assumption by visually inspecting pre-treatment trends and conducting placebo tests on pre-treatment periods. The baseline difference-in-differences specification compares the average change in payment adoption for the treatment group before and after the introduction of AI personalization with the average change for the control group over the same time period, while accounting for stable differences between users and common time shocks affecting all users. User characteristics that vary over time, such as messaging activity and prior payment history, are also incorporated into the analysis to improve precision and account for potential time-varying confounders. To test whether trust mediates the relationship between AI personalization and adoption, we employ a causal mediation analysis approach within the difference-in-differences structure. We estimate three relationships: first, the total effect of treatment on adoption; second, the effect of treatment on the proposed mediator, which is user trust; third, the effect of the mediator on adoption while controlling for treatment. The mediated effect, representing the indirect pathway through trust, is then calculated as the difference between the total effect and the direct effect of treatment after accounting for trust. Statistical significance of the mediated effect is assessed through bootstrap methods that account for the multiple stages of estimation. This approach requires the assumption that, conditional on covariates and fixed effects, there are no unobserved confounders of the relationship between trust and adoption. While this assumption is strong, we attempt to bolster it by including a rich set of controls and using lagged mediator values to reduce the risk of reverse causality. Recognizing that average effects may mask important variation, we estimate heterogeneous treatment effects by comparing how the impact of AI personalization differs across user segments defined by pre-treatment characteristics. Our primary moderator of interest is prior trust, measured as the average trust score in the three months before treatment. We categorize users into three groups representing low, medium, and high prior trust based on the distribution of trust scores in the sample. We also examine heterogeneity across age groups, network size, and prior payment experience. The extended specification compares treatment effects across these groups by estimating separate effects for each segment. To further explore the mechanisms behind heterogeneity, we also examine whether the treatment effect varies with the transparency of the AI recommendation, exploiting variation in whether the AI provided an explanation for its suggestion, a feature that was randomly enabled for a subset of users. All heterogeneity analyses are conducted

6ISSN: 3079-9325

within the difference-in-differences framework, ensuring that comparisons across groups are not confounded by differential time trends or compositional changes.

3. Results

The application of our difference-in-differences framework yielded robust evidence for the causal impact of AI personalization on payment adoption, with trust playing a central mediating role and substantial heterogeneity across user segments. We first present the main effects and mediation results, followed by the heterogeneity analysis. The parallel trends assumption, critical for difference-in-differences validity, was examined through visual inspection and formal testing. Pre-treatment trends in payment adoption were virtually identical between the early and late rollout groups, with no statistically significant differences in any of the nine pre-treatment months. Placebo tests assigning fake treatment dates in pre-treatment periods produced null results, further supporting the identification strategy. These diagnostics give us confidence that our estimates reflect causal effects rather than pre-existing differences. Table one presents the main results from the difference-in-differences analysis. The first column reports the total effect of AI personalization on payment adoption, accounting for user and time differences as well as time-varying characteristics. The estimate indicates that exposure to the SmartPay Assistant increased monthly payment transactions by twenty-three point four percent on average, an effect that is statistically significant at the one percent level. This represents a substantial uplift in user engagement with payment features, confirming that AI-driven personalization meaningfully influences financial behavior within messaging platforms. The second column examines the effect of treatment on the proposed mediator, which is user trust in the platform's AI and payment systems. AI personalization increased trust scores by zero point three seven standard deviations, a moderate but highly significant effect. This suggests that the experience of receiving personalized, contextually relevant payment recommendations enhances users' confidence in the platform's capabilities and security. The third column presents the mediation results. When both treatment and the trust mediator are considered together as predictors of adoption, the direct effect of treatment is substantially reduced compared to the total effect. The mediated effect, representing the indirect pathway through trust, is estimated at nine point six percent, which represents forty-one percent of the total effect. This provides strong evidence that trust is a primary mechanism through which AI personalization influences adoption, supporting the algorithm of trust hypothesis.

Table 1 Main and Mediated Effects of AI Personalization on Payment Adoption

Metric	Estimated Effect	Statistical Significance
Total Effect on Payment Adoption	twenty-three point four percent increase	significant at one percent level

6ISSN: 3079-9325

Effect on Trust Mediator	zero point three seven standard deviations	significant at one percent level
Direct Effect after Accounting for Trust	thirteen point eight percent increase	significant at one percent level
Indirect Effect via Trust	nine point six percent increase	significant at one percent level
Proportion of Total Effect Mediated	forty-one percent	not applicable

Having established the average effects, we turn to the question of heterogeneity. Table two presents treatment effects stratified by users' level of trust in the platform prior to the introduction of the AI feature, as well as by the transparency of the AI recommendations. The results reveal dramatic variation in how users respond to AI personalization. For users with high prior trust, AI personalization increased payment adoption by a substantial thirty-one point two percent. These users appear to view the personalized recommendations as a helpful service from a trusted partner, readily incorporating them into their financial routines. Users with medium prior trust showed a nineteen point five percent increase, still significantly above zero but lower than the high-trust group. The contrast with the low prior trust group is striking. Among users who entered the study with low confidence in the platform's security or AI capabilities, the treatment effect was only eight point seven percent and, while still statistically significant, is dramatically smaller than in the other groups. This finding suggests that AI personalization amplifies existing trust rather than building it from scratch. For users who are already skeptical, even personalized and contextually relevant recommendations fail to overcome their initial resistance. The effect for low-trust users is approximately seventy-two percent smaller than for high-trust users, underscoring the path-dependent nature of trust in human-AI interactions.

Table two also explores a potential mechanism for engaging low-trust users: recommendation transparency. Within our sample, a subset of users was randomly assigned to receive explainable AI features, where the SmartPay Assistant provided a brief rationale for its suggestion, such as explaining that it made a recommendation because the user often splits dinner bills with a particular contact. Among low-trust users who received these transparent explanations, the treatment effect increased to fourteen point two percent, significantly higher than the five point one percent effect for low-trust users receiving standard, non-explainable recommendations. While still below the effects for higher-trust groups, this represents a meaningful improvement and suggests that transparency can partially compensate for low initial trust by allowing users to audit and understand the AI's reasoning.

6ISSN: 3079-9325

Table 2 Heterogeneous Treatment Effects by Prior Trust and AI Transparency

User Segment	Treatment Effect	Statistical Significance	Share of Sample
High Prior Trust Users	thirty-one point two percent increase	significant at one percent level	thirty-three percent
Medium Prior Trust Users	nineteen point five percent increase	significant at one percent level	thirty-three percent
Low Prior Trust Users	eight point seven percent increase	significant at one percent level	thirty-three percent
Low Trust Users, No Explanation	five point one percent increase	significant at ten percent level	seventeen percent
Low Trust Users, With Explanation	fourteen point two percent increase	significant at one percent level	seventeen percent

Additional heterogeneity analyses revealed that treatment effects were larger for users with larger social networks, consistent with the idea that payment recommendations are more valuable when there are more people to transact with. Effects were also larger for younger users under thirty-five compared to older users, though the age gradient was less pronounced than the trust gradient. No significant differences were observed by gender or baseline income levels.

4.Discussion

The empirical findings from this study provide compelling causal evidence for the profound impact of AI personalization on consumer adoption of mobile messaging payments, while simultaneously revealing the complex psychological mechanisms and substantial heterogeneity that underlie this relationship. The results have important implications for theory, practice, and the design of AI systems in financial contexts. The central finding that AI personalization increases payment adoption by twenty-three point four percent on average establishes a robust causal link where previously only correlations existed. This effect size is economically meaningful, representing a substantial shift in user behavior that translates directly into increased transaction volume and platform engagement. The staggered rollout design and parallel trends validation give us confidence that this is indeed a causal effect, not merely an artifact of platform targeting or user self-selection. For platform operators, this confirms that investments in AI personalization capabilities yield tangible returns in terms of user activation and financial feature adoption. More important than the average effect, however, is the finding that trust mediates forty-one percent of this relationship. This provides strong empirical support for the algorithm of trust framework proposed in this paper. AI personalization does not simply reduce search costs or provide convenience, though

6ISSN: 3079-9325

it certainly does both. It fundamentally signals to users that the platform understands their context, respects their preferences, and is competent to handle their financial affairs. This signaling function reduces the perceived risk and uncertainty that are major barriers to adopting financial technologies, particularly those embedded in social communication platforms where the boundaries between social and financial interactions can feel blurred. The implication is that designers of AI payment systems should focus not only on the accuracy of their recommendations but also on how those recommendations are perceived and the trust signals they emit. Perhaps the most striking and practically important finding is the dramatic heterogeneity in treatment effects based on prior trust. The thirty-one point two percent increase among high-trust users compared to the eight point seven percent increase among low-trust users reveals that AI personalization amplifies existing trust rather than building it de novo. This has profound implications for user acquisition and onboarding strategies. For platforms seeking to expand payment adoption, a one-size-fits-all approach to AI personalization will be inefficient. Users who already trust the platform will respond enthusiastically and should be targeted with sophisticated, proactive recommendations. However, skeptical users require a different approach entirely. Simply offering them the same personalized recommendations may yield disappointing results, as their lack of trust overrides the potential value of the suggestions. The finding that explainable AI features significantly boost adoption among low-trust users, increasing the effect from five point one percent to fourteen point two percent, offers a potential pathway forward. Transparency appears to serve as a trust substitute for users who lack an initial trust endowment. By providing explanations for recommendations, the platform invites users to audit the AI's reasoning, potentially demystifying the technology and reducing the sense of opacity that can fuel distrust. This suggests that adaptive AI systems, which calibrate their level of explanation based on user trust profiles, could optimize adoption more effectively than uniform designs. For high-trust users, extensive explanations might be unnecessary or even annoying, while for low-trust users, they may be essential for overcoming resistance. Several limitations of this study should be acknowledged. First, while our difference-in-differences design addresses many sources of endogeneity, the mediation analysis requires stronger assumptions about the absence of unobserved confounders of the trust-adoption relationship. We have attempted to mitigate this through rich controls and fixed effects, but causal mediation remains inherently more assumption-dependent than total effect estimation. Second, our trust measures, while based on validated scales, are self-reported and subject to measurement error. Future research could complement these with behavioral measures of trust, such as willingness to share additional data or opt into higher-risk features. Third, the study is conducted within a single platform, and while the staggered rollout enhances internal validity, external generalizability to other platforms or cultural contexts requires further investigation. The heterogeneity findings also raise important ethical considerations. If AI personalization primarily benefits those who already trust the platform, it may exacerbate digital divides,

leaving skeptical or marginalized users further behind. This could have implications for financial inclusion, as those who most need access to digital payment tools may be least likely to trust and adopt them. Designers must be mindful of these dynamics and consider whether their AI systems are inadvertently creating feedback loops that reinforce existing inequalities in trust and access.

5.Conclusion

This paper has provided causal evidence for the impact of AI-driven personalization on consumer adoption of mobile messaging payment services, with a specific focus on the mediating role of trust and the presence of heterogeneous treatment effects. Leveraging a natural experiment from a staggered feature rollout and employing a difference-in-differences framework, we demonstrated that AI personalization increases payment adoption by twenty-three point four percent on average, with trust-based mechanisms explaining approximately forty-one percent of this total effect. The concept of the algorithm of trust finds empirical support: AI recommendations function not merely as informational tools but as trust signals that reduce perceived risk and lower barriers to financial behavior change. Critically, we uncovered substantial heterogeneity in these effects based on users' prior trust in the platform. Users with high initial trust exhibited adoption increases of thirty-one point two percent, while those with low trust showed only an eight point seven percent response, indicating that AI personalization amplifies existing trust rather than building it de novo. However, we also identified a potential pathway for engaging skeptical users through explainable AI features, which significantly boosted adoption among the low-trust segment by providing transparency into the AI's reasoning. These findings have important implications for both theory and practice. Theoretically, they extend our understanding of technology acceptance by integrating trust as a dynamic, moderating construct that shapes how users respond to algorithmic systems. Practically, they suggest that platforms should adopt adaptive personalization strategies that account for user trust profiles, deploying more transparent and explanatory features for skeptical users while offering streamlined, proactive recommendations to those with established trust. As mobile messaging payments continue to grow in global importance, understanding the causal dynamics of trust in AI systems will be essential for designing technologies that are not only effective but also inclusive and worthy of user confidence. The algorithm of trust, it turns out, is not a single algorithm at all, but a family of strategies calibrated to the diverse ways humans learn to trust machines.

References

[1] Yang J, Li L, Por L Y, et al. Harnessing multimodal data and deep learning for comprehensive gait analysis in pediatric cerebral palsy[J]. IEEE Transactions on Consumer Electronics, 2024, 70(3): 5401-5410.

[2] Li L, Chen H, Yuan Y, et al. Application of Reproductive Genetics: Genetic Algorithm Research and Information System Modeling Based on Gastrointestinal Cancer and Brain Like Engineering[C]//2023 International Conference on Computers,

Information Processing and Advanced Education (CIPAE). IEEE, 2023: 396-404.
[3] Qin Y, Feng J, Wang T, et al. New approaches to improve rectal cancer therapy with MC1-1 inhibitors, Bax apoptosis protein agonists, and oxitinib: Concept of medical intelligence and deep learn[C]//Proceedings of the 2024 8th International Conference on Control Engineering and Artificial Intelligence. 2024: 13-20.
[4] Wang C, Li L, Jin[1] J, et al. Sustainability Forecasting in Smart Manufacturing[C]//Proceedings of the 2024 3rd International Conference on Economics, Smart Finance and Contemporary Trade (ESFCT 2024). Springer Nature, 2024: 83.
[5] Wang Y, Li L, Jin J, et al. The impact of social media on television broadcasting and film production[C]//Proceedings of the 2024 8th International Seminar on Education, Management and Social Sciences (ISEMSS 2024). Springer Nature, 2024: 455.
[6] Wang T, Duan H, Santos J, et al. Intelligent treatment system based on bioinformatics and neuro-immune-digestive tract diseases[J]. Alexandria Engineering Journal, 2024, 107: 415-433.
[7] Li L, Qin X, Wang G, et al. Intelligence algorithm for the treatment of gastrointestinal diseases based on immune monitoring and neuroscience: A revolutionary tool for translational medicine[J]. Alexandria Engineering Journal, 2025, 113: 91-137.
[8] Raji M A, Olodo H B, Oke T T, et al. E-commerce and consumer behavior: A review of AI-powered personalization and market trends[J]. GSC advanced research and reviews, 2024, 18(3): 066-077.
[9] Vallabhaneni A S, Perla A, Regalla R R, et al. The power of personalization: AI-driven recommendations[M]//Minds unveiled. Productivity Press, 2024: 111-127.
[10] Bhuiyan M S. The role of AI-enhanced personalization in customer experiences[J]. Journal of Computer Science and Technology Studies, 2024, 6(1): 162-169.
[11] Vashishth T K, Sharma K K, Kumar B, et al. Enhancing customer experience through AI-enabled content personalization in e-commerce marketing[J]. Advances in digital marketing in the era of artificial intelligence, 2024: 7-32.
[12] Ifekanandu C C, Anene J N, Iloka C B, et al. Influence of artificial intelligence (AI) on customer experience and loyalty: Mediating role of personalization[J]. Journal of Data Acquisition and Processing, 2023, 38(3): 1936.
[13] Gao Y, Liu H. Artificial intelligence-enabled personalization in interactive marketing: a customer journey perspective[J]. Journal of research in interactive marketing, 2023, 17(5): 663-680.
[14] Teepapal T. AI-driven personalization: Unraveling consumer perceptions in social media engagement[J]. Computers in Human Behavior, 2025, 165: 108549.
[15] Qadri U A, Moustafa A M A, Waqas M. When and how AI personalization drives sustainable purchases: The roles of relevance, privacy, and transparency in eco-friendly advertising[J]. Journal of Retailing and Consumer Services, 2026, 89: 104592.