Better than Human?
Experiments with AI Debt Collectors

James J. Choi
Yale University and NBER

Dong Huang
Yale University

Zhishu Yang
Tsinghua University

Qi Zhang
Shanghai Jiao Tong University

December 6, 2024

Abstract: How good is AI at persuading humans to perform costly actions? We study calls made to persuade delinquent consumer borrowers to repay. Regression discontinuity and a randomized experiment reveal that AI is substantially less effective than human callers. Replacing AI with humans six days into delinquency closes much of the gap. But borrowers initially contacted by AI have repaid 1% less of the initial late payment one year later and are more likely to miss subsequent payments than borrowers who were always called by humans. AI's lesser ability to extract promises that feel binding may contribute to the performance gap.

We thank Zhao Jin, Jiacheng Liu, Xueming Luo, Jinfei Sheng, Haokun Sun, Yijun Zhou, and seminar participants at the Columbia & RFS AI in Finance Conference, China Financial Research Conference, China International Conference in Finance, ABFER-JFDS Conference, FMA Annual Meeting, University of Alberta, University of Chicago, Emory, and Yale for helpful discussions and comments. We thank Manlin Sun and Hongyu Wu for excellent research assistance.

Rapid progress in artificial intelligence (AI) has revived the long-standing debate on the extent to which new technologies will eliminate human jobs.[1] In this paper, we study the effectiveness of AI in a different sort of task than has previously been studied: persuading a human to perform a personally costly action. Many service and managerial jobs require performing this type of task—for example, coaxing a colleague to exert extra uncompensated effort for the good of his team (e.g., serve on a university committee), inducing a customer to make a sacrifice like switching airplane seats as a courtesy to others, or asking somebody to honestly report the details of an accident for insurance claim adjustment purposes.

The specific task we study is persuading delinquent consumer borrowers to repay their debt. Contact from a debt collector is a common experience; in 2022, 26% of U.S. adults with a credit bureau record had debt in collections.[2] The job of a debt collector is non-routine, requires social interaction, and is aided by emotional intelligence. In addition, repaying one's debts is usually seen as a moral obligation (Guiso, Sapienza, and Zingales, 2013; Bursztyn et al., 2019), which may cause AI to be less effective than humans at persuasion in this domain, since being observed committing a moral transgression by another person is more aversive than being observed by a machine (LaMothe and Bobek, 2020; Cohn et al., 2022; Kim et al., 2023).

We use debt collection data from a leading online consumer finance company in China that makes uncollateralized installment loans. Borrowers who fail to make their monthly payments on time are contacted on the phone by the company's debt collectors, urging them to repay. The company uses both human and AI callers, which allows us to evaluate AI callers' performance relative to humans and to estimate the impact of AI on the company's profits and worker productivity. The AI callers can understand the borrower's speech and generate appropriate voice replies. They provide borrowers with basic information, answer questions, and inform them of the negative consequences of defaulting. An important intermediate goal that both AI and human callers try to achieve is to extract a verbal promise to pay from the borrower. Previous research has shown that people find promises psychologically costly to break (Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006; Vanberg, 2008; He et al., 2017; Wang and Zhou, 2024).

[1] See, for example, Brynjolfsson and Mitchell (2017), Felten et al. (2020), Eloundou et al. (2023), and World Economic Forum (2020, 2023).
[2] https://apps.urban.org/features/debt-interactive-map/?type=overall&variable=totcoll (accessed April 30, 2024).

We identify the relative effectiveness of AI callers using two experiments that occurred in the firm, one natural and one intentional. The natural experiment is created because of the company's rule that nearly all newly delinquent debts with remaining principal no greater than 300 yuan (approximately 42 U.S. dollars) are permanently assigned to AI callers, whereas larger debts are transferred to human callers no later than six days after delinquency begins. Therefore, we can identify the effect of permanent versus temporary assignment to AI using a regression discontinuity design around the 300 yuan threshold. The intentional experiment is created because each month, the company takes a random 10% of newly delinquent debts with remaining principal greater than 300 yuan and assigns a randomly chosen half to be called by AI through day 5 before being called by humans thereafter (the treatment group) and assigns the other half to always be called by humans (the control group). All debts in this 10% subsample are reallocated to human callers on day 6, so this intentional experiment identifies the effect of a short-lived initial exposure to AI callers versus no exposure to AI callers.

We find in the regression discontinuity sample that when AI callers are permanently assigned to a borrower, they consistently perform worse than human callers over horizons up to one year past due, as measured by the net present value (NPV) of collected repayments scaled by the initial overdue balance. The productivity gap between AI and human callers first widens as days past due increase. It reaches its maximum around one month past due, when the NPV of repayments collected by AI callers is 9 percentage points less than that of human callers. The gap slowly narrows afterward but remains around 5 percentage points even one year past due. In addition, the gap is larger for borrowers with lower credit scores. A very stylized model of debt collectors would characterize their task as merely providing reminders and information to borrowers, and perhaps imposing nuisance costs as well—things that AI can do nearly as well as humans. Indeed, Roll and Moulton (2019) find that automated payment reminders decrease severe delinquencies. But the persistent gap in performance between AI and humans and its heterogeneity by credit quality suggest that this stylized model is importantly incomplete.

The randomized experiment shows that replacing AI callers with human callers after a few days mitigates much of the initial underperformance of AI callers. In this subsample, we continue to find that AI underperforms humans, with the NPV gap monotonically increasing to 12 percentage points by day 5. But the gap quickly narrows once human callers take over the AI cases to 2 percentage points on day 10 and 0.7 percentage points on day 30. Interestingly, the remaining

0.7 percentage point gap never closes over the next 11 months, indicating that initial contact by AI *permanently* impairs the ability of the company to collect. There may be something uniquely damaging about being contacted by AI. Repayment reductions resulting from initial contact by a less effective human call (calls on the weekend or by an inexperienced human caller) are mitigated within a few days. Using a different subset of debts, we find that permanent damage occurs even if borrowers are contacted by AI for only one day before being assigned to human callers.

AI also creates alienation beyond the initial subject of contact. One can imagine that if a department chair sends her AI to ask a colleague to serve on an onerous committee, that colleague is less likely to agree to future requests by the chair, even if made in person. In our setting, for each of the following 12 monthly payments due, borrowers previously contacted by AI are 1 to 2 percentage points more likely to be late.

We explore some potential sources of the initial AI performance gap by examining detailed outcomes of phone conversations in the randomized experiment sample, restricting to phone calls on the first day of contact. Humans call borrowers nearly one more time per day than AI callers. To remove the impact of additional phone calls, we further restrict our sample to the first call answered by borrowers. After controlling for the call's time of day, we find that AI callers are not more likely to be hung up on within 10 seconds, but they have conversations that are 31 seconds shorter on average and exhibit less variability in length, suggesting that AI callers are less capable of handling complex situations. Moreover, 21% fewer borrowers promise to repay their debts and about one-third fewer repay within two hours of answering the call if they talk to AI callers. Conditional on making a promise to repay, borrowers are less likely to keep such a promise when it is made to an AI. Therefore, AI callers appear to be worse than humans at extracting promises that feel binding. This may be because machines are not thought to be owed moral duties (Melo, Marsella, and Gratch, 2016; Petisca et al., 2020). To the extent that promises to AI do have moral force, it is less unpleasant to be commit a moral transgression in the presence of a machine than a human (LaMothe and Bobek, 2020; Kim et al., 2023).

We next consider how AI affects the productivity of human callers. During our sample period, the AI experienced five upgrades, mainly improving its speech recognition and understanding. Each upgrade was rolled out so that two consecutive versions of AI callers were used simultaneously in the same month and assigned cases at random. This arrangement allows us to measure improvements in AI productivity and their impact on human callers' performance. We observe that the AI most significantly improved between August and October 2021, increasing NPV collected through day 5 by 3 percentage points. However, using a better AI for the first five days leaves more difficult cases to day 6, so human callers on day 6 collect 3 percentage points less, resulting in similar cumulative collected NPVs. This finding is consistent with a displacement effect of AI on labor. The point estimates suggest that declines in human productivity are larger for more skilled callers, but this gradient is not statistically significant.

Finally, we address how much labor costs are saved by AI adoption. We focus on direct labor costs, i.e., workers' salaries, which consist of a fixed component and a variable component. Although the productivity deficit of AI is diminished once labor costs are accounted for, AI remains less cost-effective than human callers except for the smallest debts. Importantly, this calculation does not consider indirect labor costs, such as recruitment, training, management, pension funds, etc., nor the cost of developing the AI software. AI is relatively more effective at collecting from borrowers with high credit scores, but having human callers specialize on more challenging cases is not necessarily a winning strategy. We find that human callers who are, by chance, given 2 percentage points more low-credit-score borrowers this month are 3.7 percentage points more likely to quit within the next three months.

Our paper is related to prior work on negative human reactions to AI. Luo et al. (2019) find that AI callers can be as effective as proficient human callers at persuading borrowers to take up an attractive loan renewal offer, but AI's success rate falls by 80% if it begins the sales conversation by disclosing that it is an AI. Customers rate the AI caller to be less knowledgeable and empathetic when the disclosure occurs up front. Our work is distinct from Luo et al. (2019) in that it studies AI's ability to persuade humans to undertake an *unappealing* action and examines the long-term reaction to AI and its spillover to other choices instead of only a single response that comes within one minute. Silva, Khera, and Schwamm (2024) report that editorial board members are unable to identify which persuasive essays in their field of expertise are authored by AI versus humans, but they judge an essay to be lower-quality if they believe it to be AI-generated. Glikson and Woolley (2020) survey the literature on the determinants of human trust in AI.

Our paper is also related to the literature on the impacts of automation on labor. Previous studies find different impacts in different waves of automation.[3] They mostly find complementarity between humans and AI when AI only provides predictions and suggestions and human workers make the final decision (Gao and Jiang, 2021; Luo et al., 2021; Brynjolfsson et al., 2023; Noy and Zhang, 2023). In contrast, the company in our study delegates all or none of a phone call to AI, since it is hard for AI to assist human callers in real time during conversations. In such a setting, we find imperfect displacement effects; AI callers can replace humans but are less productive by themselves, and they do not make humans more productive when working in tandem.

Additionally, our study contributes to the literature on the performance of AI and machine learning technology (Cao et al., 2024; Erel et al., 2021; Kleinberg et al., 2018; Agrawal et al., 2019, 2023). We focus on non-routine jobs, which were previously believed to be immune to automation (Brynjolfsson and Mitchell, 2017, Felten et al., 2020) and were rarely studied in this context until recently (Gao and Jiang, 2021; Brynjolfsson et al., 2023; Noy and Zhang, 2023).

Finally, our paper contributes to an emerging literature on delinquent debt collection, a function that directly affects many individuals around the world. Drozd and Serrano-Padial (2017) and Fedaseyeu (2020) examine how variation in debt collection effectiveness driven by information technology and regulations affects credit supply. Fedaseyeu and Hunt (2015) model how reputation concerns drive the usage of third-party debt collection. Cheng et al. (2021) study how consumers fare when taken to court by debt collectors. Wang and Zhou (2024) find that a machine learning algorithm that chooses which borrowers are called by human collectors increases repayment rates relative to when these choices are made by human collection officers. Laudenbach and Siegel (2023) address the importance of personal communication in collecting loan repayments. They show that phone calls to late borrowers from bank agents are more effective than mail reminders, and bank agents with more likeable voices are especially effective.[4]

The remainder of the paper proceeds as follows. Section 1 provides institutional background about the company, its debt collection process, and its human and AI callers. Section 2 describes our data, and Section 3 specifies our experimental setups. Section 4 estimates the performance gap

[3] In the early AI and information technology revolutions, some researchers find displacement effects for low-skilled workers and increased demands for high-skilled workers (Acemoglu and Restrepo, 2020, 2022). Others find that new automation technologies are labor-augmenting (Michaels et al., 2014; Tan and Netessine, 2020).
[4] Karlan et al. (2015) also document the importance of a personal touch in text messaging in debt collection. They find that only messages with the names of the bank managers who previously serviced the clients can improve repayment.

between AI and human callers. Section 5 explores the potential sources of the performance gap. Section 6 examines the interaction between AI and human labor, and Section 7 concludes.

1 Institutional Background

1.1 The company and its lending business

The company is a leading online consumer finance service provider in China. At the end of 2022, the company had around 10 million active users with nearly 7 billion yuan (980 million USD) of outstanding loan balances. The company's main business is to originate loans to online consumers. The company targets young consumers with a short credit history but large income and consumption growth potential.[5] It operates its own online shopping platform and collaborates with third-party online retailers to offer loans at the point of sale.

The 10th percentile loan size is only 8 yuan (1 USD) and the 90th percentile is around 5,500 yuan (770 USD). The company provides two types of loans. The first is an installment loan, which the consumer repays in equal monthly installments over the next six months to three years. The second is a credit-card-like product. Consumers may apply for a credit line, which is around 7,500 yuan (1,050 USD) on average, and pay for their online order with it. "Credit card" loans have a default maturity of one month, but the borrower may choose a longer maturity up to a limit determined by the company's risk assessment of the borrower, in which case monthly payments of equal size are due during the loan's life. "Credit cards" are typically used for small payments, while installment loans are preferred for expensive purchases and durable goods, similar to the setting studied by Aydin (2022).[6] Since the company's customers are typically riskier than the population average, the interest rates are mostly 24% per annum, which is the upper limit allowed by Chinese regulators.

Each borrower is assigned a monthly repayment due date (e.g., the 8th day of each month), which may be changed by the borrower with the company's approval. Changing the due date frequently is not allowed. Borrowers who fail to pay enter the debt collection process. During delinquency, extra interest and late fees accrue on the overdue amount. Borrowers need to repay

[5] 70% of the company's customers are less than 30 years old, 65% are members of the urban working population, and 13% of them have a bachelor's degree or more. These percentages are much higher than the population-wide averages.
[6] Extremely small loans are mainly generated by "credit card" purchases. The company views each purchase, no matter how small it may be, as a standalone loan. If we aggregate all newly originated loans from both products to the borrower-month level, the monthly borrowing amount of a customer ranges from 500 yuan (70 USD) at the 10th percentile to 14,000 yuan (2,000 USD) at the 90th percentile.

the overdue amount, accrued interest, and late fees to fully resolve their delinquency.[7] Borrowers who remain delinquent for ninety days are considered to have defaulted and are reported to third-party credit report aggregators. Defaulted borrowers cannot borrow from the company again and may experience difficulty borrowing from other consumer finance companies. Defaulting may also impair the borrower's ability to make purchases even without borrowing.[8]

1.2 Human debt collection process

The company treats the first day past due as a grace period. It generally does not call delinquent borrowers on that day and just sends them reminders through text messages and phone app notifications. If the debt remains unpaid on the second day past due, the company starts calling.

The company uses different strategies depending on whether the debt is 2-10, 11-25, 26-59, 60-84, or 85+ days past due. During days 2-10, cases assigned to human callers are rotated among callers at a daily frequency. There are three blocks of time when borrowers who have not repaid or verbally promised to repay are called automatically by the system: 9 to 9:30 A.M., 3 to 3:30 P.M., and 7:15 to 7:35 P.M. Callers sit in front of a computer screen and wait for a call to be answered.[9]

Outside these three automatic call time blocks, human callers can choose which borrowers to call based on debt characteristics shown on their screen during their working hours from 9 A.M. to 8 P.M. According to the company's internal research, the filter used most by productive callers is the most recent time that the borrower logged into the app. The research also suggests that case selection skills play a minor role in explaining human callers' performance. To prevent the company's phone numbers from being blacklisted by borrowers, the company uses multiple phone numbers, and the caller can choose which one will be displayed on borrowers' caller ID.

---

[7] Extra interest accrues on the overdue amount at the same rate as the original loan. Late fees accrue on the overdue amount at 0.1% a day (or 36.5% annually). The company may forgive these charges in part or in full for borrowers over 10 days past due as an incentive to repay. Total undiscounted late charges paid by the mean (median) delinquent borrower are 0.7% (0.2%) of the initial overdue amount. A few severely delinquent borrowers paid 10% to 30% of their initial overdue amount in late charges.

[8] For example, defaulting borrowers may not use rideshares or book hotels without deposits, since some large companies use credit records for screening. If the lender can prove that the borrower is not repaying despite having enough money, it can sue the borrower. If the lawsuit is supported by the court but the borrower still refuses to repay, the borrower will be added to a blacklist of "dishonest judgment debtors" assembled by the Supreme People's Court of the People's Republic of China and prohibited from expensive consumption such as traveling by plane and purchasing real estate and luxury cars.

[9] During this procedure, a worker may receive phone calls to borrowers who were not assigned to them at the beginning of the day. Once they receive the call, the corresponding case is transferred to their own list. Our data record the actual worker who talks to the borrower. The assignment of answered calls is random across workers.

During a phone call, callers usually provide information about the loans, inform the borrower of the potential negative consequences of delinquency, and try to persuade them to repay the debts as soon as possible. Callers are provided some conversation templates but are not asked to follow them strictly. Callers may provide suggestions to borrowers, such as encouraging them to ask family members for help. These suggestions are typically short and generic in days 2-10. In later stages, the conversations are more personalized and specific.

After each phone call, the caller is required to label its outcome. If calls to a phone number are not answered multiple times, it may be labeled as potentially invalid. The answering rate is only 24%. If in a conversation, the borrower clearly and explicitly states that they will repay the amount due no later than the end of the next day, the caller will label it as a "promise to pay," and the case will be kept by the same caller for one more day. Repayment will be credited to the caller as long as it comes within the promised time; otherwise, the case will be assigned to another caller. Therefore, callers have an incentive to ask borrowers to make promises. The company uses an AI conversation examiner to make sure that false labels of "promise to pay" are penalized by salary deductions.

At later stages, the company uses different debt collection strategies. A caller will handle a debt in days 11-25 for one week before it is assigned to another person. For debts in days 26-84, this interval is typically two weeks. Beyond 85 days, we do not have detailed information about strategies, since most cases are handled by third-party agencies. Around 60% of the overdue debts are outsourced to third-party debt collection agencies during days 26-59, and around 80% are outsourced during days 60-84. The company sends almost all 85+ day late cases to third-party debt collection agencies, only keeping small cases to be collected by AI callers (see Section 3) and some very large ones for further actions like lawsuits. Borrowers are typically contacted less intensely as the repayment gets further past due (see Online Appendix Figure C1).

1.3 AI caller

To cope with the high volume of cases and to reduce labor costs, the company introduced AI callers in 2018. The rules determining whether a borrower is assigned to AI will be explained in Section 3. The AI callers understand borrowers' speech and generate appropriate answers, speaking with a synthetic voice which is easily recognized as not human. The AI can provide basic information about the overdue loans, address potential negative consequences of delinquency, and respond to simple questions and explanations about the delinquency.

Table 1 illustrates the conversation process and some sample scripts that the AI caller typically uses. The conversation is divided into four stages by design. In the first stage, the AI caller greets the borrower and confirms their name. If the AI caller dialed the wrong number, it apologizes and hangs up the phone. Otherwise, the AI caller continues to Stage 2 to inform the borrower about the overdue debt. The information provided by the AI caller includes the principal amount, overdue amount, bill date, days past due, and the new due date or time. The borrower is usually asked to repay within two hours or by the end of the day. The AI caller also emphasizes potential negative consequences if the borrower fails to repay: worsening credit records, late fees, difficulties in future borrowing and consumption, and even lawsuits from the company. The AI may also mention the possibility of informing the borrower's "emergency contacts," who are typically their parents and colleagues, imposing social pressure.[10]

The AI caller then waits for the borrower's responses and sees if they have any further questions. The software classifies possible responses into five broad categories. In Case A, the borrower agrees to repay today or asks for an extension. The AI caller will then confirm the new due time with the borrower, tell them that their promise has been recorded, and ask them to keep their word. In Case B, the borrower is unable to repay the debt and may explain their difficulties—for example, they do not have enough money at hand, or they are too busy to deal with the debt. The AI caller can understand these explanations and reply accordingly. For example, for liquidity problems, the AI caller may ask the borrower to borrow money from their family or friends. In contrast, to busy borrowers, the AI caller may say that it understands that they are busy but reminds them of the negative consequences of default. In Cases C and D, the borrower claims that they do not have any debt with the company, or they have already repaid or have set up auto-payment. The AI caller will then ask the borrower to recall their borrowing history and to double-check their accounts or auto-payment settings. In addition, the AI caller can answer inquiries about basic information about the debts, such as late fees (Case E).

[10] Each borrower is asked to list two "emergency contacts" as their "guarantors" at the time of registration. These "emergency contacts" do not have any legal obligation to repay the loans if the borrower defaults. The company uses these "emergency contacts" if the borrower defaults and refuses to talk to the debt collectors. Some emergency contacts maybe willing to repay the borrower's debt. The company usually does not call the "emergency contacts" until day 5 of delinquency. After that, about 30% of calls are made to these contacts. The fraction rises to 50% after one month of delinquency but then gradually falls to 20% as the duration of delinquency extends. Calling emergency contacts is often viewed as an escalation of the debt collection actions and is tightly restricted by the regulator.

Finally, when the borrower has no more questions about their loans, the AI caller will conclude the conversation by reiterating the negative impacts of delinquency and asking the borrower to contact customer service for further information. Similar closing words will also be used to end the conversation when the AI cannot recognize the borrower's responses (due to a long silence, loud noises, strong accents, etc.) or when the borrower's responses cannot be classified into one of the five pre-specified cases (for example, the borrower yells at the caller or complains about the annoying phone calls).

2 Data Description

Our data provide us with comprehensive information about the debt collection process in the company between April 2021 and December 2023. To ensure that we can track each delinquent debt for at least one year, we restrict our analysis to cases entering collection before December 2022, which gives us more than 22 million cases. Consistent with the company's practice, multiple debts of an individual borrower are merged into one entry during collection.

We have loan and borrower characteristics for all delinquent debts, including loan size at delinquency, borrower internal credit score, age, gender, and education level. The company uses two different measures of loan size: the overdue payment amount and the remaining principal. The internal credit score is based on the probability of default estimated by the company. The company divides all delinquent borrowers into deciles and assigns them an integer score from 1 to 10, where 1 is the highest decile of default probability. This score is updated daily, incorporating the phone call outcomes of the previous day and the daily loan sizes. Education levels are self-reported, although the company can verify some of them if borrowers have uploaded their degree certificates and transcripts when registering their accounts.

We also have daily records of debt collection status and repayment actions. We know the number of days overdue, the caller handling the loan, whether the borrower has promised to repay, and how much the borrower repays each day. We also have information about callers' efforts to contact borrowers, including the number of phone calls they make, the number of phone calls answered by the borrower, and the total duration of the calls.

Finally, we have data about callers' demographic information, monthly performance, and compensation. Callers' demographic information includes their age, gender, city of birth, and whether they are in-house or with a third-party collection agency. For in-house callers, we have their job titles and their tenure (in months) with the company. Performance measures include the total amount of money collected, monthly target collection amount, performance ranking, and the ratio of the actual amount collected by the caller to her target. We know the salary amount each caller received, as well as the portion that is performance-based and the amount that is deducted due to penalties.

Table 2 Panel A reports summary statistics for the loan and borrower characteristics in the full sample. The characteristics are measured on day 2 past due, the first day when cases enter the debt collection process. The average delinquent debt has an overdue amount of 1,128 yuan (160 USD) and a remaining principal of 6,474 yuan (910 USD), which are larger than the corresponding moments for the population of outstanding loans, as larger loans are more likely to default. The medians are smaller than the means: the median overdue amount is 653 yuan (92 USD), and the median remaining principal is 4,248 yuan (600 USD). The average internal credit score is around 5. Among delinquent borrowers, 70% are males, 13% have a bachelor's degree or more, and the average age is 27 years.

Case sizes are heavily right-skewed: the maximum remaining principal is 1 million yuan (about 140,000 USD). Extremely large debts are typically nonstandard contracts with specific customers for special purposes. They are treated separately by the company, so we want to exclude them from our analysis. Since separately treated cases are not labeled in our data, we exclude cases with remaining principal above the 99th percentile. The left tail does not require trimming since extremely small cases are excluded in our experimental design, as discussed in the next section.

3 Experimental Setup

To identify the productivity difference between AI and human callers, we utilize the company's rules for assigning cases between AI and human callers. Figure 1 illustrates the assignment procedure. First-time delinquent borrowers are always assigned to human callers. Starting with the second delinquency, borrowers can be assigned to either AI or human callers.

The company initially allocates all cases with overdue amounts no greater than 20 yuan or remaining principal no greater than 300 yuan to AI callers. In rare situations, which we discuss later in this section, these small cases are assigned to human callers after day 25.

Larger cases are either unconditionally or conditionally randomly assigned to AI or human callers. The company randomly selects 10% of these larger second-delinquency cases every month for testing and monitoring purposes."[11] In this subsample, a random half of cases are assigned to human callers on day 2, while the other half are assigned to AI callers on days 2 to 5 before being reallocated to human callers on day 6 onwards. Once a given delinquency is handled by a human caller, it typically will not be given back to an AI caller.

For subsequent delinquencies, the borrower's assignment to be initially called by a human or an AI remains the same as it was for his second delinquency unless the subsequent delinquency is small (overdue amount $\leq$ 20 yuan or remaining principal $\leq$ 300 yuan), in which case it always gets assigned to AI. Therefore, only the assignment in the second delinquency can be viewed as orthogonal to potential outcomes within our sample; the type of person who reappears in our data as delinquent a third time after always being called by a human might be different on average from the type who becomes delinquent a third time after being called by AI. Thus, for larger cases, our analyses focus only on borrowers in their second delinquency, which are about 11% of the full sample. We call this subsample of second delinquencies the "completely randomized subsample."

The remaining 90% of larger second-delinquency cases are assigned between AI and human callers randomly conditional on case characteristics; that is, the probability of a case being assigned to AI varies by its characteristics, as does the timing of when humans take over from AI (anywhere from day 2 to day 6). AI treatment effect estimates within this conditionally randomized subsample are similar to those in the completely randomized subsample, so we do not report them.

Whereas the completely randomized subsample allows us to identify the effect of replacing human callers with AI callers from days 2 to 5, the discontinuity in the company's assignment rule for small cases creates an opportunity to use a regression discontinuity (RD) design to identify the local treatment effect of replacing humans with AI for a much longer time. The 20-yuan overdue amount threshold is extremely small—almost at the 1st percentile of the full-sample distribution. The 300-yuan remaining principal threshold is somewhat less extreme—at around the 5th percentile of the full-sample distribution. Therefore, in the RD analysis, we exclude cases with less than 20 yuan of overdue payments and apply the standard RD methodology with one running variable, the remaining principal.

[11] Larger cases that enter the debt collection process on the last few days of each calendar month are always assigned to human callers because there are fewer cases initiated at the end of each calendar month. These cases are excluded from our analyses.

Figure 2 Panel (a) plots the fraction of cases assigned to AI as a function of remaining principal on day 2. Consistent with the stated assignment rules, cases below 300 yuan of remaining principal are all assigned to AI callers, while only about 80% of cases above the cut-off are assigned to AI. The discontinuity in the AI fraction is sharp.

Figure 2 Panel (b) shows the fraction of cases assigned to AI callers on either side of the threshold from days 1 to 25. The fractions for "Under 300" are calculated based on cases in the (295, 300] yuan interval, while the fractions for "Above 300" are calculated based on cases in the (300, 305] yuan interval. Small cases are all handled by AI callers in the first 25 days. In contrast, on days 2-3, only 80% of the larger cases are assigned to AI callers. The fraction falls to around 60% on days 4-5. From day 6 onwards, all larger cases are handled by human callers. Panel (c) extends the horizon to day 360. Cases above 300 yuan remain under human treatment for the entire extended period. For cases below 300 yuan, a small fraction of them are assigned to human callers after day 25, mainly due to the introduction of third-party collection agencies. When the company delegates to a third-party agency, it randomly selects some cases, maybe conditional on some loan characteristics, and assigns them to the agency. The assignment of some small cases to humans biases against finding significant collection differences across the remaining principal threshold.

4 AI versus Human Caller Performance

4.1 Measure of debt collection productivity: Net present value of collected cash flows

We use the net present value (NPV) of cash flows collected from day 2 past due onwards as the measure of caller productivity. For each case, we calculate how much money is paid towards the initial delinquent payment on each day, including late fees. We then discount these cash flows to day 2 using a 24% per annum (24/365 = 0.066% per day) discount rate, which is close to the average APR of the loans originated by the company. It is also the maximum legal APR allowed by Chinese regulators. It can be viewed as the opportunity cost of uncollected money, which could have been lent to other borrowers and generated interest at a 24% APR if it were collected on time.[12] Finally, the NPV is scaled by the amount overdue on day 2. The computed NPV can grow beyond 1, since borrowers may pay late fees (see footnote 7).

[12] Setting the discount rate to zero has little impact on the results, since most payments are collected in the early days of delinquency.

4.2 Small cases subsample: Regression discontinuity design

In this subsection, we compare the productivity of AI callers to human callers by utilizing the discontinuity in the company's AI deployment strategy at the 300-yuan cutoff in remaining principal.

Table 2 Panel B reports summary statistics for loan and borrower characteristics in our subsample for the RD design: cases with remaining principal between 100 and 500 yuan, which gives us over 1 million cases. Although loan sizes are much smaller than in the full sample, as expected, the gender composition, average age, and the fraction of borrowers with a bachelor's degree or more are all close to those in the full sample. The internal credit score is somewhat lower in the RD sample than in the full sample but is still very close to 5.

Table 3 Panel A reports continuity tests on five predetermined loan and borrower characteristics around the 300-yuan remaining principal threshold. Local linear regressions with uniform kernels over the coverage error rate (CER)-optimal bandwidths are used in the estimation, and RD bias-corrected z-statistics estimated by local quadratic regressions are reported.[13] The average loan characteristics on the two sides of the cutoff (columns 2 and 3) are quite similar, and their differences are small (column 4) and not statistically distinguishable from zero (columns 5-7). Figure 3 Panel (a) shows these tests graphically.[14] In Online Appendix A, we check if there is manipulation around the cutoff by examining the density of observations. Although the remaining principal amount has some tendency to cluster at 300 yuan (and also at 200 and 400 yuan), the density functions can be considered continuous at the threshold.

Given the validity of our RD design, we compare the average collected NPV difference on each side of the 300-yuan cutoff, which gives the treatment effect of AI callers on debt collection productivity. Table 3 Panel B presents estimates of the NPV differences at various horizons. Since we are now interested in point estimates of the productivity gap, the mean squared error (MSE)-optimal bandwidths are used in the regressions.

The differences between the mean NPV estimated from the left (AI) and the right (Human) are all negative and significant at the 1% level, regardless of the evaluation horizon. These gaps

---

[13] As suggested by Cattaneo et al. (2019), the CER-optimal bandwidth is used because, for testing the null hypothesis of continuity, we are interested in inference (the confidence interval) instead of point estimates.

[14] The number of bins is set to 40 (i.e., each bin is 5 yuan wide) on either side of the cutoff, which is close to the Integrated Mean Squared Error (IMSE)-optimal number of bins of around 44. The IMSE-optimal number of bins minimizes the IMSE of local mean estimators. It is useful for assessing the overall shape of the function.

are also economically significant. On day 2, the NPV gap of 0.04 is a 19% productivity loss relative to the human mean NPV of 0.21. The gap grows to 0.09 by day 30 before starting to shrink because human callers are not able to collect much more beyond day 30, whereas AI continues to make some significant collection progress. Nonetheless, even after 360 days, AI's productivity loss relative to humans remains large: 5%.[15] Figure 3 Panel (b) presents the RD plots of collected NPVs through various horizons after the initial due date. Figure 4 shows the productivity gap over time graphically. Re-estimating the treatment effects while controlling for the five loan and borrower characteristics in Table 3 Panel A has little impact on the estimates (Table 3 Panel B column 8). Using undiscounted cashflows as the outcome measure makes little difference (Online Appendix Figure C2).

Next, we examine how AI's performance gap varies with borrowers' credit quality in Figure 5, using the same specification as in Figure 4. Low, medium, and high groups refer to internal credit scores of 1-3, 4-7, and 8-10. AI initially underperforms more with high-score borrowers. However, the gap between AI and humans shrinks quickly for the high group, approaching -0.03 in the long run. In contrast, the performance gaps for lower-score borrowers keep expanding until around 30 days. The magnitudes of the long-run productivity gaps are monotonically decreasing in credit scores. High-score borrowers may mainly need reminders, which AI callers can provide adequately, while low-score borrowers likely find repaying more costly and may therefore require greater persuasion that AI callers are less capable of performing.

4.3 Completely randomized subsample

The previous subsection shows that AI callers alone are less productive than human callers alone. The company usually has AI callers supplement human caller efforts in the early stages of delinquency. Specifically, some cases are assigned to AI callers at first. If the delinquency has not been cured soon, human callers take over.

To identify the performance of the "AI + Human" strategy, we utilize the 10% completely randomized subsample. In this subsample, the company randomly selects half of the cases and

[15] Recall that many cases above 300 yuan are allocated to AI callers before day 6. In Appendix **Table C1**, we restrict cases on the right of the cutoff to the arm of the 10% completely randomized subsample that is only called by humans. Relative to the estimates in Table 3, the treatment effect on day 2 is almost double and the one-year gap of -0.06 is also larger.

assigns them to human callers from day 2 onwards (the control group), while the remainder are assigned to AI callers on day 2 and reallocated to human callers from day 6 onwards (the treatment group). Table 2 Panel C shows summary statistics on the completely randomized subsample. Since small cases with remaining principal no greater than 300 yuan are excluded from this subsample, the overdue amount and remaining principal here are on average larger than in the full sample. Other borrower characteristics are similar to the full sample.

As a first step, we validate that the treatment and control groups are comparable to each other. We regress predetermined loan and borrower characteristics onto a treatment group indicator and calendar month of delinquency dummies. Table 4 Panel A shows that the coefficient on the treatment group indicator is insignificant when the dependent variable is overdue amount, remaining principal, internal credit score, gender, age, and education level. This shows that the two groups are statistically indistinguishable from each other ex ante.

We then estimate the productivity gap between "AI + Human" and the always-human control using the same regression specification, whose results are shown in Table 4 Panel B. Columns 2 and 3 show the average cumulative NPV collected of the treated (AI) and control (Human) groups, respectively. Column 4 reports the difference (AI minus Human), and the next column reports the t-statistic of this difference.

For all evaluation horizons, the "AI + Human" treatment group significantly underperforms the always-human control group. The gap is 0.09 on day 2, the first day of contact, which corresponds to a 33% productivity loss relative to the always-human control, and expands to 0.12 on day 5, a 24% productivity loss. Once human callers take over after day 5, the performance of the two groups converges quickly, so that the NPV difference is only 0.02 on day 10. Nevertheless, the "AI + Human" group never repays as much as the control group; even after a year, the gap is 0.007. On the one hand, this is only about a 1% relative productivity loss. On the other hand, it is remarkable that only five days of exposure to AI callers permanently impairs the company's ability to collect.

The last column re-estimates the differences with loan observations weighted by their initial overdue payment amount on day 2. These value-weighted differences, which represent the total monetary losses from AI treatment, are larger in the early days but similar in the long run. Figure 6a presents these results in graphical form.[16]

Figure 6b and Figure 6c plot the NPV gap over time by internal credit score and loan size, respectively. The low-score cases suffer the least productivity loss from AI initially, which is similar to what we saw in the RD analysis. We also learned from the RD design that the NPV gap of low-score cases would keep growing and exceed the gaps of the other two groups if AI callers continued working on them. In the completely randomized subsample, however, human callers intervene on day 6, halting the damage. Therefore, low-score cases also have the least performance damage over longer horizons. High-score cases experience the largest permanent damage from initial AI contact. On the loan size dimension, cases with larger overdue amounts generally have larger performance gaps initially, consistent with our expectation that they require more persuasion that AI is less able to handle. In the long run, larger loans are more damaged by initial AI contact than smaller loans, but mid-sized loans with overdue amounts between 800 and 1500 yuan are the most damaged.

5 Understanding AI's underperformance

5.1 Underperformance on day 2

Why does AI underperform humans in collecting payments? In order to gain insight into this gap, we begin by examining other outcomes of phone calls made on day 2—the duration of the call, the fraction of borrowers who promise to repay, and the fraction of promisers who make payments shortly after the call.

Table 5 Panel A reports the average outcomes of all phone calls made by AI and human callers on day 2 within the 10% completely randomized subsample. Human callers make 0.85 more phone calls per day to each borrower than AI callers and thus are answered 0.35 more times per day. But for both types of callers, the phone answering rates are 24%, since borrowers cannot tell whether a call is made by an AI or human caller until they pick it up.

To analyze differences in the ability of AI versus human callers that are separate from the frequency with which they make calls, we next restrict our sample to the first call answered by each borrower. The results are reported in Table 5 Panel B. The time of first-answered calls is on

[16] In Appendix Figure C3, we reproduce Figure 6a using the sum of undiscounted cash flows as the productivity measures and obtain similar results.

average a little earlier for human callers than for AI callers: 11:31 AM versus 11:47 AM.[17] The reason is that, as mentioned in Section 1.2, there is a half-hour automatic call period from 9 A.M. to 9:30 A.M. when all cases assigned to human callers are called once. In contrast, calls from AI callers are distributed more evenly across the day. To control for this disparity, we estimate timing-adjusted results that control for one-hour-interval time-of-call fixed effects.[18]

There is a significant 1.3 second difference between the two types of callers in how long the phone rings before it is answered, but this disparity disappears after controlling for time-of-call fixed effects. On the other hand, the duration of phone calls significantly differs whether or not time-of-call fixed effects are controlled for. The unadjusted mean duration of an AI call is only 28 seconds, which is 19 seconds less than for calls by human callers. The gap widens to 31 seconds after the timing adjustment. This finding suggests that AI callers may be able to provide only limited information, are less engaging, and are unable to handle complicated situations, leading to short conversations. Online Appendix D contains some sample AI call transcripts.

Online Appendix Figure C4 shows the histograms of phone call durations for the two types of callers separately after reweighting the calls so that the time-of-call distributions of the two caller types match. A potential interpretation of the differences in average call times described above is that borrowers hang up quickly upon realizing that an AI is calling. However, the figure shows that AI calls are almost equally likely to terminate within the first 10 seconds as human calls, which suggests that immediate hangups do not explain the difference between AI and human caller performance. AI phone call lengths are concentrated around 30 seconds, while the duration of human calls has greater variation—potentially a proxy for flexibility in response to complexity.

We additionally find that 21% fewer borrowers make a promise to repay their debts when talking to AI callers. This result must be interpreted carefully because AI callers classify a call as

[17] In the table, we convert the time of the call to a decimal number representing hours from midnight. For example, 2:15 PM is converted to 14.25.

[18] The difference in average call time of day is small, but strictly speaking, a borrower who first answers a human call at 11:00 A.M. has missed more calls than a borrower who first answers an AI call at 11:00 A.M., so controlling for time of day may not control for all unobservable borrower characteristics. In Online Appendix Table C2 Panel A and Online Appendix Table C3, we restrict the sample to the first call attempted between 9 and 9:30 A.M. The results are similar: AI callers receive fewer promises to repay and these promises are less likely to be kept. In this subsample, human calls are shorter than AI calls because human callers must dial all borrowers assigned to them during this 30-minute automatic call period, so they have to finish each conversation quickly. More borrowers in this subsample repay their debts within 15 minutes when talking to an AI than to a human. Appendix Table C2 Panel B splits the sample by the overdue payment amount at 1,000 yuan (around the median size) and finds that this 15-minute pattern is mostly driven by small borrowers, with whom human callers have especially short conversations on average.

having resulted in a promise if they believe there was a promise to pay on the *same* day as the call, whereas human callers classify a call as having resulted in a promise if there was a promise to pay no later than the end of the *following* day.[19] Nonetheless, we see in the first two columns of Table 6 that within 30 minutes of the call—that is, well before the end of the following day—borrowers who made a promise to a human are 1.5 percentage points more likely to have repaid than borrowers who made a promise to an AI. By the end of the same day the call was made, this difference is 19 percentage points. Keeping in mind the caveat that borrowers endogenously choose whether to make a promise in a manner that may be correlated (differently depending on the caller type) with their propensity to pay without having made a promise, it does not seem that humans are able to extract promises from more borrowers simply because the deadline for a payment promise to a human is later than the deadline for a promise to an AI. Relatedly, one might have expected that if AI callers are bad at extracting promises, they would disproportionately receive promises from borrowers who are likely to pay promptly anyway—the low-hanging fruit. But borrowers are *less* likely to pay after a promise made to an AI than after a promise made to a human.

It is also interesting that in the absence of a promise, the difference in collection probability between AI and human callers is muted: 0.2 percentage points higher for AI within 30 minutes of the call and 8 percentage points higher for humans by the end of the day of the call. Borrowers who make a payment promise to an AI are 13 percentage points more likely to pay by the end of the day than borrowers who speak to an AI but do not make such a promise, which is only half of the promise versus no promise difference for borrowers who speak to a human—25 percentage points. These numbers suggest that borrowers making a promise to an AI creates less of a sense of obligation than a promise to a human. Table 5 Panel C shows that integrating across both borrowers who do and do not make a promise, an answered call from an AI is 18 percentage points less likely to result in a same-day payment than an answered call from a human.

[19] To check the accuracy of the AI's classification, a native Chinese-speaking research assistant read 200 AI call transcripts, blinded to the AI's classification, and classified whether the borrower made a promise to pay by the end of the day. Out of 100 transcripts where the AI judged a promise to have been made, the research assistant labeled 84 as containing a promise (80 to pay on the same day, 4 to pay by the next day). Out of the other 100 transcripts where the AI judged no promise to have been made, the research assistant labeled 95 as containing no promise. Human callers have an incentive to exaggerate how many callers have made payment promises, so the Type I error rate for promises made to human callers is also likely to be higher than the corresponding Type II error rate.

How much does the difference in phone call frequency between AI and humans explain the performance gap between them on day 2? We provide suggestive evidence in Table 7. Because call frequency to a given borrower may depend on the borrower's behavior, we avoid reverse causality confounds by not directly examining how a borrower's repayment relates to the frequency with which she is called. Instead, we sort human callers into terciles based on their call frequency in the *previous* month. We find that callers' average call frequency in the current month is positively correlated with their previous month's frequency. Borrowers are randomly assigned to callers, so this persistence is due to caller fixed effects rather than borrower characteristics. We then identify the effect of call frequency on borrower repayment by seeing how repayment this month responds to being randomly assigned a caller who called others more versus less frequently in the prior month.[20]

We find that the number of phone calls per borrower-day, the probability of calling again on the same day that a promise is made, and the average minutes until the follow-up call after a promise (conditional on following up on the same day) do not have statistically robust relationships with payments collected on day 2. Only the average time interval between consecutive calls has significant predictive power for callers' performance on day 2. Specifically, human callers in the lowest tercile wait on average 11 minutes less between calls than callers in the highest tercile, which is associated with 1.6 percentage points more collected NPV on day 2. In unreported results, a binned scatter plot of day 2 scaled repayment against the average time between calls indicates that the relationship is close to linear. AI callers take 15 minutes longer between calls than human callers in the median tercile, so the time gap between calls only explains 2.2 percentage points of the gap between AI and the median tercile's performance. By comparison, the average productivity gap on day 2 between AI and all human callers is 8.8 percentage points.[21]

In summary, the evidence suggests that while some of the performance deficit of AI relative to human callers can be attributed to AI's lower call frequency, AI also has a lesser ability to extract

---

[20] This identification will be confounded if a caller's call frequency strategy is correlated with unobserved features of the caller that affect borrower repayment.

[21] If we take seriously the point estimates of all monotonic relationships between caller behavior and collected money, whether or not the relationship is statistically significant, and extrapolate the relationship estimated among human callers to AI, we would attribute an additional 2.8 percentage points of the AI productivity shortfall to follow-up call behavior towards borrowers who made a promise to repay. Forty-four percent of borrowers make a promise to AI, and within that subgroup, AI's lower follow-up rate accounts for $0.006/0.019 \times (0.825 - 0.878) = -1.67\%$. AI's longer time to follow-up conditional on receiving a payment promise and following up accounts for $-0.009/18.78 \times (215.70 - 94.61) = -5.80\%$. Weighting by the fraction of callers affected, we get $0.44 \times -0.0167 + 0.44 \times 0.825 \times -0.0580 = -2.80\%$.

promises to repay and create pressure to keep those promises. AI additionally appears to be less able to communicate with borrowers, engage their attention, and/or handle complex situations.

5.2 Permanent collection impairment

In Section 4.3, we documented that collection is permanently impaired for borrowers who are first contacted by AI callers instead of human callers. Although it is possible that borrowers inferred from the AI contact that the company was less serious about collecting the debt, it is hard to believe that 354 days of subsequent contact from human callers would not eliminate differences in this belief relative to borrowers who were always contacted by humans.

Table 8 shows that borrowers contacted by AI experience a more general alienation from the company; not only are they less likely to ever repay the initial delinquent payment, but they are more likely to miss each of their next 12 monthly payments. Conditional on being delinquent for a second time (and thus entering the completely randomized subsample) and having at least one more monthly payment due before their loan matures, 60% of borrowers always contacted by humans miss that next monthly payment. This probability of delinquency is 2.3 percentage points higher for borrowers first contacted by AI. Among borrowers who have at least 12 more monthly payments due, borrowers first contacted by AI are 1.4 percentage points more likely to miss the twelfth payment due than borrowers always contacted by humans. Thus, whatever damage AI is doing is not narrowly isolated to the payment it initially tried to collect.

To assess whether there is something uniquely damaging about AI, or whether *any* less-productive collection method used in the early days of delinquency also causes permanent impairment, we examine the long-run effect of lower *human* caller productivity in the initial stage of debt collection.

First, we exploit the fact that collection calls that occur on weekends are less effective. Since a debt's due date normally occurs on the same day of each month (e.g. the 11th) and cannot be changed frequently, whether a borrower is first contacted about a late payment on a weekend should be uncorrelated with borrower and debt characteristics. We confirm this orthogonality by regressing overdue payment amount, remaining principal, internal credit score, gender, age, or attainment of a bachelor's degree measured on day 2 after the due date (i.e., the first day of contact) on an indicator for if that day is a Saturday or Sunday. We also control for week fixed effects so that only variation within the same week is used for identification. The sample is the completely randomized subsample. In untabulated results, we find that no *t*-statistic on the weekend dummy has a magnitude greater than 1.29.

In Table 9, we see that borrowers who are first contacted on weekends repay 1.5 percentage points less in normalized NPV on the day of contact than borrowers first contacted on business days. However, a significant gap lasts for only three days and becomes insignificant afterward. Thus, the productivity loss from first contacting a borrower on a weekend is transient, unlike initial contact from an AI caller.

We next explore variation in caller working experience, measured by the number of months since the caller joined the company. The company assigns debts randomly among human callers every day, so some debts are handled by more experienced callers on day 2. In line with the company's operating and managing practices, we define senior callers as callers who joined the company more than four months ago. We regress variables of interest onto a senior-caller indicator with month fixed effects using debts in the completely randomized subsample. The sample is further restricted to callers who specialize in debts that are in their first five days past due. In untabulated results, we find that differences in overdue payment amount, remaining principal, internal credit score, gender, age, or attainment of a bachelor's degree measured on day 2 after the due date between borrowers assigned to junior versus senior callers have *t*-statistics whose magnitude is 1.47 or less.

Table 10 shows that junior callers collect 1.1 percentage points less normalized NPV on day 2, but the gap disappears in the following days as the debts are rotated to other callers on each day. Again, there is no permanent damage associated with a less productive initial human contact. In Panel B, we further split the group of junior workers by their working experience and find that, although underperformance is larger and persists for a few more days among workers with no more than one month of experience, the productivity gap is mostly offset in the long run.

In the above two analyses, less productive human calls occur for only one or two days, whereas AI calls for four days. To evaluate the impact of only one day of AI treatment on day 2, Table 11 utilizes a subset of the *conditionally* randomized subsample, to which 90% of sufficiently large cases are randomly assigned. In February, March, May, and June 2022, all cases in this subsample whose initial overdue amounts exceeded 400 yuan and whose internal credit scores were in deciles 4 to 7 were assigned to AI callers for only day 2 and then reallocated to human callers afterward. We use these cases as the treatment group and use contemporaneous cases with the same characteristics that were always assigned to human callers in the completely randomized subsample as the control group. For this restricted sample, the difference between human and AI callers is 8.8 percentage points on day 2, which is the same point estimate we found within the completely randomized subsample. Unlike before, as human callers take over the cases on day 3, the difference shrinks by almost a half. The gap loses statistical significance by one month of further contact.

However, after splitting the sample by internal credit score, we find that even only one day of AI contact creates lasting damage for borrowers below the median credit score. Borrowers with a credit score of 4 or 5 who are called by AI pay back 2 percentage points less NPV after one year. In contrast, the collection gap among high-score borrowers loses statistical significance after day 5.

Online Appendix Table C4 repeats the weekend effect and inexperienced worker effect analyses on the restricted subsample. It finds no significant long-run performance gap between AI and human callers.

These results indicate that there is something uniquely damaging about being contacted by an AI caller. However, the evidence is only suggestive because the less productive initial human contacts we test are still much more effective than AI callers. We cannot rule out that calls from a human who is just as ineffective as an AI caller would result in similar long-term damage.

6 Interactions between AI and Human Callers

6.1 Impact of AI upgrades on human callers

The AI caller software was upgraded several times during our sample period, which gives us an opportunity to examine how improvement in AI productivity affects human callers. This question is particularly important in light of the current rapid development of AI technology.

Figure 7 illustrates the AI upgrade process by showing the fractions of cases assigned to different versions of AI callers each month in our sample period. We study six versions of AI callers. We call the first version in our sample period "V1," although this is not the very first version of the AI deployed by the company in 2018. Subsequent versions are labeled "V2" to "V6," according to their order of introduction.

As Figure 7 shows, the company introduced new versions of AI callers progressively. V1 was the predominant version used at the beginning of the sample period in April 2021, but the company was already simultaneously deploying V2. The first cases were assigned to V3 in May 2021, and V2 was phased out after June 2021. V4 was introduced in September 2021 and took over all cases starting in November 2021. The company began testing V5 in January 2022 and retired V4 after February 2022. The final version in our sample period, V6, began testing in June 2022 and took over all cases starting in October 2022.

Since assignment to different active versions of AI callers was random within each time period, measuring their relative productivity is straightforward. In the following analyses, we restrict our sample to the completely randomized subsample. Figure 8 shows the time series of average collected NPVs on day 2 or over the first five days past due for each AI caller version, along with the corresponding collections of human callers. The gap between AI and human callers remains wide over time, despite the multiple AI upgrades.

Table 12 formally tests the improvement of AI performance, comparing cumulative NPV collected over two- to ten-day horizons between pairs of contemporaneously active AI callers. Recall that in the completely randomized subsample, the AI caller only works from days 2-5. For each NPV horizon, the test is implemented by regressing collected NPV onto a set of indicators for which AI version was assigned the case and calendar month dummies. The results show that the most salient improvement occurred when upgrading from V1 to V4. The increase in collected NPV is 0.030 over the first five days, corresponding to an 8% increase in AI productivity relative to the sample-wide average NPV in the last row. On day 6, however, human callers take over, immediately closing the cumulative gap to a statistically insignificant 0.012 between borrowers called by V4 and V1. This means that human callers taking over from V4 collect 0.030 – 0.012 = 0.019 less NPV on day 6 than humans taking over from V1.[22]

The day 6 finding suggests that the more effective AI callers are in the first five days, the tougher the cases that are left for human callers on day 6, and the less human callers can collect. Across the five AI upgrades, there is no statistically robust evidence that the total amount collected

[22] The company says that the AI improvements were concentrated in the speech recognition and language understanding algorithms. Appendix
Table **C5** compares several phone call outcomes across different versions of AI. From V1 to V4, the average phone call duration falls, but there are also fewer calls shorter than 10 seconds. Given that most long AI calls consist of repetitious conversations, this result indicates a potential increase in the information efficiency of the conversations. In addition, V4 is better at extracting promises to repay from borrowers. Finally, the V4 has fewer calls classified as ineffective, which includes calls with only silence after picking up, voicemail or auto-reply voice messages, and conversations that the AI cannot understand or cannot classify into pre-specified scenarios.

by humans partnering with AI expands. This result is consistent with a displacement effect of AI callers.

Table 13 further examines the displacement effects of the AI upgrade from V1 to V4. The sample is restricted to the completely randomized subsample treated by AI callers in the first five days during September and October 2021, when V1 and V4 coexist. We study how human callers perform after receiving these AI-treated cases. We restrict the sample of human callers to those specializing in day 2-10 cases, who are the major group of callers working on day-6 cases and within whom the cases are randomly assigned. The variable of interest is the human caller's performance on day 6, which is measured as the increase in collected NPVs from days 5 to 6, i.e., ΔNPV6. Hence, we focus on the cases that remain unpaid at the end of day 5 past due.

In Column 1, we regress human callers' day-6 performance, ΔNPV6, onto an indicator that the cases are treated by AI V4 instead of V1 in the first five days. The estimated coefficient on the V4 indicator is -0.04, suggesting that callers perform worse on cases previously treated by the better version of AI due to the increase in task difficulty.

Column 3 regresses ΔNPV6 onto human callers' ability, which is proxied by their historical performance on day-6 cases (average ΔNPV6) in the full sample during August 2021, before V4 was introduced. We split human callers into terciles based on their historical performance and use tercile dummies in the regressions. The significant positive coefficients on the dummies indicate persistence in callers' day-6 performance, which can be viewed as indicative of callers' ability. In Online Appendix Table C6, we perform balance tests to confirm that case assignments among callers are random: observable loan characteristics are orthogonal to callers' previous performance.

Column 5 interacts the AI V4 indicator with callers' previous performance tercile dummies to see how human callers' productivity loss varies with their ability. We find that, although the coefficients on the interaction terms are far from significant, the point estimates hint that better callers tend to be more negatively affected by the improvements of AI callers. This would be the case if the AI caller is improved by learning from and mimicking the skills of the best human callers.

6.2 Impact of increased task difficulty and the extent of AI application

We saw in Sections 4.2 and 4.3 that AI callers' underperformance is especially large when trying to collect from low-credit-score borrowers. A natural managerial response could be to have AI specialize in high-credit-score borrowers and humans specialize in low-credit-score borrowers.

On the other hand, a rise in task difficulty for humans could make the job more unpleasant, damaging morale, increasing job turnover, and raising the required compensation to retain workers. These are indirect costs of AI adoption that a company needs to take into consideration. We explore this point in this subsection by exploiting ex-post imbalances in case difficulty among callers.

Debts are randomly assigned among human callers, so random variation causes some callers to receive more debts that are hard to collect than other callers. For each caller in each month, we measure this ex-post imbalance by the fraction of debts with internal credit scores of 3 or lower that are assigned to the caller. The larger the imbalance measure, the more difficult debts the caller is assigned. The sample is restricted to callers who specialize in debts in the first five days past due. We also require the caller to work for at least 20 days in a month to be included in the sample, so that callers with extreme imbalances are not disproportionately callers with fewer working days and thus fewer assigned debts.[23]

We regress callers' monthly performance and compensation onto the imbalance measure with and without caller characteristic controls (working status classification as junior, senior, or returning callers who previously left the company; age; gender; and working experience in months). All specifications include month fixed effects. Table 14 reports the coefficients on the imbalance measure. Based on the results from the specifications without caller characteristic controls, a caller whose imbalance is at the 95th percentile and thus receives 2 percentage points more difficult debts in a month relative to the sample-wide average that month achieves a repayment rate[24] that is 0.46 percentage points lower, and her relative performance ranking is 10 percentage points lower. Worse performance translates into 477 yuan (67 USD) less total salary, which is equivalent to around 10% of the average compensation. Finally, we examine the caller's tendency to remain at the company at the end of the current month or at the end of the next three months with the same specification estimated with logit regressions. Average marginal effects[25] are reported in the last rows. An increase in imbalance of 2 percentage points causes a 3.7 percentage point higher chance of leaving the company within the next three months.

[23] The typical number of working days of callers is 25 days a month, or equivalently, about six days a week.

[24] Repayment rate is a metric the company uses to determine its performance-based compensation. It is the undiscounted sum of money collected in a month divided by the sum of overdue balances assigned to the caller that month.

[25] We calculate a marginal effect for each caller at his covariate values, and report the average of these caller-level marginal effects.

Thus, if AI is found to be cost-effective on tasks that humans find relatively pleasant, and cost-ineffective on tasks that humans find unpleasant, it may decrease profits to have AI specialize in the former and humans specialize in the latter.

6.3 AI productivity net of labor cost savings

AI callers perform significantly worse than human callers. On the other hand, AI callers have almost zero marginal costs when making phone calls. Therefore, to fully evaluate how well AI may replace human callers, we need to subtract labor costs in our NPV calculations.

Aggregate caller salary consists of two parts. One part is fixed, depending only on the total number of callers and unrelated to how much money they collect. The other part is variable, a function of the total amount of money that callers collect. Online Appendix C Section 1 provides more information about the salary scheme. Although there is nonlinearity and variation in salary schemes across callers in different stages of the debt collection process, for a simple back-of-envelope calculation, we assume constant rates for the fixed and variable components of labor costs. Following the procedure described in Online Appendix C Section 2, we estimate that to employ callers to talk to one delinquent borrower for one minute, the fixed cost is 1.1565 yuan (0.16 USD). The per-minute fixed cost rate is converted to the borrower-day level by multiplying it by the average phone call length per borrower on the corresponding day after delinquency. In addition, for every yuan collected, the variable cost is 0.0051 yuan (0.00072 USD).[26] In the NPV calculation, we subtract out labor costs on the same day that the associated collection effort occurs.

Figure 9 shows the average differences in collected NPV between AI and human callers as a function of days past due after adjusting for caller salary. Panel (a) uses the RD subsample, and Panels (b) and (c) use the completely randomized subsample equally weighted and weighted by the initial overdue payment amount on day 2, respectively. The estimation methods are the same as what we use for the corresponding subsamples in Figure 4 and Figure 6. For comparison, we also show the differences in unadjusted NPV.

---

[26] A back-of-envelope calculation shows that the direct labor cost for debt collection is comparable to the average labor cost in China. An average caller's daily fixed cost is 78 yuan: 130 assigned borrowers per day × 1 answered call per borrower-day × 0.52 minutes per call × 1.1565 yuan of salary per minute phone call. The daily average variable cost is 124 yuan, which equals 24,400 yuan of money collected per day × 0.0051 yuan salary for every yuan collected. Therefore, the total daily direct labor cost for an average caller is 202 yuan, with more than 60% from the variable costs. The national average disposable income of urban residents in China is 49,283 yuan per year, or equivalently, 197 yuan daily (assuming 250 working days).

After accounting for direct labor costs, the collected NPV gaps between AI and human callers become narrower or even flip sign. For small cases with less than 300 yuan of remaining principal, the difference between AI and humans becomes positive after about three months of collection effort. Because calling one borrower incurs the same labor cost regardless of the debt's size, the return to human labor is relatively low on smaller cases. In the completely randomized subsample, where the case sizes are larger, the adjusted productivity gaps become smaller, but AI remains significantly less cost-effective than humans even in the long run when cases are value-weighted. This suggests that AI callers are cost-effective on small cases but the benefits do not outweigh the damages on larger cases.

Importantly, we only have considered direct labor costs, i.e., salary paid to human callers. To hire and manage more than 2,000 callers, the company also needs to spend money on many indirect costs, such as worker recruitment, training, management, pension funds, etc. On the other hand, we also do not include in our calculation the cost of developing and improving the AI software. The cost-effectiveness of AI will also depend on the prevailing wages in the local labor market.

7 Conclusion

An important task in many service and managerial jobs is to persuade another human to perform a personally costly action. In this paper, we study how well AI does in one instance of such a task: persuading borrowers to pay their delinquent debt. We find that currently, AI is significantly worse than humans at getting borrowers to repay. AI appears to have a lesser ability to extract verbal promises to repay and create a sense of obligation to keep those promises. AI additionally seems less able to communicate with borrowers, engage their attention, and/or handle complex situations.

Losses from using AI can be substantially mitigated if human callers take over cases from AI after a few days. Nonetheless, even in this collaborative arrangement, borrowers who were initially called by AI have paid less after a year than borrowers who were always called by humans. The alienation created by AI contact extends beyond the initial late payment in question; AI-contacted borrowers are more likely to miss each of their next 12 monthly payments. Thus, AI callers create semi-permanent damage to the company's relationship with its borrowers.

Of course, AI technology will continue advance, and we may soon reach a point where it is difficult for people to discern whether a caller is a human or AI. At that point, people may rationally assume that *all* callers are AIs. If an important driver of the negative reaction to AI is simply the knowledge that it is not a human, then call recipients may become less cooperative with all callers, regardless of whether the caller is actually human.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2023. "When should you adjust standard errors for clustering?." *Quarterly Journal of Economics,* 138(1): 1-35.

Acemoglu, Daron, and Pascual Restrepo. 2018. "The race between man and machine: Implications of technology for growth, factor shares, and employment." *American Economic Review*, 108(6): 1488-1542.

Acemoglu, Daron, and Pascual Restrepo. 2019. "Artificial intelligence, automation, and work." In *The Economics of Artificial Intelligence: An Agenda* (pp. 197-236). University of Chicago Press.

Acemoglu, Daron, and Pascual Restrepo. 2020. "AIs and jobs: Evidence from US labor markets." *Journal of Political Economy*, 128(6): 2188-2244.

Acemoglu, Daron, and Pascual Restrepo. 2022. "Tasks, automation, and the rise in us wage inequality." *Econometrica*, 90(5): 1973-2016.

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb. 2019. "Artificial intelligence: the ambiguous labor market impact of automating prediction." *Journal of Economic Perspectives*, 33(2): 31-50.

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz. 2023. "Combining human expertise with artificial intelligence: Experimental evidence from radiology." *NBER Working Paper* No. w31422.

Aydin, Deniz. 2022. "Consumption response to credit expansions: Evidence from experimental assignment of 45,307 credit lines." *American Economic Review*, 112(1): 1-40.

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. 2023. "Generative AI at work." *NBER Working Paper* No. w31161.

Brynjolfsson, Erik, and Tom Mitchell. 2017. "What can machine learning do? Workforce implications." *Science*, 358(6370): 1530-1534.

Burstyn, Leonardo, Stefano Fiorin, Daniel Gottlieb, and Martin Kanz. 2019. "Moral incentives in credit card debt repayment: Evidence from a field experiment." *Journal of Political Economy*, 127(4): 1641-1683.

Cao, Sean, Wei Jiang, Junbo L. Wang, and Baozhong Yang. 2024. "From man vs. machine to man+ machine: The art and AI of stock analyses." *Journal of Financial Economics*, forthcoming.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and partnership." *Econometrica*, 74(6): 1579-1601.

Cheng, Ing-Haw, Felipe Severino, and Richard R. Townsend. 2021. "How do consumers fare when dealing with debt collectors? Evidence from out-of-court settlements." *Review of Financial Studies*, 34(4): 1617-1660.

Cialdini, Robert B. 2001. "The science of persuasion." *Scientific American*, 284(2): 76-81.

Cohn, Alain, Tobias Gesche, and Michel André Maréchal. 2022. "Honesty in the digital age." *Management Science*, 68(2): 827-845.

Drozd, Lukasz A., and Ricardo Serrano-Padial. 2017. "Modeling the revolving revolution: the debt collection channel." *American Economic Review*, 107(3): 897-930.

Ellingsen, Tore, and Magnus Johannesson. 2004. "Promises, threats and fairness." *Econnomic Journal*, 114(495): 397-420.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. "GPTs are GPTs: An early look at the labor market impact potential of large language models." *arXiv*: 2303.10130.

Erel, Isil, Léa H. Stern, Chenhao Tan, and Michael S. Weisbach. 2021. "Selecting directors using machine learning." *Review of Financial Studies*, 34(7): 3226-3264.

Fedaseyeu, Viktar. 2020. "Debt collection agencies and the supply of consumer credit." *Journal of Financial Economics*, 138(1): 193-221.

Fedaseyeu, Viktar, and Robert M. Hunt. 2015. "The economics of debt collection: Enforcement of consumer credit contracts." *FRB of Philadelphia Working Paper* No. 15-43.

Felten, Edward W., Manav Raj, and Robert Seamans. 2020. "The occupational impact of artificial intelligence: Labor, skills, and polarization." *SSRN Working Paper* No.3368605.

Frank, Morgan R., David Autor, James E. Bessen, Erik Brynjolfsson, Manuel Cebrian, David J. Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro, Dashun Wang, Hyejin Youn, and Iyad Rahwan. 2019. "Toward understanding the impact of artificial intelligence on labor." *Proceedings of the National Academy of Sciences*, 116(14): 6531-6539.

Gallegos, Demetria. 2024. "Is it OK to be mean to a chatbot?" *Wall Street Journal*, February 15. https://www.wsj.com/tech/ai/artificial-intelligence-chatbot-manners-65a4edf9

Gao, Zihan, and Jiepu Jiang. 2021. "Evaluating human-AI hybrid conversational systems with chatbot message suggestions." In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*: 534-544.

Glikson, Ella, and Anita Williams Woolley. 2020. "Human trust in artificial intelligence: Review of empirical research." *Academy of Management Annals* 14(2): 627-660.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2013. "The determinants of attitudes toward strategic default on mortgages." *Journal of Finance*, 68(4): 1473-1515.

He, Simin, Theo Offerman, and Jeroen Van De Ven. 2017. "The sources of the communication gap." *Management Science*, 63(9): 2832-2846.

Karlan, Dean, Melanie Morten, and Jonathan Zinman. 2015. "A personal touch in text messaging can improve microloan repayment." *Behavioral Science & Policy*, 1(2): 25-31.

Kim, Tae Woo, Hyejin Lee, Michelle Yoosun Kim, SunAh Kim, and Adam Duhachek, 2023. "AI increases unethical consumer behavior due to reduced anticipatory guilt." *Journal of the Academy of Marketing Science*, 51: 785-801.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human decisions and machine predictions." *Quarterly Journal of Economics*, 133(1): 237-293.

Laudenbach, Christine, and Stephan Siegel. 2023. "Personal communication in an automated world: Evidence from loan repayments." *Journal of Finance*, forthcoming.

Luo, Xueming, Siliang Tong, Zheng Fang, Zhe Qu. 2019. "Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases." *Marketing Science*, 38(6): 937-947.

Luo, Xueming, Marco Shaojun Qin, Zheng Fang, and Zhe Qu. 2021. "Artificial intelligence coaches for sales agents: Caveats and solutions." *Journal of Marketing*, 85(2): 14-32.

Maslej, Nestor, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Parli, Yoav Shoham, Russell Wald, Jack Clark, and Raymond Perrault. 2023. *The AI Index 2023 Annual Report.* AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

Melo, De Celso, Stacy Marsella, and Jonathan Gratch. 2016. "People do not feel guilty about exploiting machines." *ACM Transactions on Computer-Human InteractionI,* 23: 1–17.

Michaels, Guy, Ashwini Natraj, and John Van Reenen. 2014. "Has ICT polarized skill demand? Evidence from eleven countries over twenty-five years." *Review of Economics and Statistics*, 96(1): 60-77.

Noy, Shakked, and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science*, 381(6654): 187-192.

Petisca, Sofia, Ana Paiva, and Francisco Esteves. 2020. "Perceptions of people's dishonesty towards AIs." *International Conference on Social AIics*: 132–143. Springer.

Roll, Stephen P., and Stephanie Moulton. 2019. "The impact of automated reminders on credit outcomes: Results from an experimental pilot program." *Journal of Consumer Affairs*, 53(4): 1693-1724.

Silva, Gisele S., Rohan Khera, and Lee H. Schwamm. 2024. "Reviewer experience detecting and judging human versus artificial intelligence content: The *Stroke* journal essay contest." *Stroke* 55: 2573-2578.

Tan, Tom Fangyun, and Serguei Netessine. 2020. "At your service on the table: Impact of tabletop technology on restaurant performance." *Management Science*, 66 (10): 4496-4515.

Vanberg, Christoph. 2008. "Why do people keep their promises? An experimental test of two explanations." *Econometrica*, 76(6): 1467-1480.

Wang, Qingchen, and Yijun Zhou. 2024. "Artificial intelligence and debt collection: Evidence from a field experiment." Working paper.

World Economic Forum. 2020. *The Future of Jobs Report 2020*. Switzerland.

World Economic Forum. 2023. *The Future of Jobs Report 2023*. Switzerland.

Figures

Figure 1. Case assignment between AI and human callers on day 2 past due.

This figure shows how cases are assigned between AI and human callers. "Almost always AI" means that more than 95% of cases are always handled by AI callers over the life of the cases, and less than 5% of cases are assigned to human callers after day 25. For the conditionally randomized subsample, the probability of being assigned to a human caller and when humans take over from AI depend on case characteristics and the calendar date.

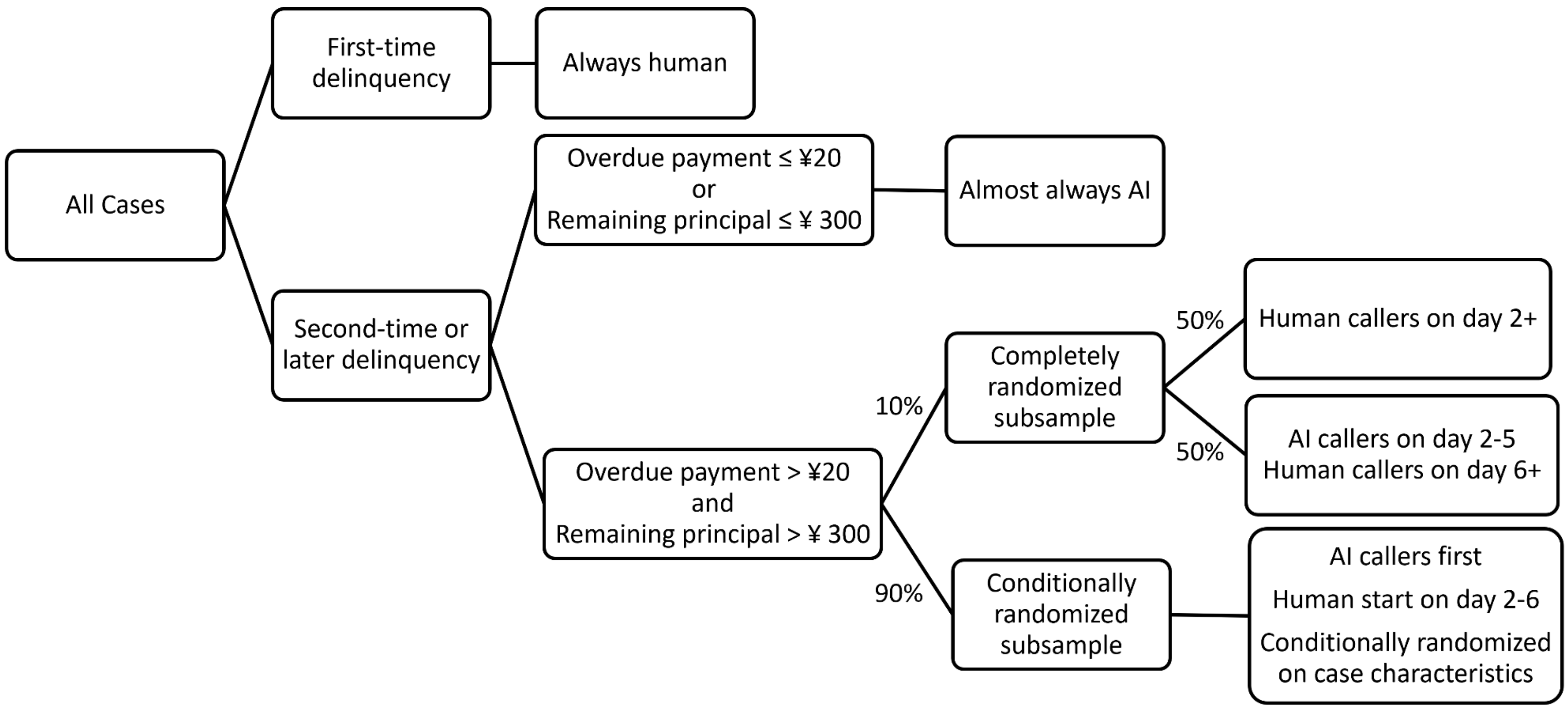
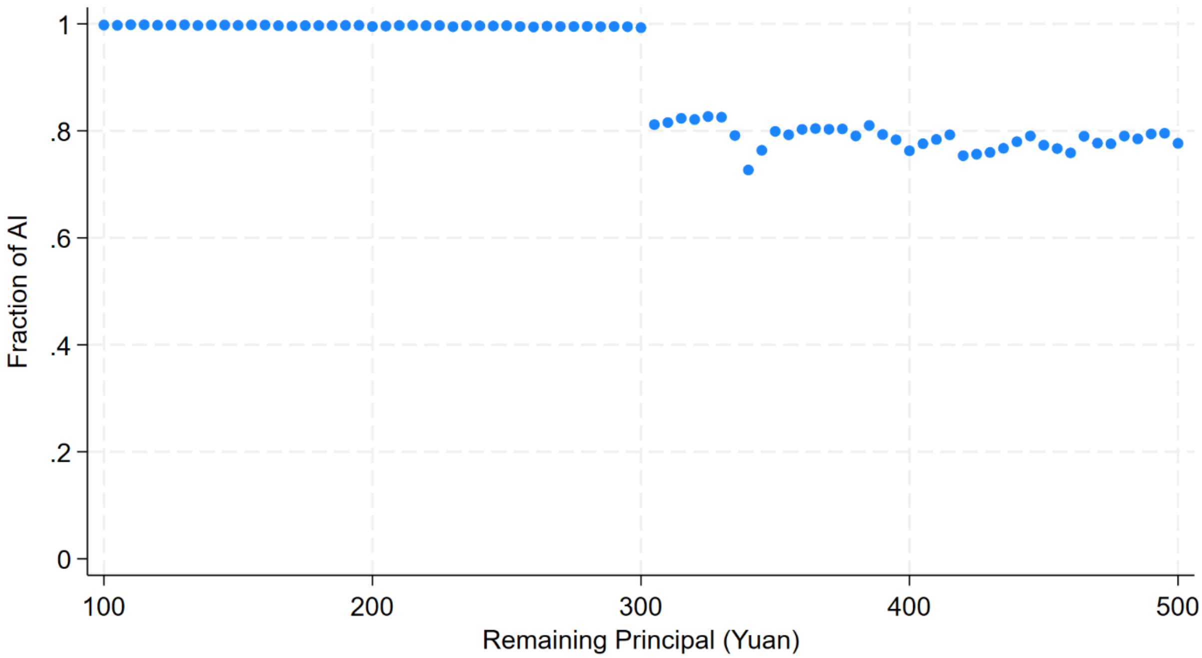


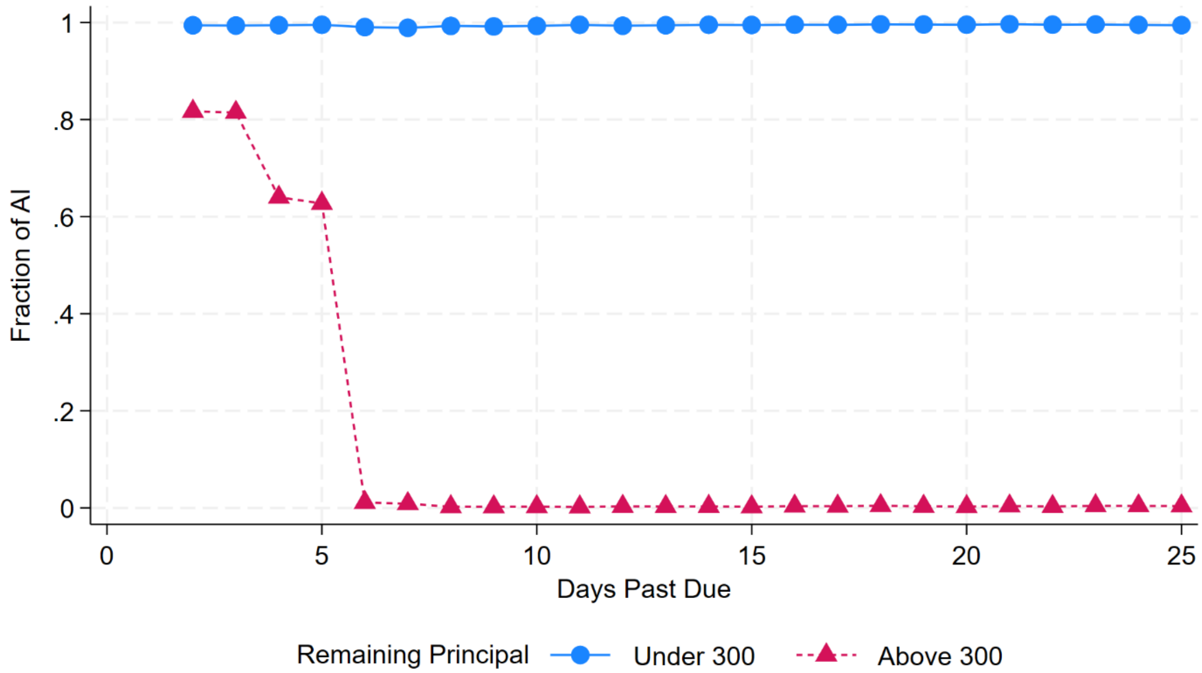
Figure 2. Fraction of cases assigned to AI callers around the 300-yuan remaining principal threshold.

Panel (a) shows the fraction of cases assigned to AI callers on day 2 past due as a function of principal remaining on day 2. Panel (b) shows the AI fractions for cases in the (295, 300] and (300, 305] yuan day-2 principal remaining intervals on days 2-25 past due. Panel (c) extends the horizon of panel (b) to day 360.

(a) Fraction of AI cases by remaining principal on day 2 past due



(b) Fraction of AI cases just below and above the 300-yuan threshold (days 2-25 past due)



(c) Fraction of AI cases just below and above the 300-yuan threshold (days 2-360 past due)

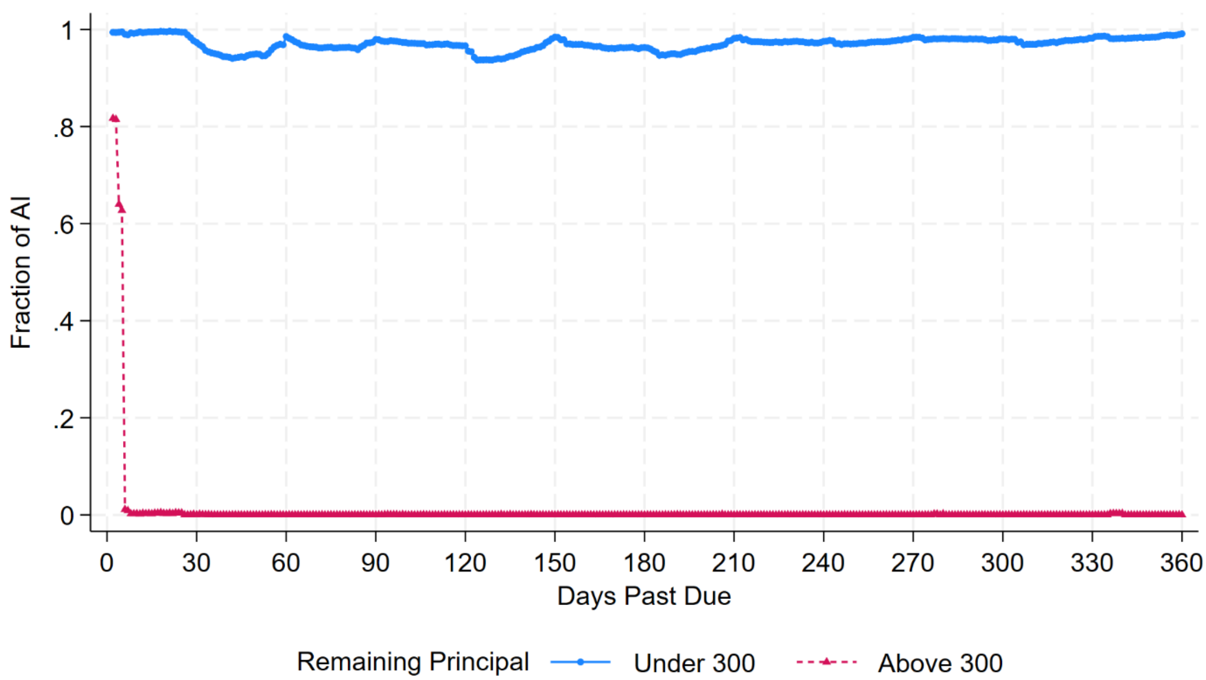
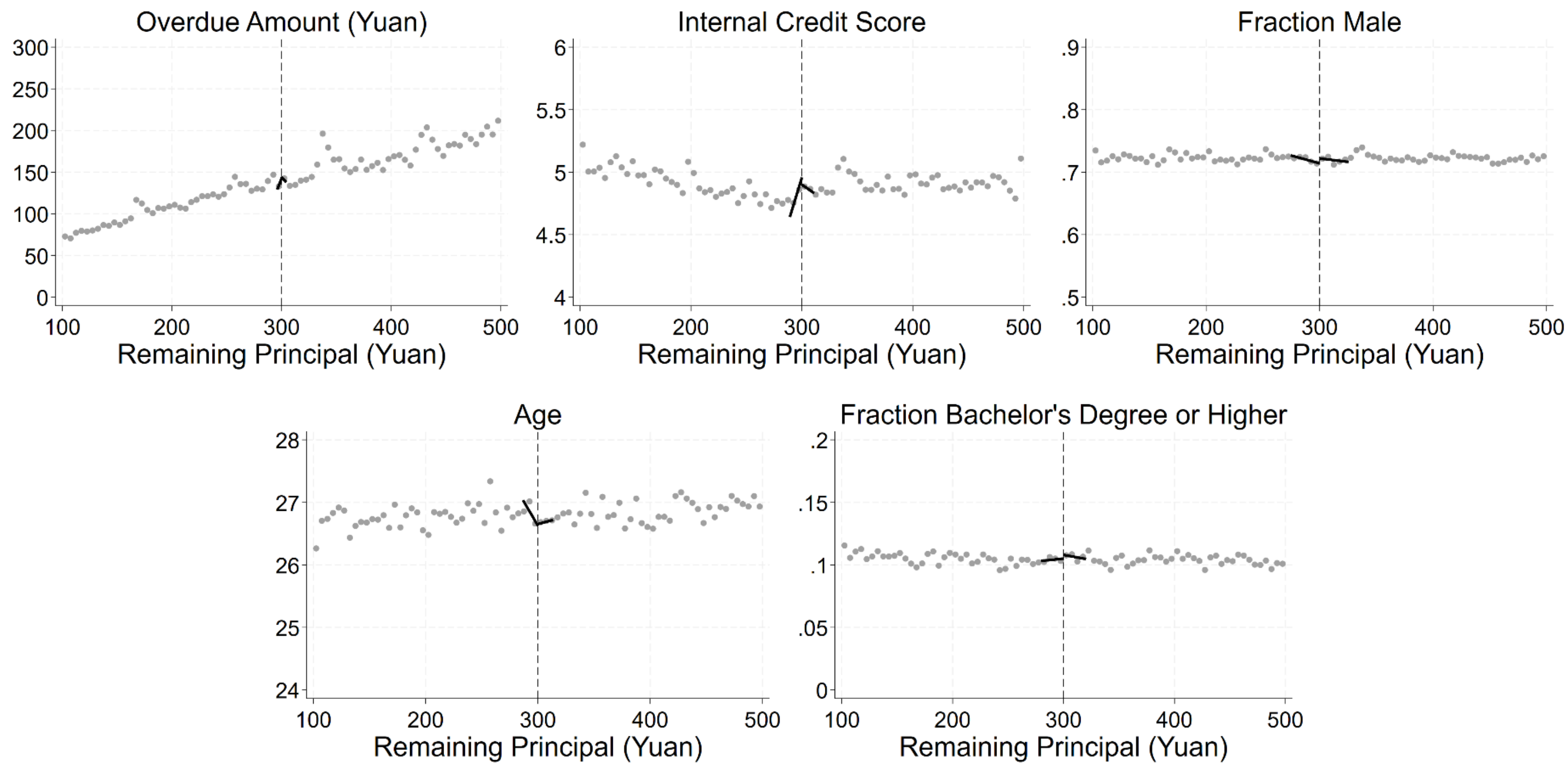


Figure 3. Loan and borrower characteristics and collected NPVs around the 300-yuan remaining principal cutoff.

Panel (a) plots average loan and borrower characteristics measured on day 2 past due as a function of remaining principal on day 2 past due. Panel (b) plots average cumulative NPVs of payments collected up until day 2, 5, 10, 30, 90, or 360 past due, divided by the balance due on day 2, as a function of remaining principal on day 2 overdue. The grey dots are binned averages, and the black lines are local linear fits within the regression discontinuity estimation bandwidths on each side. Coverage-error-rate-optimal bandwidths are used in Panel (a) and mean-squared-error-optimal bandwidths in Panel (b).

(a) Loan characteristics



(b) Cumulative collected NPV

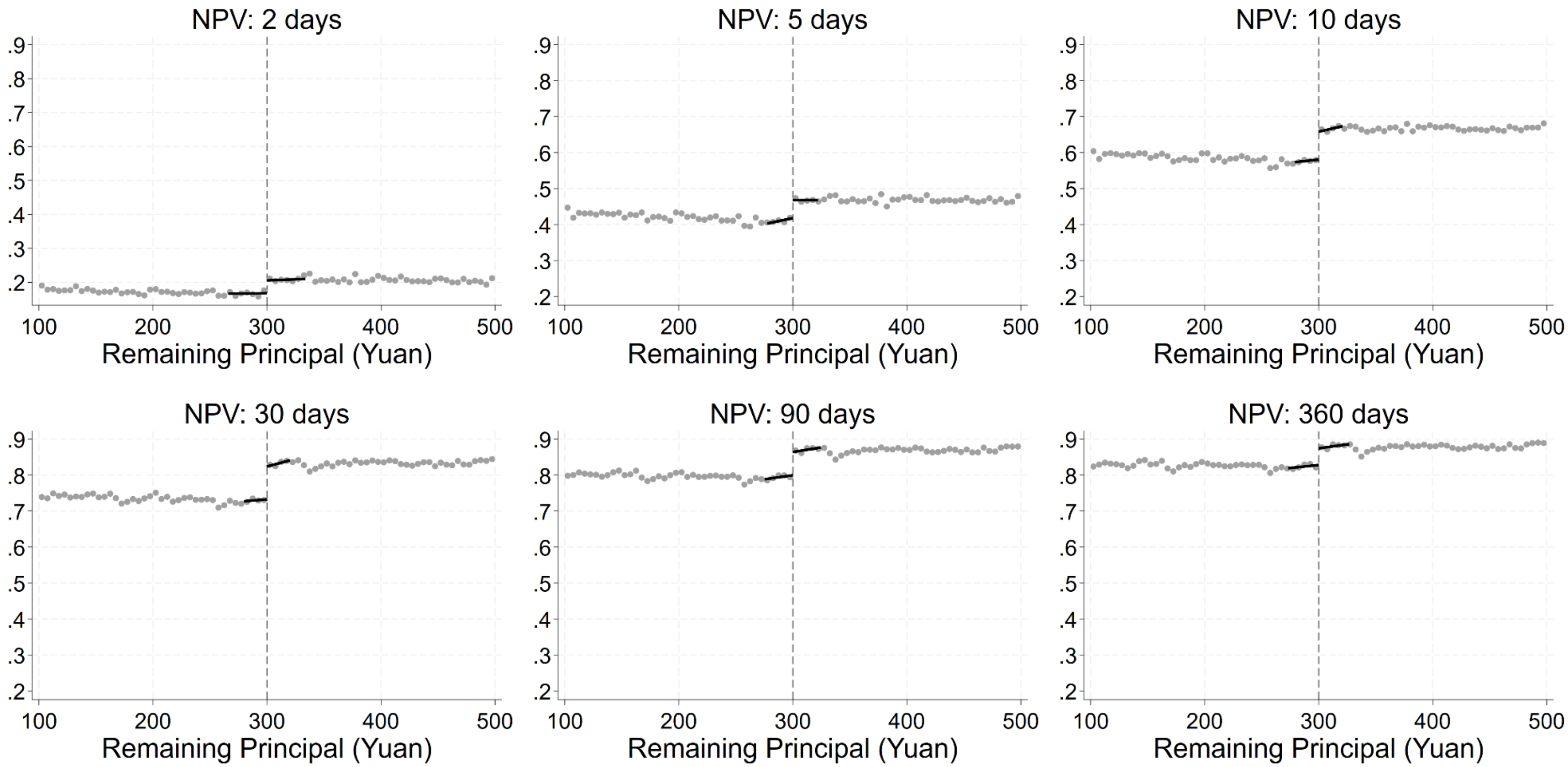


Figure 4. AI minus human caller cumulative collected NPV by horizon – small cases RD.

This figure shows the difference in average cumulative NPVs of payments collected divided by balance due on day 2 between AI and human callers as a function of days past due. The differences are estimated by regression discontinuity around the 300-yuan remaining principal threshold for permanent AI treatment. The bars indicate 95% robust regression discontinuity confidence intervals. Differences are plotted every three days before day 60 and every 10 days after day 60.

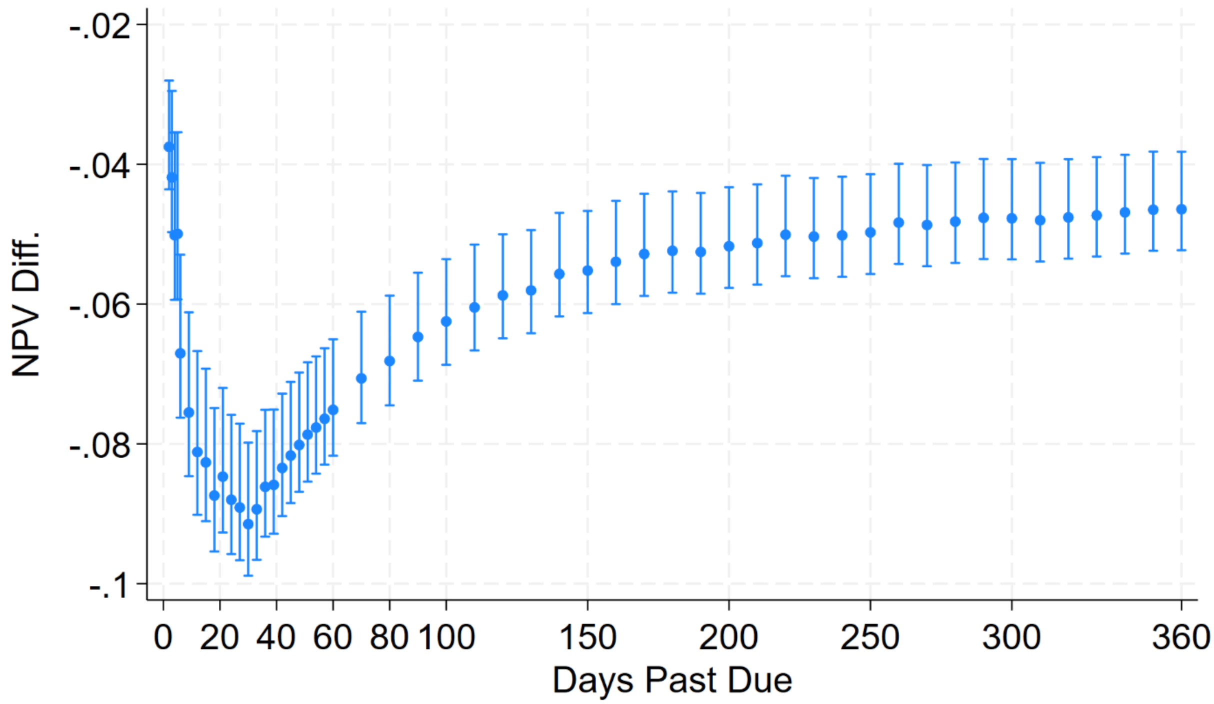


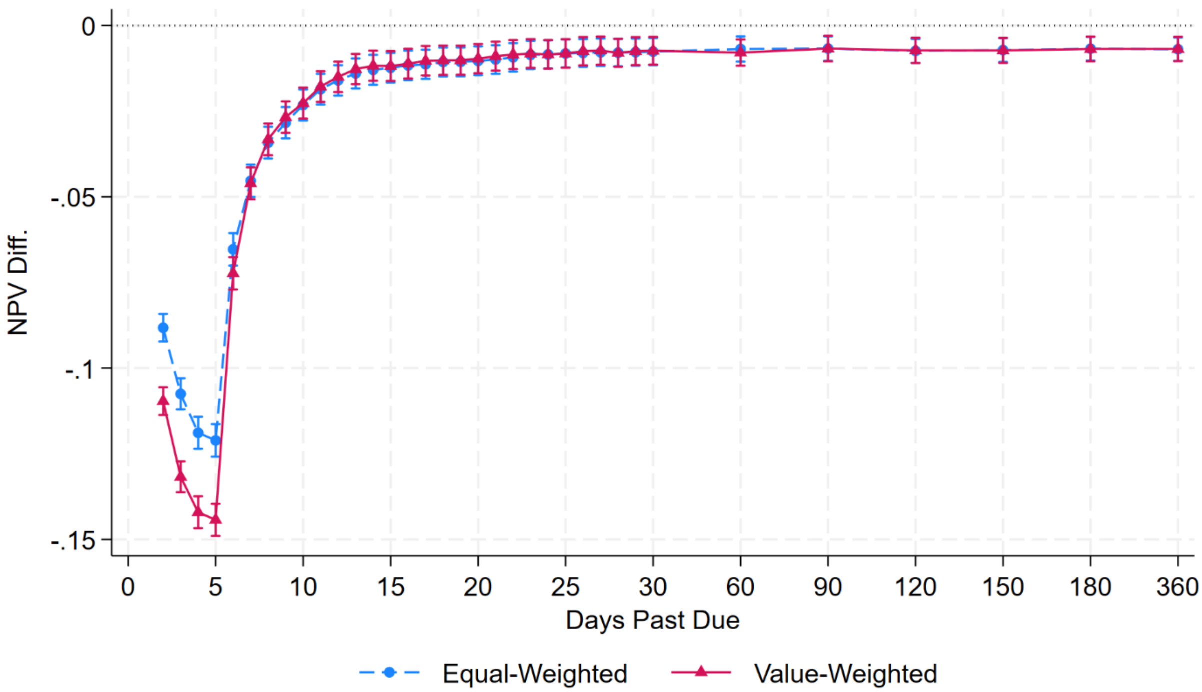
Figure 5. AI minus human caller cumulative collected NPV by horizon, by internal credit score – small cases RD.

This figure shows the differences in average cumulative payment NPVs collected divided by balance due on day 2 between AI and human callers as a function of days past due, separately for borrowers in three groups of internal credit scores. The differences are estimated by regression discontinuity around the 300-yuan remaining principal threshold for permanent AI treatment. “Low”, “Med”, and “High” refer to cases with internal credit scores in deciles 1-3, 4-7, and 8-10, respectively. Differences are plotted every three days before day 60 and every 10 days after day 60.

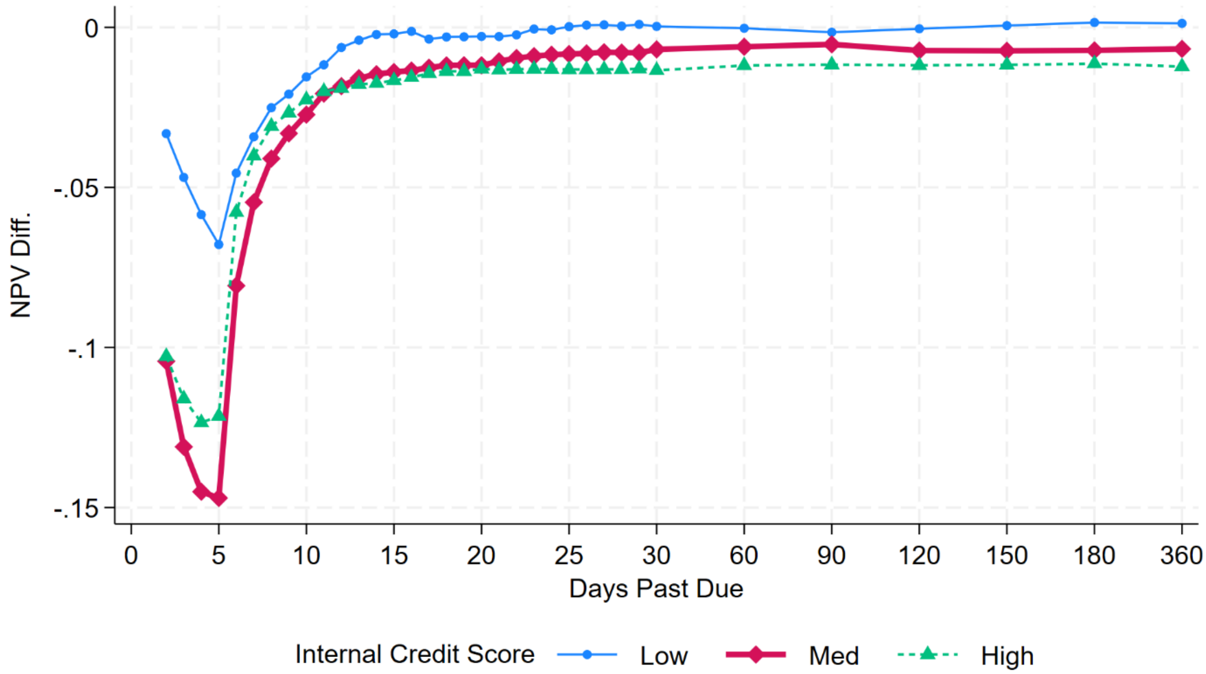
Figure 6. AI + human minus always-human cumulative collected NPV by horizon – Completely randomized subsample.

This figure shows the difference in average cumulative NPVs of payments collected divided by balance due on day 2 between the AI + human and always-human groups as a function of days past due, using the completely randomized subsample. The differences control for month-of-delinquency fixed effects. The bars indicate 95% confidence intervals. Differences are plotted daily before day 30 and every 30 days afterwards. Panel (a) pools all cases together. Panels (b) and (c) split cases by internal credit score and overdue payment size, respectively, and show equal-weighted differences. In Panel (b), “Low”, “Med”, and “High” refer to internal credit scores in deciles 1-3, 4-7, and 8-10, respectively.

(a) All cases.



(b) By internal credit score.



(c) By loan size.

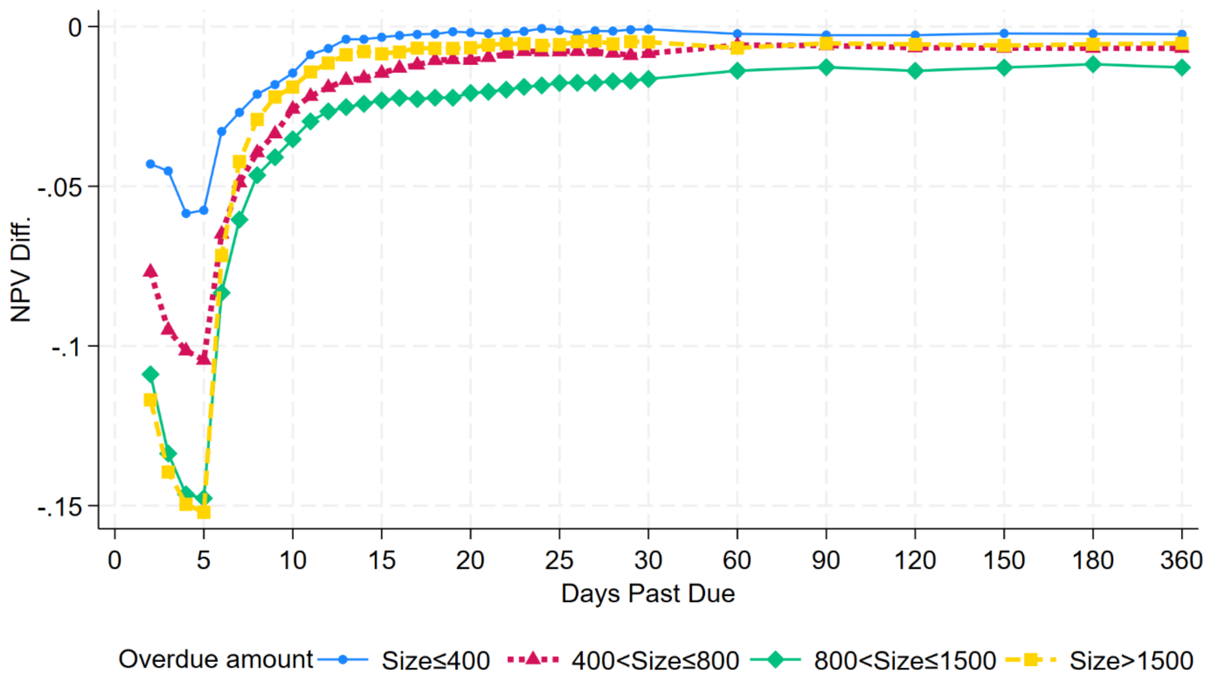


Figure 7. Deployment of different versions of AI callers over time.

This figure shows the fraction of cases assigned to each version of AI callers during each month of our sample period. The first version in our sample period is labeled as "v1," but it is not the very first version of AI caller used by the company. The fractions are calculated within the 10% completely randomized subsample.

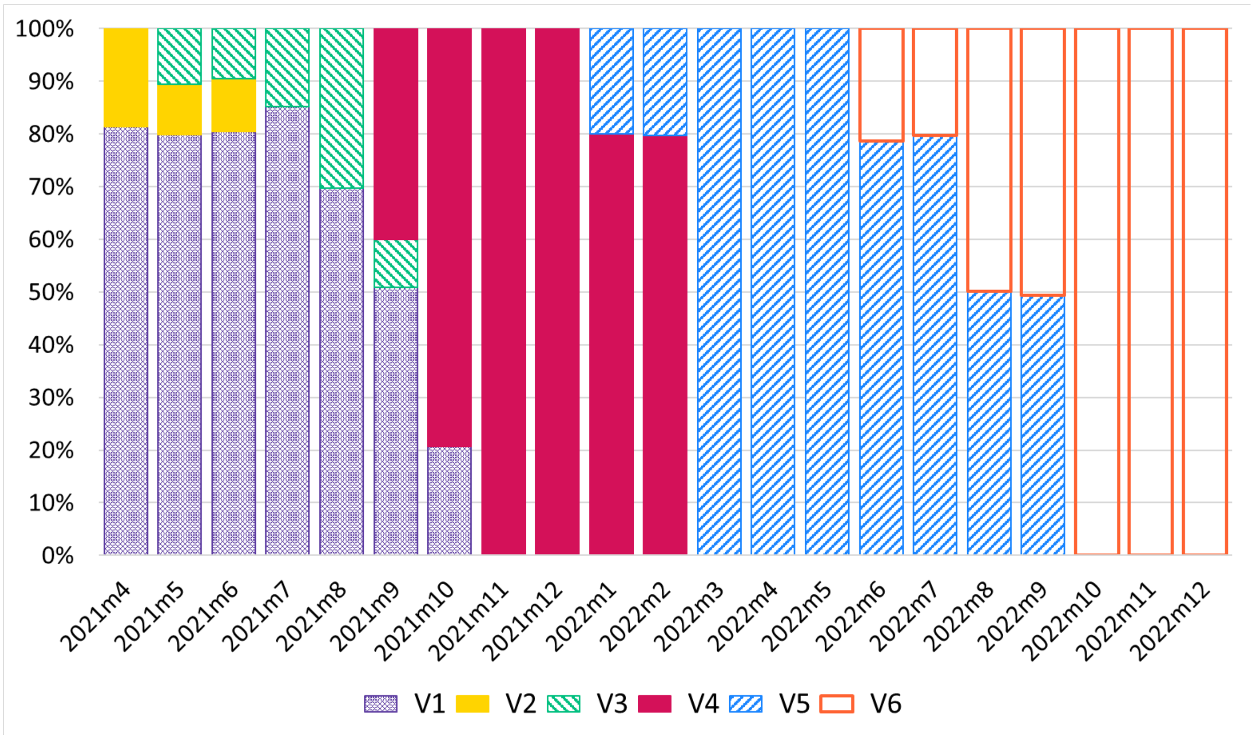
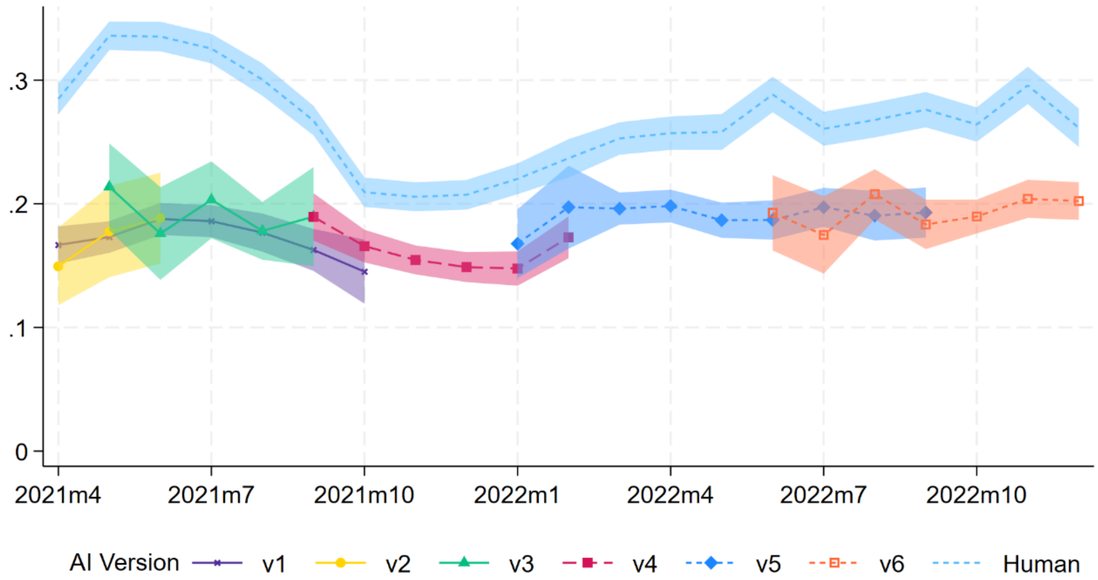


Figure 8. Performance of different versions of AI callers over time.

This figure shows the monthly time-series of performance of different versions of AI callers and human callers, measured by the average cumulative collected NPV on day 2 past due (Panel (a)) or through day 5 past due (Panel (b)). The shaded areas represent 95% confidence intervals.

(a) NPV on day 2



(b) NPV over first 5 days

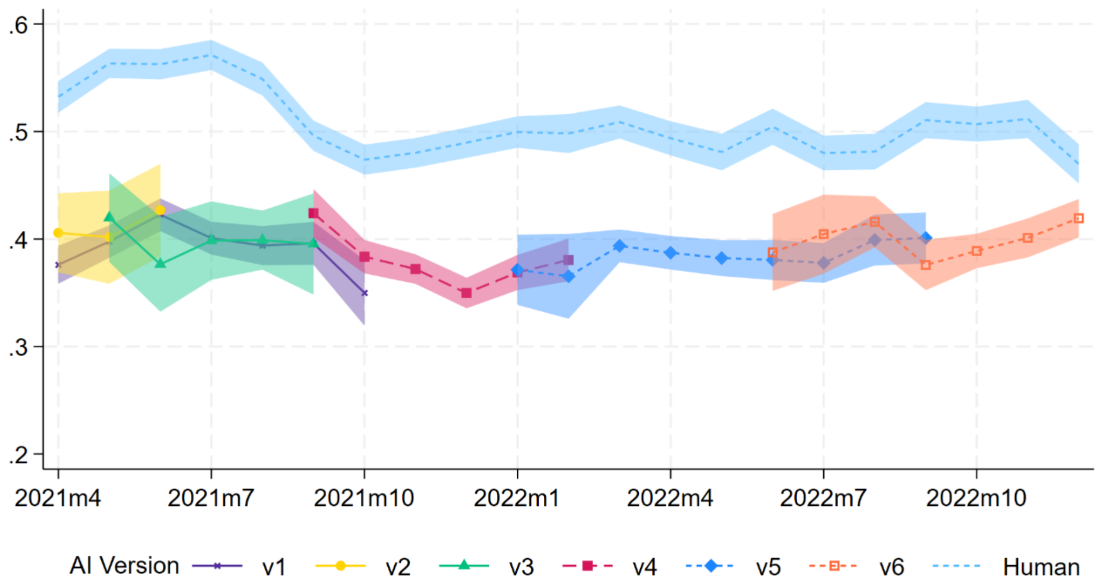
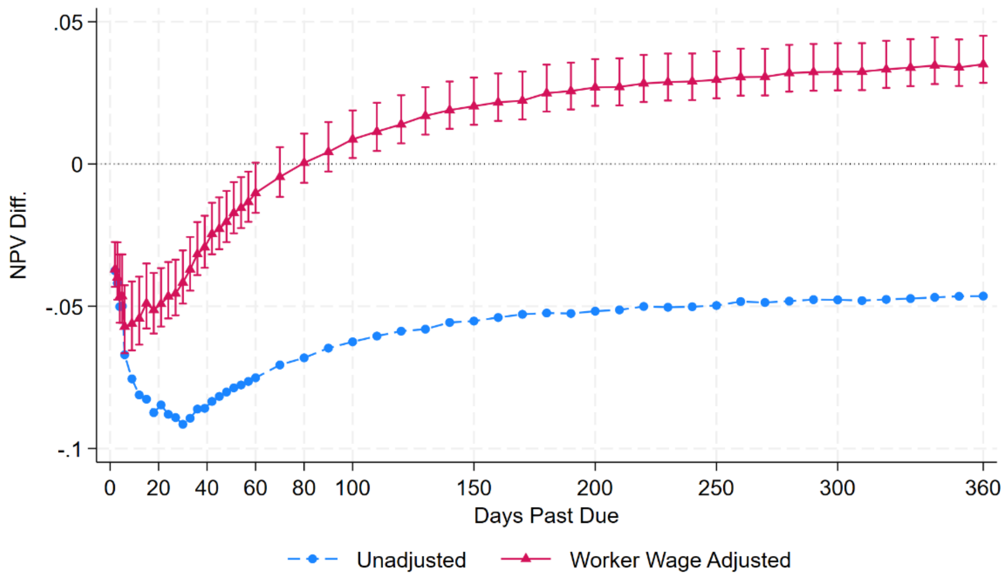


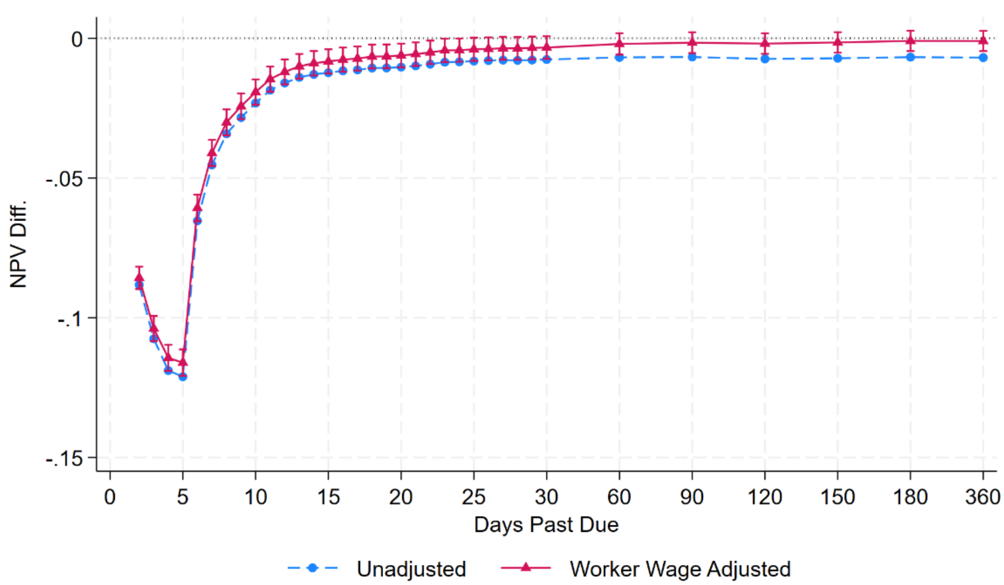
Figure 9. Collected NPV differences between AI and human callers over time, adjusted for labor costs.

This figure shows the difference in average cumulative NPVs of payments collected net of labor costs divided by balance due on day 2 between AI and human callers as a function of days past due. The bars indicate 95% confidence intervals. Panel (a) shows results from small cases using a regression discontinuity design. Panels (b) and (c) use the completely randomized subsample, with the former equally weighting all cases and the latter weighting cases by the overdue payment amount on day 2.

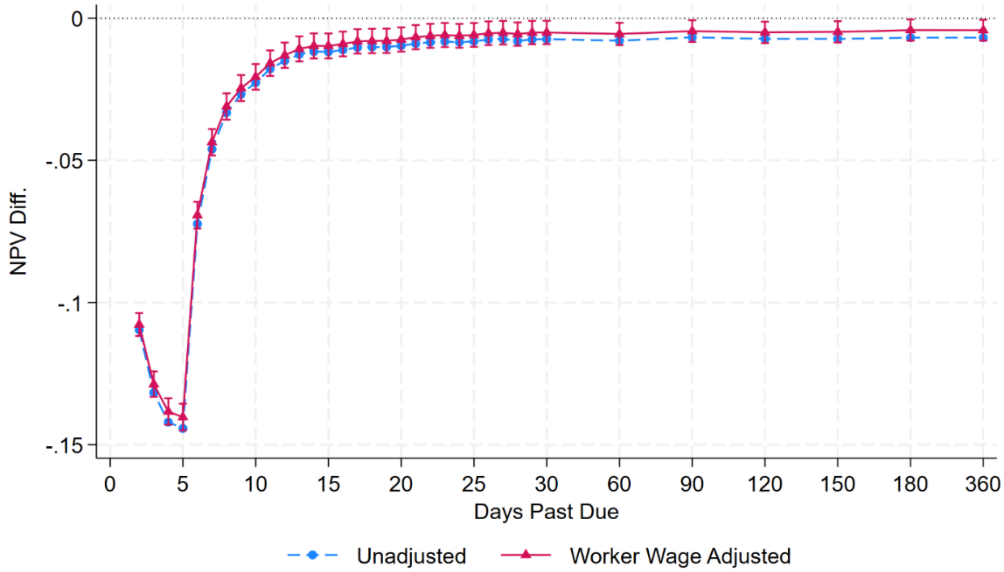
(a) Small cases – RDD.



(b) Completely randomized subsample, equal-weighted.



(c) Completely randomized subsample, value-weighted.



Tables

Table 1. Sample scripts of AI caller.

This table presents sample scripts that the AI caller typically uses in a phone call for debt collection. The third stage may include multiple rounds of questions and responses.

Stage	Stage Content	Sample Scripts
1	(1) Greeting. (2) Confirm the identity of the borrower.	AI: Hello, is that Mr./Ms. ABC (borrower's full name) speaking? Borrower: Yes (AI moves forward to the second stage) Borrower: No (AI apologizes and hangs up the phone)
2	(1) Inform the borrower about the delinquent debt. (2) Address the negative outcomes of default.	AI: Mr./Ms. ABC, this is XYZ company. You previously borrowed x yuan from our company, and you should pay the monthly repayment of y yuan in full by [bill date]. But as of today, you have been missing the monthly repayment for z days. Please repay the amount by the end of today / within h hours. Any overdue payment can lead to declining credit scores, large late fees, difficulties in future borrowing and consumption, and even lawsuits. We may also call your emergency contacts if you fail to repay.
3	*Case A1*. Borrower agrees to repay today. AI confirms the promise.	Borrower: Yes, I will repay the debt. AI: Could you promise to repay the overdue amount in full by the end of today / within h hours? Borrower: Yes. AI: Great! I have recorded your promise in our system. Please keep your promise. Otherwise, we will continue calling you.
	Case A2. Borrower agrees to repay later. AI confirms the promise and addresses potential negative consequences.	Borrower: No, I cannot do it today. How about tomorrow? AI: Your loans are related to your credit records. If you continue to delay, your case will enter the debt collection process. Can you repay *today*? Borrower: Tomorrow is the best I can do. AI: Okay. We have recorded your promise. No late fees will be charged for today. But if you continue to default, your credit records will be negatively affected.
	Case B1. Borrower refuses to repay and does not explain the reasons. (1) AI emphasizes the negative outcomes again, asks for reasons	Borrower: No, I cannot repay the debt right now. AI: It is your obligation to repay the debt. Any overdue payment can lead to [potential negative consequences]. Could you tell me the reason why you cannot repay the debt? Do you have any difficulties in repaying the debts? (Move to Case B2.)

	Case B2. Borrower refuses to repay and explains the reason. AI replies accordingly and addresses the negative outcomes.	Borrower: I cannot repay today because I have no money to repay / I am very busy today. AI: You can always find a way to raise money. For example, you can borrow from your family members and friends. / I understand you are busy. Please keep in mind that … [potential negative consequences].
	Case C. Borrower denies having debts with the company. AI asks the borrower to recall.	Borrower: I don't know XYZ company. / I have never borrowed money from your company. AI: Please recall carefully if you have ever borrowed money from XYZ company. The company's name is spelled as "X-Y-Z." Please be advised that late repayment can lead to negative outcomes.
	Case D. Borrower claims that he/she has repaid the debt fully or has set up auto-payment.	Borrower: I have already paid back the debt this morning, haven't you received the money yet? AI: But there is still y yuan on your balance. Borrower: I have set up auto-payment. AI: Okay. We will charge y yuan from your linked bank / WeChat / Alipay account shortly. Please make sure that you have sufficient balances in your bank account.
	Case E. Borrower asks for additional information.	Borrower: Do I need to pay any late fees? AI: Late fees include overdue interest and principal, credit evaluation fees, guarantee fees, and so on. Details can be found in our app and your loan contract. Borrower: If I can repay today, do I need to pay late fees? AI: Okay, we will temporarily suspend additional debt collection actions. You can make sure you will repay within h hours, right? Borrower: How long has it been overdue? AI: You have been z days past due. We have sent you several text messages before.
4	When the borrower has no more questions, the borrower's questions do not belong to the above cases, or the AI cannot understand the borrower's response.	AI: Okay. Please be advised that you will be responsible for any negative consequences of default. If you have any other questions, feel free to contact our customer service. Bye!

Table 2. Summary statistics of delinquent loans.

This table reports summary statistics for the full sample of delinquent loans and two different subsamples used in our analyses. Variables are measured on day 2 past due. Panel A shows the mean, standard deviation, selected percentiles, and observation counts for the full sample of delinquent loans. Panel B shows statistics for delinquent loans with remaining principal between 100 yuan and 500 yuan. Panel C shows statistics for the completely randomized subsample, which is restricted to borrowers in their second delinquency.

Panel A. Full sample

Variable	Mean	S.D.	Min	P1	P25	P50	P75	P99	Max	# Obs.
Overdue amount (yuan)	1,128.1	1,822.4	0.01	14.7	316.0	653.5	1,304.6	7,688.8	808,666.7	22,122,179
Remaining principal (yuan)	6,474.0	7,330.0	0.01	48.6	1,792.5	4,248.1	8,500.0	34,448.4	1,000,000.0	22,122,179
Internal credit score	5.42	2.85	1	1	3	5	8	10	10	22,122,179
Male indicator	0.70	0.46	0	0	0	1	1	1	1	22,122,179
Age	27.43	6.36	18	19	23	26	31	46	60	22,122,179
Bachelor's degree or more indicator	0.13	0.34	0	0	0	0	0	1	1	22,122,179

Panel B. RDD subsample

Variable	Mean	S.D.	Min	P1	P25	P50	P75	P99	Max	# Obs.
Overdue amount (yuan)	142.13	112.53	20.01	22.12	58.60	106.09	188.68	503.16	848.18	1,010,540
Remaining principal (yuan)	304.74	112.45	100.00	104.03	209.00	308.03	400.66	496.19	499.99	1,010,540
Internal credit score	4.91	2.77	1	1	3	4	7	10	10	1,010,540
Male indicator	0.72	0.45	0	0	0	1	1	1	1	1,010,540
Age	26.81	5.98	18	19	22	25	30	46	60	1,010,540
Bachelor's degree or more indicator	0.10	0.31	0	0	0	0	0	1	1	1,010,540

Panel C. Completely randomized subsample

Variable	Mean	S.D.	Min	P1	P25	P50	P75	P99	Max	# Obs.
Overdue amount (yuan)	1,522.9	1,846.4	20.2	86.3	554.8	1,018.0	1,849.4	8,653.9	35,639.9	147,424
Remaining principal (yuan)	8,593.9	6,966.4	300.1	467.5	3,438.0	6,600.1	11,667.8	30,968.8	34,919.6	147,424
Internal credit score	5.97	2.71	1	1	4	6	8	10	10	147,424
Male indicator	0.70	0.46	0	0	0	1	1	1	1	147,424
Age	27.77	6.79	18	19	22	26	32	47	59	147,424
Bachelor’s degree or more indicator	0.10	0.31	0	0	0	0	0	1	1	147,424

Table 3. Treatment effect of permanent assignment to AI callers – RD results.

Panel A reports tests of continuity at the 300-yuan remaining principal threshold for predetermined loan and borrower characteristics measured on day 2 past due. Columns 2 and 3 report the regression-fitted value of the variables of interest at the threshold from the left side (permanent AI) and the right side (human). Column 4 reports the differences between the left and right fitted values (AI minus human), with *z*-statistics, *p*-values, and 95% robust RD confidence intervals in the following columns. Panel B reports the treatment effect of being permanently assigned to an AI caller on the present value of cash flows collected from the case within a given horizon after the original due date, scaled by the day 2 overdue balance. In addition to the first seven columns as in Panel A, Panel B column 8 re-estimates the differences around the threshold while controlling for all five covariates in Panel A. Local linear regressions with uniform kernels are used in the estimation in all rows. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)		(8)
	Variable	Left Mean (AI)	Right Mean (Human)	Diff. (L–R)	*z*-stat.	*p*-val.	95% Bias-Corrected C.I.		Diff. with Covariates
Panel A. Loan and borrower characteristics									
(1)	Overdue payment	142.9	143.8	-0.94	-0.40	0.686	-6.83	4.50	
(2)	Credit score	4.96	4.90	0.06	1.30	0.194	-0.03	0.15	
(3)	Male	0.714	0.722	-0.008	-1.51	0.132	-0.018	0.002	
(4)	Age	26.63	26.65	-0.027	-0.22	0.824	-0.199	0.159	
(5)	Bachelor's degree or higher	0.105	0.108	-0.003	-0.73	0.468	-0.011	0.005	
Panel B. Cumulative NPV collected									
(6)	2 days	0.168	0.206	-0.038***	-9.02	<0.001	-0.044	-0.028	-0.040***
(7)	5 days	0.418	0.468	-0.050***	-7.77	<0.001	-0.059	-0.035	-0.049***
(8)	10 days	0.581	0.658	-0.078***	-12.73	<0.001	-0.087	-0.064	-0.076***
(9)	30 days	0.733	0.824	-0.092***	-18.40	<0.001	-0.099	-0.080	-0.091***
(10)	60 days	0.782	0.857	-0.075***	-17.24	<0.001	-0.082	-0.065	-0.074***
(11)	90 days	0.800	0.864	-0.065***	-16.06	<0.001	-0.071	-0.056	-0.064***
(12)	180 days	0.820	0.872	-0.052***	-13.84	<0.001	-0.058	-0.044	-0.052***
(13)	360 days	0.828	0.875	-0.046***	-12.59	<0.001	-0.052	-0.038	-0.046***

Table 4. Treatment effect of five days of AI callers – Completely randomized subsample.

This table compares two groups in the 10% completely randomized subsample: the treatment group handled by AI callers from days 2 to 5 past due before being given to human callers on day 6, and the control group handled by human callers from day 2 onwards. Panel A reports tests that the two groups are balanced in their pre-determined loan and borrower characteristics measured on day 2. Columns 2 and 3 report the average of the variables of interest among cases assigned to the treatment and the control groups, respectively. Column 4 reports the differences between the averages, with *t*-statistics in the following column. Panel B reports the performance of the two groups measured by collected NPV scaled by the day 2 overdue balance. The estimates come from linear regressions of the variable of interest onto a treatment group indicator and day-2 calendar month fixed effects. In columns 2-5, cases are equally weighted in the regression, and in column 6, cases are weighted by their overdue payment amount on day 2. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
	Variables	Mean (AI)	Mean (Human)	Diff: AI – Human	*t*-stat.	Diff: AI – Human (value-wgt.)
Panel A. Loan and borrower characteristics						
(1)	Overdue amount	1523.7	1522.2	1.5	0.15	
(2)	Remaining principal	8585.2	8604.7	-19.5	-0.54	
(3)	Internal credit score	5.970	5.961	0.009	0.66	
(4)	Male	0.701	0.703	-0.002	-0.83	
(5)	Age	27.75	27.79	-0.043	-1.25	
(6)	Bachelor's degree or higher	0.103	0.104	-0.001	-0.36	
Panel B. Cumulative NPV collected						
(7)	2 days	0.180	0.268	-0.088***	-42.89	-0.110***
(8)	5 days	0.389	0.510	-0.121***	-50.07	-0.144***
(9)	10 days	0.612	0.635	-0.023***	-10.02	-0.023***
(10)	30 days	0.733	0.741	-0.0076***	-3.77	-0.0074***
(11)	60 days	0.767	0.774	-0.0069***	-3.63	-0.0079***
(12)	90 days	0.783	0.790	-0.0067***	-3.69	-0.0067***
(13)	180 days	0.796	0.803	-0.0068***	-3.86	-0.0068***
(14)	360 days	0.802	0.809	-0.0069***	-4.02	-0.0069***

Table 5. Phone call outcomes of AI and human callers on day 2.

This table compares outcomes for phone calls made on day 2 past due by AI callers versus human callers within the 10% completely randomized subsample. Columns 2 and 3 report the average of the variables of interest within the AI and human caller groups, respectively. Column 4 reports the differences between the averages, with *t*-statistics in the following column. The estimations are based on linear regressions of the outcome variable onto an AI-caller indicator and calendar month fixed effects. Panel A uses data on all phone calls made on day 2 past due, while Panels B and C restrict the sample to the first call answered by each borrower on day 2. The time of the call is represented by hours since midnight in decimals. The timing adjustment accounts for the time of calling by including fixed effects for the time of calling every hour. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Panel A. All calls on day 2 past due.

	(1) Variables	(2) Mean (AI)	(3) Mean (Human)	(4) Diff: AI – Human	(5) *t*-stat.
(1)	# Phone calls per borrower	4.62	5.47	-0.85***	-56.54
(2)	# Phone calls answered	0.65	1.00	-0.35***	-57.63
(3)	% Phone calls answered	0.236	0.236	0.000	0.06

Panel B. First answered calls.

	(1) Variables	(2) Mean (AI)	(3) Mean (Human)	(4) Diff: AI – Human	(5) *t*-stat.
(1)	Time of calls (hours since midnight)	11.79	11.51	0.28**	12.51
(2)	Ringing time to answer (sec)				
	Unadjusted	19.47	20.72	-1.25**	-11.66
	Timing-adjusted			-0.02	-0.17
(3)	Duration (sec)				
	Unadjusted	28.12	47.13	-19.02***	-41.41
	Timing-adjusted			-30.96***	-61.25
(4)	% Promise to repay				
	Unadjusted	0.441	0.651	-0.211***	-52.44
	Timing-adjusted			-0.211***	-45.86
(5)	Prob. answering the next call				
	Unadjusted	0.412	0.406	0.006	1.54
	Timing-adjusted			0.055***	9.42

Panel C. Repayment after first answered calls (all timing-adjusted).

(1) Variables	(2) Mean (AI)	(3) Mean (Human)	(4) Diff: AI – Human	(5) *t*-stat.
Repay (fully or partially) after the first answered call within …				
15 minutes	0.045	0.048	-0.003	-1.32
30 minutes	0.060	0.077	-0.017***	-6.94
1 hour	0.078	0.120	-0.042***	-14.48
2 hours	0.107	0.188	-0.081***	-23.72
5 hours	0.169	0.310	-0.141***	-34.63
the same day	0.270	0.451	-0.181***	-40.03

Table 6. Repayment after the first answered calls, conditional on promises to repay.

This table reports the fraction of borrowers who repay their debts (fully or partially) within various periods after the first answered phone call from AI or human callers on day 2 after the due date, conditioning on whether the borrowers make a promise to repay their debt during the conversation. The analysis uses the 10% completely randomized subsample. The estimates are from linear regressions of the outcome variable onto an AI-case indicator and calendar month and hour-of-call fixed effects. The numbers in parentheses are *t*-statistics for the difference between the proportions in the indicated columns. *** p<0.01, ** p<0.05, * p<0.1.

	With a Promise		Without a Promise				
	(a) AI	(b) Human	(c) AI	(d) Human	(e) Diff: (a) - (c)	(f) Diff: (b) - (d)	(g) Diff: (e) - (f)
15 minutes	0.061	0.059	0.033	0.027	0.029***	0.032***	-0.003
					(11.05)	(12.07)	(-0.92)
30 minutes	0.081	0.096	0.043	0.041	0.039***	0.055***	-0.016***
					(12.71)	(17.62)	(-3.74)
1 hour	0.107	0.149	0.055	0.065	0.052***	0.084***	-0.032***
					(14.46)	(22.83)	(-6.25)
2 hours	0.149	0.232	0.073	0.105	0.076***	0.127***	-0.050***
					(18.02)	(29.17)	(-8.3)
5 hours	0.230	0.380	0.120	0.179	0.110***	0.201***	-0.091***
					(21.92)	(39.15)	(-12.72)
the same day	0.343	0.537	0.212	0.290	0.132***	0.247***	-0.115***
					(23.67)	(43.41)	(-14.55)

Table 7. Relationship between phone call strategies and performance on day 2.

This table shows how human callers' phone call frequency and follow-up call frequency are related to their collection performance on day 2 past the due date. The sample is restricted to human callers specializing in the first five days past due. For each frequency metric, human callers are grouped into terciles by their average call behavior across all their assigned cases in month $t-1$. The column "Avg. in Month t" contains the average day 2 call behavior in month t. AI callers' average call behavior is reported in the last row of each panel. The columns "Avg. Diff." and "NPV2 Diff." contain the difference in average call behavior or collected scaled payments between the medium or high tercile relative to the lowest tercile. The borrower sample is the 10% completely randomized subsample. Panel A considers all phone calls and all borrowers, while Panel B restricts the sample to borrowers who make a promise to repay on day 2. Numbers in parentheses are t-statistics. *** p<0.01, ** p<0.05, * p<0.1.

Panel A. Phone call frequency.

	# Phone Calls (Per Borrower-Day)			Average Time Interval Between Calls (Minutes)		
Human Tercile in Month $t-1$	Avg. in Month t	Avg. Diff. (t-stat.)	NPV2 Diff. (t-stat.)	Avg. in Month t	Avg. Diff. (t-stat.)	NPV2 Diff. (t-stat.)
1 (Low)	6.62			94.35		
2 (Med)	6.82	0.20***	-0.013**	99.18	4.83***	-0.013**
		(4.41)	(-2.38)		(4.99)	(-2.37)
3 (High)	6.98	0.36***	0.009	105.39	11.04***	-0.016***
		(7.97)	(1.63)		(11.28)	(-2.81)
AI	4.72			114.27		

Panel B. Follow-up frequency within the same day, conditional on a promise to repay on day 2.

	% Follow-Up Within the Same Day			Average Time Interval to Follow-Up After Receiving a Promise (Minutes)		
Human Tercile in Month $t-1$	Avg. in Month t	Avg. Diff. (t-stat.)	NPV2 Diff. (t-stat.)	Avg. in Month t	Avg. Diff. (t-stat.)	NPV2 Diff. (t-stat.)
1 (Low)	0.862			88.38		
2 (Med)	0.878	0.016*	0.001	94.61	6.23**	-0.003
		(1.73)	(0.17)		(2.24)	(-0.54)
3 (High)	0.881	0.019**	0.006	107.16	18.78***	-0.009*
		(2.09)	(1.05)		(6.69)	(-1.67)
AI	0.825			215.70		

Table 8. Probability of delinquency on subsequent monthly payments.

This table shows borrowers' probability of missing subsequent monthly payments after they are initially assigned to the treatment group (five days of AI followed by humans) or the control group (always human) in the completely randomized subsample. Subsequent monthly payments due are identified by the number of months between the payment in question and the month borrowers entering the completely randomized subsample. Columns 2 and 3 report the fractions of borrowers who miss the *p*th payment in the treatment and the control groups, respectively. Column 4 reports the difference (AI minus human) and column 5 the corresponding *t*-statistic. The sample for the *p*th payment only includes loans with at least *p* months of payments between the month of entry into the completely randomized subsample and the loan's maturity. The number of such loans are reported in the last column. *** p<0.01, ** p<0.05, * p<0.1.

(1) Subsequent Monthly Payment	(2) AI	(3) Human	(4) Diff.: AI - Human	(5) *t*-stat.	(6) # Loans
1	0.627	0.604	0.023***	8.99	143,482
2	0.601	0.579	0.022***	8.48	139,212
3	0.590	0.570	0.020***	7.23	133,979
4	0.575	0.556	0.019***	7.02	129,021
5	0.563	0.545	0.018***	6.15	122,279
6	0.558	0.539	0.019***	6.42	116,723
7	0.545	0.529	0.016***	5.45	111,998
8	0.535	0.517	0.018***	5.81	105,728
9	0.527	0.511	0.016***	5.11	98,268
10	0.513	0.499	0.014***	4.17	89,934
11	0.505	0.494	0.011***	2.99	80,074
12	0.506	0.492	0.014***	3.64	72,829

Table 9. Day of the week of first call and debt collection outcomes.

This table compares borrowers who are first called by humans on weekends versus weekdays in the completely randomized subsample. Columns 1 and 2 contain average outcomes of borrowers first contacted on weekends and weekdays, respectively. Column 3 is the difference between the two types of debts, and column 4 reports the corresponding *t*-statistic, based on linear regressions of the variable of interest onto a weekend indicator and week fixed effects. Standard errors are adjusted for clustering at the calendar month level. Columns 5 and 6 re-estimate the outcome differences controlling for overdue payment amount, remaining principal, internal credit score, gender, age, and attainment of a bachelor's degree measured on day 2 after the due date. *** p<0.01, ** p<0.05, * p<0.1.

	(1) Weekend Average	(2) Weekday Average	(3) Diff: (1) - (2)	(4) *t*-stat.	(5) Diff: (1) - (2) w/ Covar.	(6) *t*-stat.
Call outcomes on day 2:						
% Calls answered	0.233	0.238	-0.005	-1.33	-0.004	-1.17
# Calls per borrower	5.541	5.431	0.110**	2.22	0.115**	2.78
# Calls answered	0.996	1.001	-0.005	-0.25	0.000	0.00
Call duration (sec)	72.42	72.06	0.42	0.22	1.02	0.60
% Promise to repay	0.669	0.665	0.003	0.38	0.005	0.56
Cumulative NPV collected:						
2 days	0.257	0.272	-0.015***	-3.86	-0.014***	-3.92
3 days	0.375	0.391	-0.017***	-3.86	-0.016***	-4.24
4 days	0.453	0.464	-0.011**	-2.64	-0.011**	-2.71
5 days	0.507	0.512	-0.005	-1.09	-0.004	-0.98
6 days	0.555	0.555	0.000	0.10	0.001	0.26
10 days	0.633	0.635	-0.002	-0.46	-0.001	-0.38
15 days	0.681	0.681	-0.000	-0.03	0.000	0.03
30 days	0.741	0.742	-0.001	-0.26	-0.001	-0.3
60 days	0.774	0.775	-0.001	-0.38	-0.001	-0.51
90 days	0.790	0.790	0.000	0.04	-0.000	-0.07
180 days	0.803	0.803	-0.001	-0.15	-0.001	-0.27
360 days	0.808	0.809	-0.001	-0.27	-0.001	-0.39

Table 10. Caller working experience and debt collection outcomes.

This table compares borrowers first called by senior versus junior callers on day 2. The sample is the completely randomized subsample restricted to human callers specializing in debts in the first 5 days past due. In Panel A, Columns 1 and 2 contain average characteristics of debts first contacted by the two types of callers. Column 3 shows the difference between the two, and column 4 contains the corresponding *t*-statistics. The estimations are based on linear regressions of the variable of interest onto a senior-caller indicator with month fixed effects, clustering standard errors at the month level. Columns 5 and 6 re-estimate the outcome differences additionally controlling for overdue payment amount, remaining principal, internal credit score, gender, age, or attainment of a bachelor's degree measured on day 2. Panel B further splits junior callers into two groups according to their working experiences, and reports differences relative to senior callers that do not control for additional covariates. *** p<0.01, ** p<0.05, * p<0.1.

Panel A. Debt collection outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Junior Avg.	Senior Avg.	Diff: (1) - (2)	*t*-stat.	Diff: (1) - (2) w/ Covar.	*t*-stat.
Cumulative NPV collected:						
2 days	0.241	0.250	-0.009	-1.66	-0.011**	-2.35
3 days	0.363	0.368	-0.005	-0.74	-0.008	-1.25
4 days	0.439	0.441	-0.002	-0.34	-0.005	-0.93
5 days	0.492	0.490	0.002	0.39	-0.001	-0.20
6 days	0.539	0.537	0.002	0.43	-0.001	-0.15
10 days	0.622	0.619	0.003	0.64	0.001	0.19
30 days	0.728	0.730	-0.002	-0.33	-0.003	-0.56
60 days	0.763	0.762	0.001	0.20	0.000	0.02
90 days	0.777	0.779	-0.002	-0.52	-0.003	-0.70
180 days	0.791	0.793	-0.002	-0.55	-0.003	-0.69
360 days	0.796	0.798	-0.002	-0.58	-0.003	-0.71

Panel B. Splitting the group of junior callers.

	(1)	(2)	(3)	(4)	(5)
Working Experience:		0 to 1 month		2 to 3 months	
	Senior Avg.	Diff.	*t*-stat.	Diff.	*t*-stat.
Cumulative NPV collected:					
2 days	0.250	-0.015***	-2.91	-0.009	-1.59
3 days	0.368	-0.011*	-1.75	-0.006	-0.82
4 days	0.441	-0.013**	-2.17	-0.001	-0.15
5 days	0.490	-0.006	-0.92	0.002	0.28
6 days	0.537	-0.005	-0.68	0.001	0.20
10 days	0.619	-0.002	-0.36	0.002	0.42
30 days	0.730	-0.007	-1.24	-0.001	-0.18
60 days	0.762	-0.005	-0.94	0.003	0.71
90 days	0.779	-0.008	-1.59	-0.000	-0.08
180 days	0.793	-0.006	-1.43	-0.001	-0.25
360 days	0.798	-0.006	-1.64	-0.001	-0.22

Table 11. Collected NPV difference between AI and humans after one day of AI treatment.

This table examines how only one day of AI calls on day 2 followed by human calls affects cumulative loan repayment NPV over various horizons. The treatment group is a subset of cases in the conditionally randomized subsample whose initial overdue amount exceeds 400 yuan and whose internal credit score is in deciles 4 to 7. The sample period is February, March, May, and June 2022. The control group consists of cases with the same ranges of size and internal credit score in the completely randomized subsample that are always assigned to human callers in the same time periods. Column 1 reports the cumulative NPV differences between treatment and control. Columns 3 and 5 split the sample by internal credit score, and columns 4 and 6 report the corresponding *t*-statistics. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
			Low Score: 4 & 5		High Score: 6 & 7	
Horizon	Difference	*t*-stat.	Difference	*t*-stat.	Difference	*t*-stat.
2 days	-0.088***	-13.50	-0.065***	-6.71	-0.102***	-11.76
3 days	-0.048***	-5.89	-0.044***	-3.50	-0.049***	-4.72
4 days	-0.033***	-3.95	-0.035***	-2.61	-0.031***	-2.95
5 days	-0.026***	-3.07	-0.032**	-2.38	-0.021**	-2.03
6 days	-0.022***	-2.65	-0.034**	-2.49	-0.015	-1.42
8 days	-0.022***	-2.72	-0.033**	-2.48	-0.015	-1.52
10 days	-0.016**	-2.05	-0.029**	-2.21	-0.008	-0.87
15 days	-0.012*	-1.66	-0.022*	-1.78	-0.007	-0.72
30 days	-0.010	-1.39	-0.019	-1.62	-0.004	-0.50
60 days	-0.009	-1.35	-0.020*	-1.81	-0.002	-0.30
90 days	-0.008	-1.33	-0.019*	-1.81	-0.002	-0.28
180 days	-0.010	-1.64	-0.023**	-2.21	-0.003	-0.37
360 days	-0.008	-1.31	-0.021**	-2.04	-0.001	-0.09

Table 12. Performance of different versions of AI callers.

This table reports the cumulative collected NPV differences between consecutive versions of AI callers at the horizons of 2-6, 8, and 10 days past due. The differences are estimated by linear regressions of collected NPV onto indicators for which AI version the case was assigned to and calendar month fixed effects. The last row reports the sample-wide average NPV collected at each horizon. *t*-statistics are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

	NPV Horizon (days past due)						
Version Diff.	2	3	4	5	6	8	10
V2 - V1	-0.0052	0.0084	0.0063	0.0145	0.0064	0.0040	0.0071
	(-0.53)	(0.68)	(0.51)	(1.17)	(0.49)	(0.32)	(0.57)
V3 - V1	0.0127*	-0.0041	-0.0060	-0.0018	-0.0109	-0.0119	-0.0144
	(1.74)	(-0.46)	(-0.65)	(-0.2)	(-1.12)	(-1.29)	(-1.58)
V4 - V1	0.023***	0.0242**	0.0234**	0.0303***	0.0118	0.005	-0.0013
	(2.67)	(2.34)	(2.17)	(2.71)	(1.05)	(0.46)	(-0.12)
V5 - V4	0.0218*	0.0133	0.0072	-0.0047	0.0045	0.01	0.0052
	(1.92)	(1.04)	(0.49)	(-0.33)	(0.33)	(0.74)	(0.37)
V6 - V5	-0.001	-0.0034	0.0062	0.0043	0.0119	0.0144	0.0163*
	(-0.14)	(-0.38)	(0.75)	(0.45)	(1.31)	(1.56)	(1.91)
Average NPV	0.180	0.279	0.342	0.389	0.490	0.570	0.612

Table 13. Human caller performance on day 6 after AI callers were upgraded to V4.

This table examines the impact of the AI caller upgrade from V1 to V4 on human callers' performance on day 6 past due. The sample of cases is restricted to the completely randomized subsample in September and October 2021. The sample cases are also required to remain unpaid on day 6. The sample of callers is restricted to callers specializing in cases 2-10 days past due. Columns 1 and 2 regress human caller performance on day 6 (i.e., NPV6 minus NPV5, denoted by "ΔNPV6") onto an indicator for being treated by V4 AI callers in the first five days and month fixed effects. Columns 3 and 4 regress callers' day-6 performance in the current month onto tercile dummies for their performance on the same type of cases in August 2021. Columns 5 and 6 interact the AI V4 dummy with the past performance tercile dummies. Numbers in parentheses are *t*-statistics clustered at the caller level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

	(1) ΔNPV6	(2) ΔNPV6	(3) ΔNPV6	(4) ΔNPV6	(5) ΔNPV6	(6) ΔNPV6
AI V4	-0.041***	-0.040***			-0.034	-0.031
	(-2.90)	(-2.88)			(-1.48)	(-1.42)
Prev Perf Med			0.035**	0.038**	0.039	0.040
			(2.12)	(2.28)	(1.42)	(1.51)
Prev Perf High			0.053***	0.050***	0.058**	0.057**
			(3.48)	(3.32)	(2.30)	(2.37)
AI V4 × Prev Perf Med					-0.007	-0.004
					(-0.21)	(-0.15)
AI V4 × Prev Perf High					-0.009	-0.014
					(-0.29)	(-0.46)
Constant	0.197***	0.332***	0.140***	0.284***	0.161***	0.306***
	(17.55)	(3.116)	(11.97)	(2.70)	(8.45)	(2.88)
AI V4 + AI V4 × Prev Perf Med:					-0.040*	-0.036
					(-1.71)	(-1.56)
AI V4 + AI V4 × Prev Perf High:					-0.043*	-0.045**
					(-1.86)	(-2.00)
Loan Char. Control	No	Yes	No	Yes	No	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	3,328	3,328	3,328	3,328	3,328	3,328
# Callers	544	544	544	544	544	544
R-squared	0.003	0.049	0.004	0.050	0.006	0.053

Table 14. Ex-post imbalance in assigned case difficulty and caller outcomes.

This table examines the relationship between ex-post imbalance in assigned case difficulty and worker outcomes. For each caller in each month, the ex-post imbalance measure is the fraction of debts assigned to the caller with internal credit score of 3 or lower. The sample is restricted to callers who specialize in debts in the first five days past the due date and who work for at least 20 days in the month. The table shows estimates from regressions of worker outcomes onto the imbalance measure with and without worker characteristics (working status classification as junior, senior, or returning callers who previously left the company; age; gender; and working experience in months) as covariates. The coefficients are for the imbalance measure, and the marginal effect of a 2% increase in imbalance is based on the regressions without covariates. All specifications include month fixed effects and are estimated by OLS except when the dependent variable is a "quit next month" indicator, in which chase we estimate with logit regressions. Numbers in parentheses are *t*-statistics clustered at the month level. *** p<0.01, ** p<0.05, * p<0.1.

	Sample Average	Coefficients on the Imbalance Measure		Marginal Effect for a +2% Imbalance (Without Covar.)
		w/o Covar.	w/ Covar.	
Repayment rate	0.183	-0.230***	-0.223***	-0.005
		(-5.07)	(-5.14)	
Performance ranking	0.504	-5.164***	-4.927***	-0.103
		(-6.36)	(-6.63)	
Total salary	4464.30	-23868.8***	-22917.0***	-477.38
		(-5.46)	(-5.77)	
Retention rate at the end of …				
Current month *t*	0.890	-11.03**	-10.43*	-0.021
		(-2.01)	(-1.90)	
Month *t* + 1	0.857	-11.56*	-10.91*	-0.028
		(-1.89)	(-1.86)	
Month *t* + 2	0.843	-13.57*	-13.10*	-0.035
		(-1.86)	(-1.89)	
Month *t* + 3	0.829	-13.43*	-12.96*	-0.037
		(-1.66)	(-1.66)	

Online Appendix A. Additional Results on the Regression Discontinuity Design

1 Tests of manipulation at the threshold

An important assumption in a RD design is that agents do not exert precise control (Lee, 2008) over whether they are above or below the threshold. Since the company never discloses its debt collection assignment rules to the public, borrowers are unlikely to manipulate their remaining principal to avoid human callers or the opposite.

We validate the no-manipulation assumption by examining the distribution of observations around the cut-off. Figure A1 shows the results of the RD density test. Figure A1(a) tests for a discontinuity at the true cut-off of 300 yuan. The bars indicate the density of the running variable, remaining principal. Since the assignment rule is right-continuous, the intervals include their right ends but not their left ends. Therefore, 300-yuan cases belong to the right-most bar to the left of the cut-off. The histogram shows that there is an increase in observation density just below the cut-off. We believe that this is because borrowers and lenders tend to round to multiples of 100 yuan: lenders may want to issue loan amounts with a 100-yuan step size, and borrowers may prefer to keep a balance at an integer multiple of 100 yuan when repaying their principal. Despite such a tendency, the density function (solid lines) estimated by local quadratic regressions shows no significant jump at the cut-off; the robust RD *t*-statistic for the discontinuity is only -1.01.

The tendency to round is also observed at 200 yuan and 400 yuan, as shown in Panels (b) and (c) in Figure A1. In placebo tests, we use an artificial cut-off of 200 or 400 yuan and perform the same test as in Panel (a). We find somewhat larger discontinuities in density at these artificial cut-offs, so the bunching at the 300-yuan cut-off is not abnormal. In addition, our continuity tests in Table 3 Panel A suggest that the tendency to round at 300 yuan is unrelated to observed loan and borrower characteristics.

Following Cattaneo et al. (2019), we also implement a binomial test at the cut-off. The test compares the number of observations just below versus just above the cut-off within a given symmetric neighborhood around the cut-off. If there is no manipulation at the cut-off, 50% of cases should be below the cut-off. The binomial test examines whether the empirical fraction is significantly different from 50%. Table A1 reports the results. For a neighborhood radius smaller than 2 yuan, there are significantly more cases less than or equal to 300 yuan. As we consider larger radii between 2.5 to 5 yuan around the cut-off, the distribution becomes balanced. This can be explained by the decreasing tendency to round to 300 yuan as people move further away from the cut-off. We therefore conclude that there is no intentional manipulation related to AI caller usage at the 300-yuan cut-off. In the following section, we show that our results are robust to excluding potentially rounded observations.

2 Robustness checks

Table A2 shows robustness checks of the RD regression results from varying the specification. As a benchmark, column 1 repeats our main results from Table 3 Panel B, which uses the MSE-optimal bandwidth and uniform kernel. Columns 2 and 3 change the kernel choice to a triangular kernel and an Epanechnikov kernel, respectively. Column 4 uses the CER-optimal bandwidth. Column 5 doubles the MSE-optimal bandwidth and column 6 shrinks it by half. These variations generate results similar to the main specification, confirming that our results are robust to bandwidth and kernel choices.

The last three columns in Table A2 show the results of a "donut-hole" test, which checks the robustness of our results with respect to observations close to the cut-off. This approach can evaluate the sensitivity of the results to manipulation, even if it is not suspected, as well as sensitivity to the unavoidable extrapolation in local linear regressions. In the test, observations within $\pm w$ of the cutoff are excluded before running the same RD regressions. We set w to be 0.5, 1, and 2—neighborhoods with potential rounding. The results are quite similar to the original ones, alleviating concerns about manipulation and rounding.

3 Placebo tests

Finally, Table A3 shows two placebo tests using the artificial cut-offs of 200-yuan and 400-yuan of remaining principal. We use CER-optimal bandwidths in these RD regressions, since they give the most power when making inferences about the null hypothesis that there is no jump in the outcome variables (Cattaneo et al., 2019). The results do not reject the null hypothesis for both artificial cut-offs and at all evaluation horizons.

References

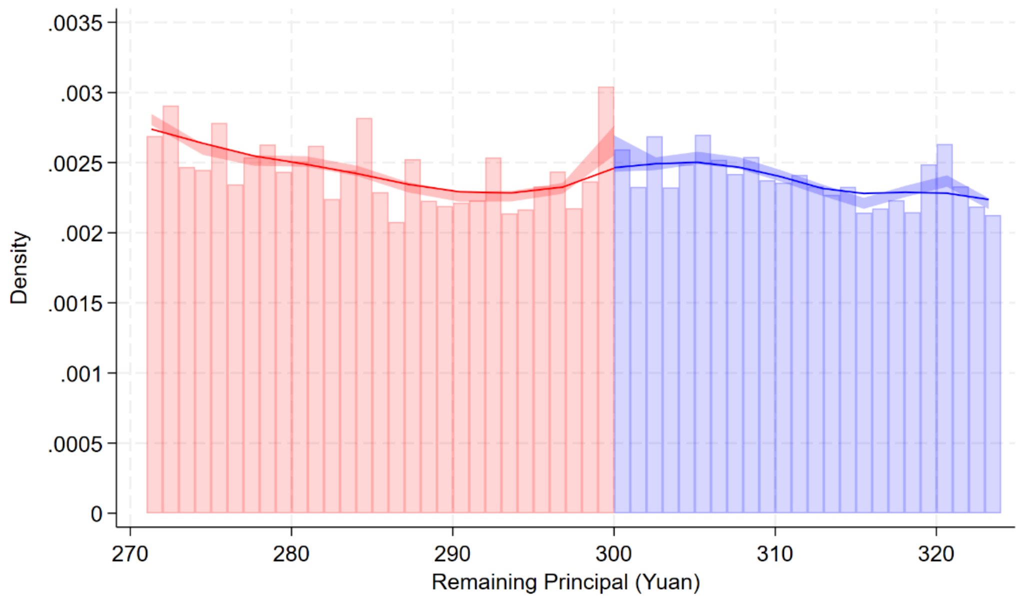
Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.

Lee, David S., 2008. "Randomized experiments from non-random selection in U.S. House elections." *Journal of Econometrics* 142: 675-697.

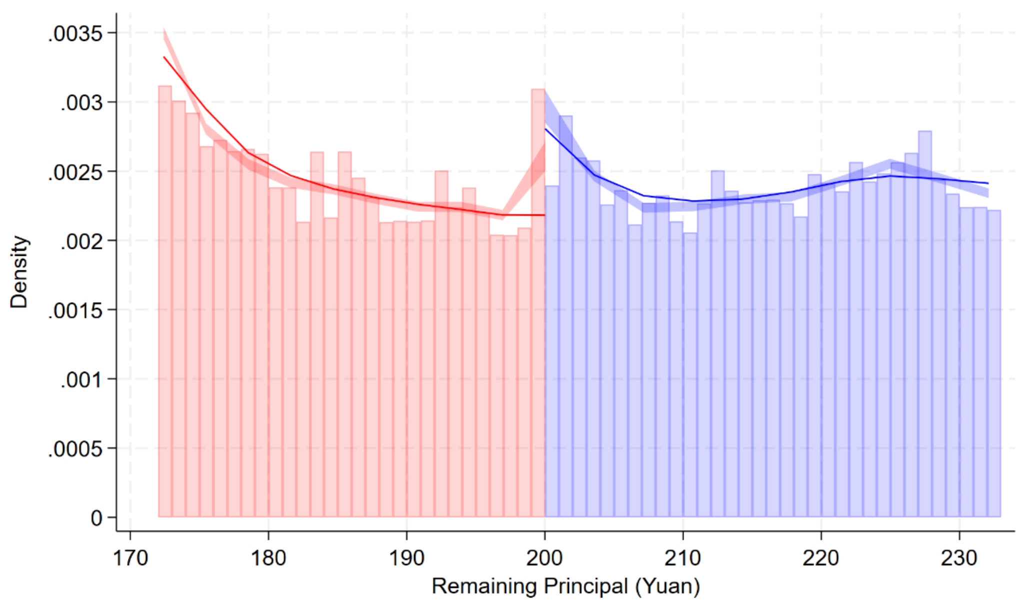
Figure A1. RD density test around the threshold.

This figure reports the results of the RD density test to detect potential manipulation around the threshold. The figure first shows the histogram of the running variable—remaining principal—around the threshold. Each interval in the histograms includes the right end but not the left end. We estimate the density functions on both sides of the threshold separately using local quadratic regressions, which are displayed by the solid lines. The shaded areas around the fitted lines mark the 95% robust RD confidence intervals using local cubic regressions. All local regressions use the triangular kernel with the IMSE-optimal bandwidth.

(a) 300-yuan threshold.



(b) Placebo test: 200-yuan threshold.



(c) Placebo test: 400-yuan threshold.

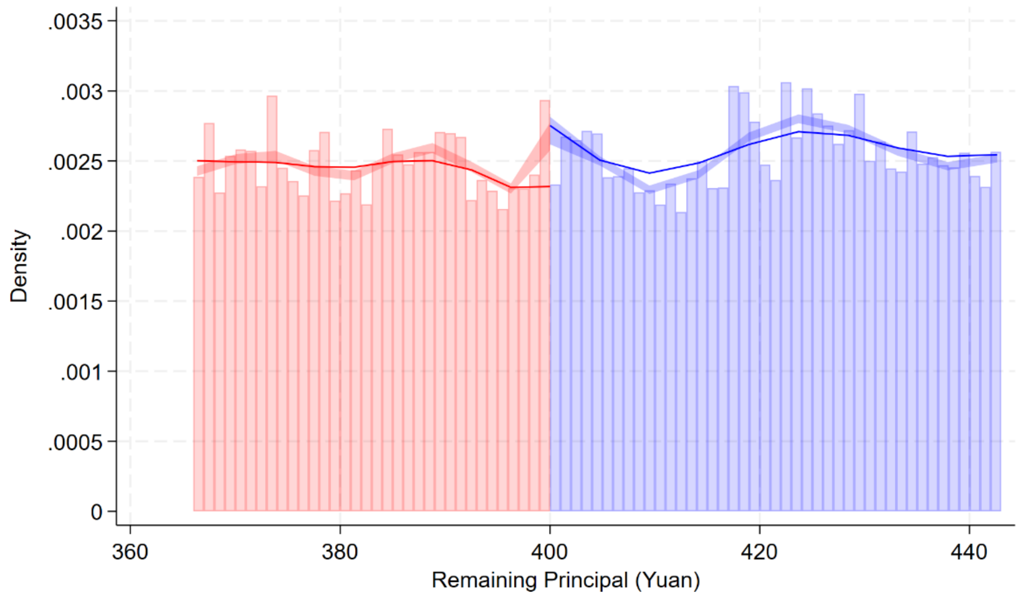


Table A1. Binomial test of manipulation at the threshold.

This table reports the results from the Binomial test of manipulation at the threshold. For a neighborhood of width $2x$ around the 300-yuan cut-off, the test counts the numbers of observations below and above the threshold and calculates the fraction of observations below the threshold. When there is no manipulation at the threshold, the null hypothesis holds that the fraction of observations below the threshold is 0.5, so the distribution of observations on both sides of the threshold can be considered as random.

Neighborhood Radius x	# Obs. in (300-x,300]	# Obs. in (300,300+x]	% Below	p-val.
0.5	2,042	1,317	60.8%	<0.001
1.0	3,075	2,644	53.8%	<0.001
1.5	4,365	3,739	53.9%	<0.001
2.0	5,466	4,996	52.2%	<0.001
2.5	6,567	6,454	50.4%	0.326
3.0	7,663	7,699	49.9%	0.778
3.5	8,969	8,848	50.3%	0.369
4.0	10,125	10,060	50.2%	0.652
4.5	11,365	11,180	50.4%	0.220
5.0	12,480	12,569	49.8%	0.578

Table A2. Collected NPV difference between AI and human callers: Robustness checks.

This table shows robustness checks on the RD design regression results in Table 3 Panel B, which estimates the average difference in collected NPV between AI and human callers at horizons of 2 to 360 days. The first column reports the main results, which are the same as the results in Table 3 Panel B. Columns 2 and 3 change the kernel. Columns 4 to 6 modify the bandwidth. Columns 7 to 9 exclude observations within ±*w* of the cutoff. "MSE" and "CER" stand for the optimal bandwidths that minimize the mean squared error and the coverage error probability, respectively. "Epan." stands for Epanechnikov kernel. RD robust *z*-statistics are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Main setup	Kernel Choice		Bandwidth Choice			Donut-Hole		
							$w = 0.5$	$w = 1$	$w = 2$
NPV 2d	-0.0375***	-0.0382***	-0.0407***	-0.0359***	-0.0449***	-0.0355***	-0.0406***	-0.0442***	-0.0386***
	(-9.02)	(-10.97)	(-12.33)	(-6.79)	(-8.53)	(-2.70)	(-11.00)	(-12.02)	(-9.55)
NPV 5d	-0.0500***	-0.0501***	-0.0509***	-0.0462***	-0.0500***	-0.0461***	-0.0588***	-0.0611***	-0.0600***
	(-7.77)	(-8.15)	(-8.29)	(-5.72)	(-8.24)	(-2.30)	(-9.53)	(-9.55)	(-8.30)
NPV 10d	-0.0777***	-0.0767***	-0.0769***	-0.0760***	-0.0851***	-0.0760***	-0.0812***	-0.0834***	-0.0786***
	(-12.73)	(-12.42)	(-12.54)	(-9.70)	(-12.77)	(-6.09)	(-12.81)	(-12.77)	(-10.59)
NPV 30d	-0.0915***	-0.0892***	-0.0893***	-0.0908***	-0.105***	-0.0911***	-0.0921***	-0.0924***	-0.0938***
	(-18.40)	(-15.97)	(-15.95)	(-14.23)	(-17.64)	(-9.61)	(-18.40)	(-18.28)	(-18.06)
NPV 60d	-0.0751***	-0.0732***	-0.0728***	-0.0758***	-0.0845***	-0.0764***	-0.0746***	-0.0749***	-0.0726***
	(-17.24)	(-14.55)	(-14.36)	(-13.72)	(-17.79)	(-10.76)	(-15.47)	(-14.76)	(-13.57)
NPV 90d	-0.0647***	-0.0660***	-0.0655***	-0.0674***	-0.0761***	-0.0675***	-0.0660***	-0.0642***	-0.0637***
	(-16.06)	(-13.66)	(-13.44)	(-13.39)	(-18.32)	(-10.70)	(-14.34)	(-13.79)	(-12.66)
NPV 180d	-0.0524***	-0.0541***	-0.0534***	-0.0551***	-0.0632***	-0.0552***	-0.0523***	-0.0514***	-0.0510***
	(-13.84)	(-11.68)	(-11.41)	(-11.79)	(-16.72)	(-9.69)	(-11.81)	(-11.65)	(-10.60)
NPV 360d	-0.0464***	-0.0486***	-0.0477***	-0.0489***	-0.0568***	-0.0492***	-0.0464***	-0.0447***	-0.0440***
	(-12.59)	(-10.79)	(-10.45)	(-10.77)	(-15.40)	(-9.61)	(-10.78)	(-10.33)	(-9.17)
Bandwidth	MSE	MSE	MSE	CER	2 × MSE	½ × MSE	MSE	MSE	MSE
Kernel	Uniform	Triangular	Epan.	Uniform	Uniform	Uniform	Uniform	Uniform	Uniform

Table A3. Collected NPV difference between AI and human callers: Placebo tests.

This table reports placebo test results estimating the average difference in collected NPV across the artificial cut-offs of 200-yuan and 400-yuan of remaining principal at horizons of 2 to 360 days. The specifications of the RD regression are the same as those in Table 3 Panel B. RD robust *z*-statistics estimated by local quadratic regressions are in parentheses. Since the point estimates and the *z*-statistics are estimated by local linear and quadratic regressions, respectively, their signs can differ. "MSE" stands for the optimal bandwidth that minimizes the mean squared error. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)
	Artificial cutoff c	
	c = 200	c = 400
NPV 2d	-0.0060	0.0064
	(-0.76)	(1.33)
NPV 5d	-0.0030	0.0050
	(-0.63)	(1.00)
NPV 10d	-0.0027	0.0009
	(-0.37)	(-0.04)
NPV 30d	-0.0099	-0.0002
	(-1.92)	(-0.01)
NPV 60d	-0.0002	0.0004
	(-0.37)	(-0.17)
NPV 90d	0.0015	-0.0002
	(0.02)	(-0.35)
NPV 180d	0.0042	-0.0002
	(0.55)	(-0.23)
NPV 360d	0.0049	-0.0013
	(0.67)	(-0.46)
Bandwidth	MSE	MSE
Kernel	Uniform	Uniform

Online Appendix B. Estimation of Unit Labor Costs

1 Caller salary scheme

Individual callers' monthly salary consists of two components: ranking salary and completion salary. Both components are determined by the caller's monthly performance, measured as the total amount of outstanding balance collected.

Ranking salary is based on the caller's performance rank among a group of callers who have similar tenures and workloads and who work in the same stage of the debt collection process. The relationship between rank and ranking salary is increasing and convex. Figure B1 shows performance ranking salary as a function of caller rank in May 2022 for callers specializing in borrowers who are 2-10 days past due. Ranking salary has six tiers, and the salary within each tier changes linearly. The function has the steepest slope in Tier 1. The top caller receives 5,500 yuan as ranking salary while the lowest 5% of callers receive nothing. In addition, the company may divide callers into several groups and encourage competition between groups to maintain callers' morale. The winning group can receive from 100 yuan to a few hundred yuan as an extra ranking bonus.

Completion salary is determined by the amount of money the caller collects scaled by a pre-specified target. At the beginning of each month, the company sets a target collection amount for each caller based on the predicted total outstanding balance the company may need to deal with, as well as the number of callers and caller working experience. Junior callers can have a 10% lower target in their first four months with the company. At the end of the month, the target completion rate is calculated as the ratio of the actual amount collected by the caller to her target. Completion salary is an increasing piecewise linear function of completion rate. Figure B2 illustrates the relationship between completion rate and completion salary that the company applied in May 2022. The target amount was 448,526 yuan in May 2022. The completion salary jumps at the completion rates 0.7, 0.8, 0.9, 0.95, and 1. The slope also slightly increases with the completion rate across the intervals. Callers can earn more than 3,500 yuan if they achieve the target and nothing if they collect less than 70%. The average completion rate is 1.01.

Finally, the company also has a minimum wage of around 3,000 yuan per month, varying slightly with time and employment location. If the sum of a caller's ranking salary and completion salary is below the minimum wage, the company will pay the caller the minimum wage.

Figure B3 presents the salary amount that callers specializing in days 2-10 received in May 2022. Figure B3(a) plots ranking salary paid as a function of callers' performance ranking. The shape of the curve closely tracks the formula in Figure B1, with small variations that reflect extra bonuses from group-level performance competitions. Figure B3(b) shows completion salary paid as a function of the completion rate, which precisely follows the formula in Figure B2.

Finally, Figure B3(c) presents the relationship between overdue money collected and total salary paid, which is the sum of ranking and completion salary after some adjustments that we will discuss shortly. The upper "surface" of the scatter dots equals the theoretical maximum salary that callers can receive given the amount collected. It is upward sloping above 3,000 yuan, the minimum wage.[1] The slope is about 0.045 yuan of salary per one yuan collected. In practice, callers typically receive a salary below the theoretical maximum for several reasons, including penalties for absence from work or late arrivals, for example.

The most significant penalty is for violation of rules regarding conversations with overdue borrowers. To comply with government regulations and to maintain a positive image with the public, the company has several rules about what callers cannot say to borrowers. Prohibitions include swear words, threats, discrimination, false information, and unwarranted promises to borrowers. The company uses an AI examiner to go through all phone call records and identify misconduct every month. For each caller in each month, the company calculates a "quality control (QC) ratio" defined as the fraction of appropriate conversations. The actual salary that a caller receives is the theoretical maximum multiplied by the QC ratio. The average QC ratio is about 0.953, but 10% of callers have a QC ratio below 0.87.

This quality adjustment helps explain why, although there is a jump in completion salary at a 100% completion rate (which corresponds to about 450,000 yuan of money collected), total salary paid has no significant jump at the cut-off. This is because, to exceed the target, callers just below the threshold tend to violate the rules more. Therefore, despite receiving a jump in completion salary for crossing the threshold, they are penalized by a low QC ratio, leading to their receiving total salaries that are not discretely higher than if they were just below the threshold.

[1] Some callers received salaries lower than the minimum wage or even zero salaries because they left the company in the middle of the month and received only partial salaries proportional to the actual number of working days or borrowers contacted, minus penalty deductions as explained in the text.

2 Estimation of unit labor costs

For each individual caller, both salary components depend on performance. However, no matter how much money each caller collects, the ranking salary per caller is the same for the company. We estimate the fixed labor cost of calling one borrower for one minute by dividing the total ranking salary paid by the company to callers who specialize in borrowers who are 2-10 days overdue by the total minutes of phone calls made by these callers. We multiply the fixed costs per minute of phone calls by the average length of phone calls per borrower-day to obtain the fixed labor costs per borrower-day. Figure B4(a) shows the monthly time series of average ranking salary per minute of phone calls over time. We use the time-series average of 1.1565 yuan as our estimate of the fixed labor cost to talk to one borrower for one minute.

The completion salary can be viewed as a variable labor cost for the company, as it is related to the actual amount of money collected. To get the average variable salary that the company has to pay per yuan collected, we divide the total completion salary paid to the above group of callers by the total amount of money collected by them in each month. Figure B4(b) reports the monthly time series of this ratio. The average completion salary displays an increasing trend: the average was 0.004 yuan in June 2021 and was raised by 75% to 0.007 yuan in May 2022. We use the time-series average, which is about 0.0051 yuan per one yuan collected, as our estimate of the variable labor cost of collecting an additional yuan.

Figure B1. The relationship between ranking and ranking salary.

This figure visualizes the formula used in May 2022 to calculate an individual caller's ranking salary as a function of their performance ranking. Callers are ranked by their total money collected in a month within a group of callers in the same stage of debt collection and with similar tenure.

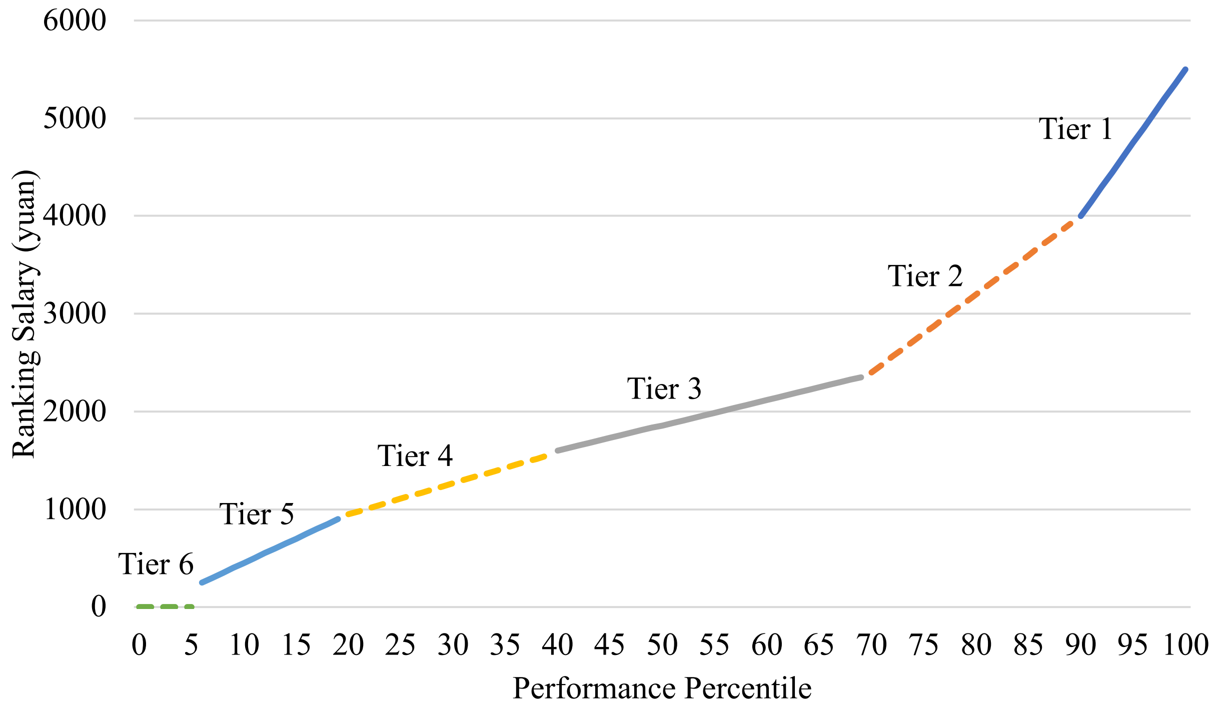


Figure B2. The relationship between completion rate and completion salary.

This figure visualizes the formula used to calculate an individual caller's completion salary as a function of their target completion rate in May 2022. The target completion rate is defined as the ratio of money collected in the month to a target of money to be collected specified by the company at the beginning of the month. The target amount was 448,526 yuan in May 2022.

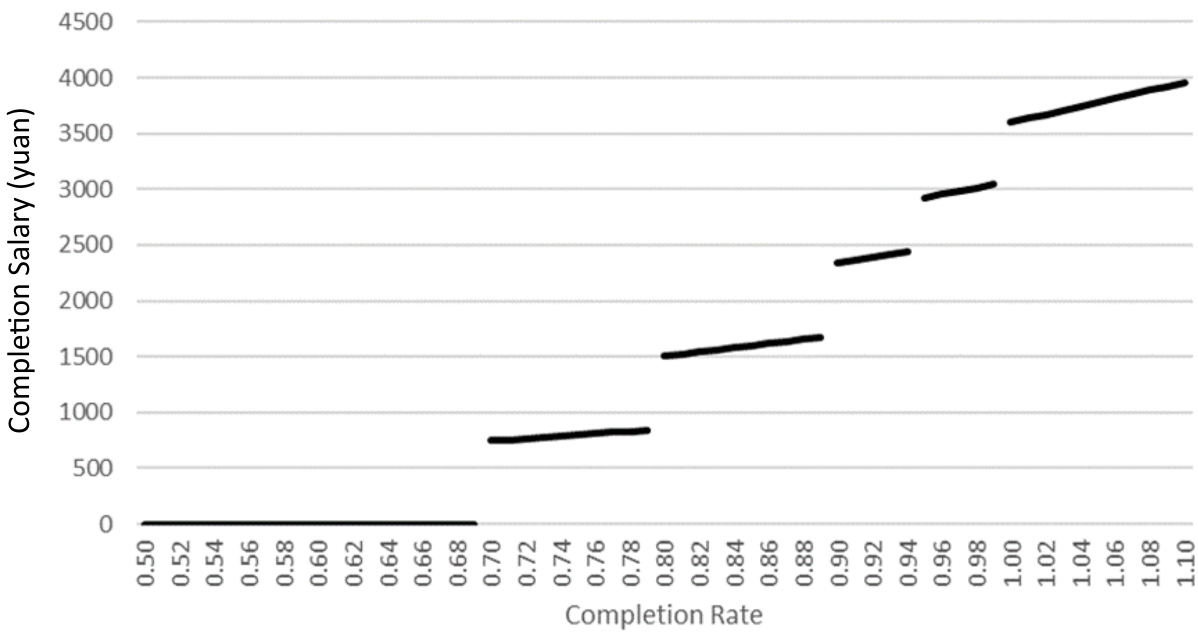
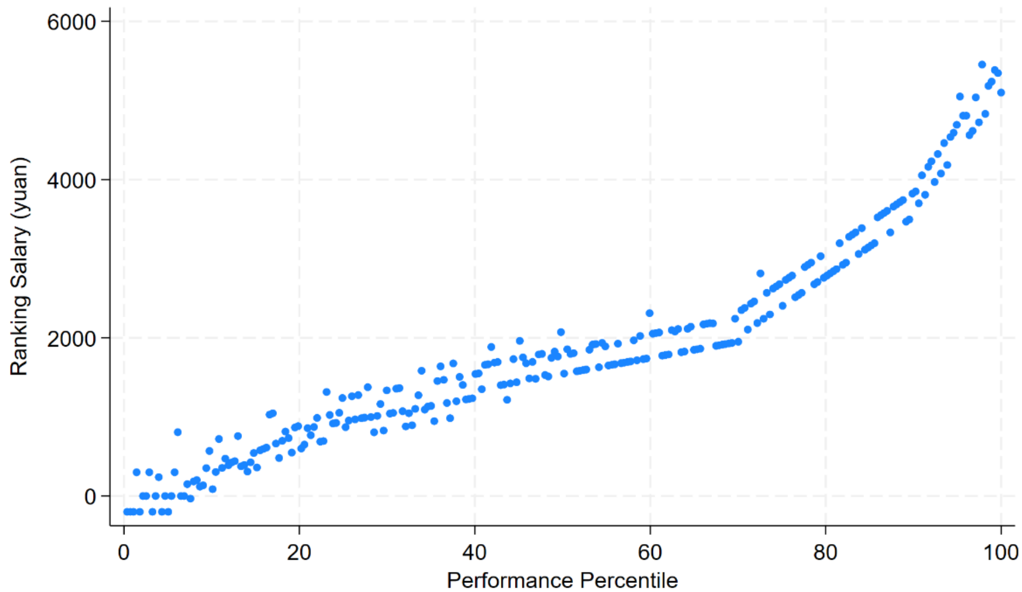


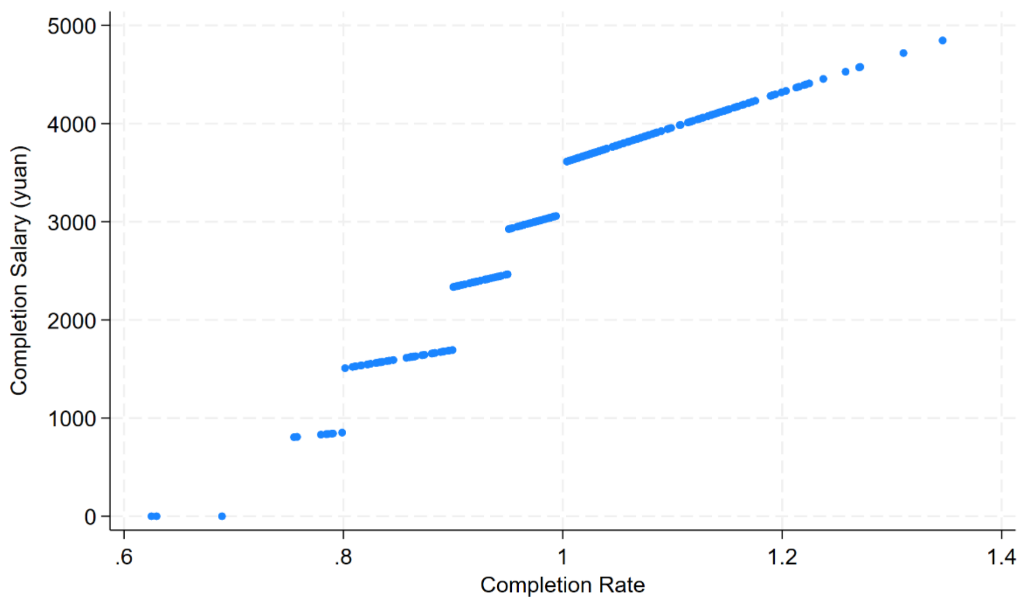
Figure B3. Actual salary received by callers.

This figure reports the actual salary received by senior callers in the "M1 Early" stage in May 2022. Panel (a) reports the actual ranking salary received by callers as a function of caller performance ranking. Panel (b) shows the actual completion salary received by callers as a function of their completion rate. Panel (c) shows the actual total salary received by callers as a function of the amount of money collected. The total salary is the sum of the ranking salary and completion salary, capped by the minimum wage, and adjusted for additional penalties and bonuses.

(a) Ranking salary as a function of ranking.



(b) Completion salary as a function of completion rate.



(c) Total salary as a function of money collected.

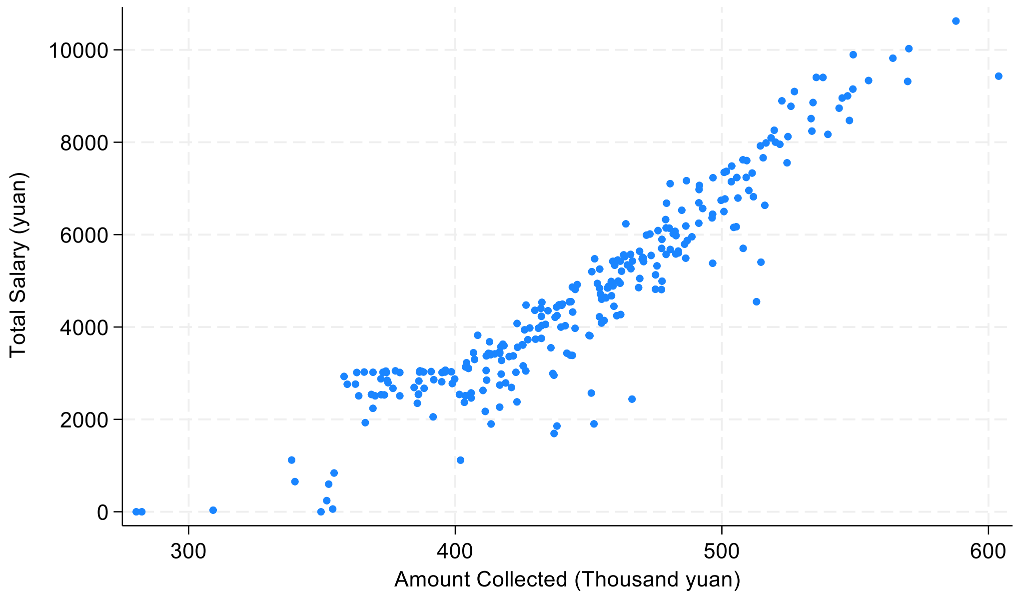
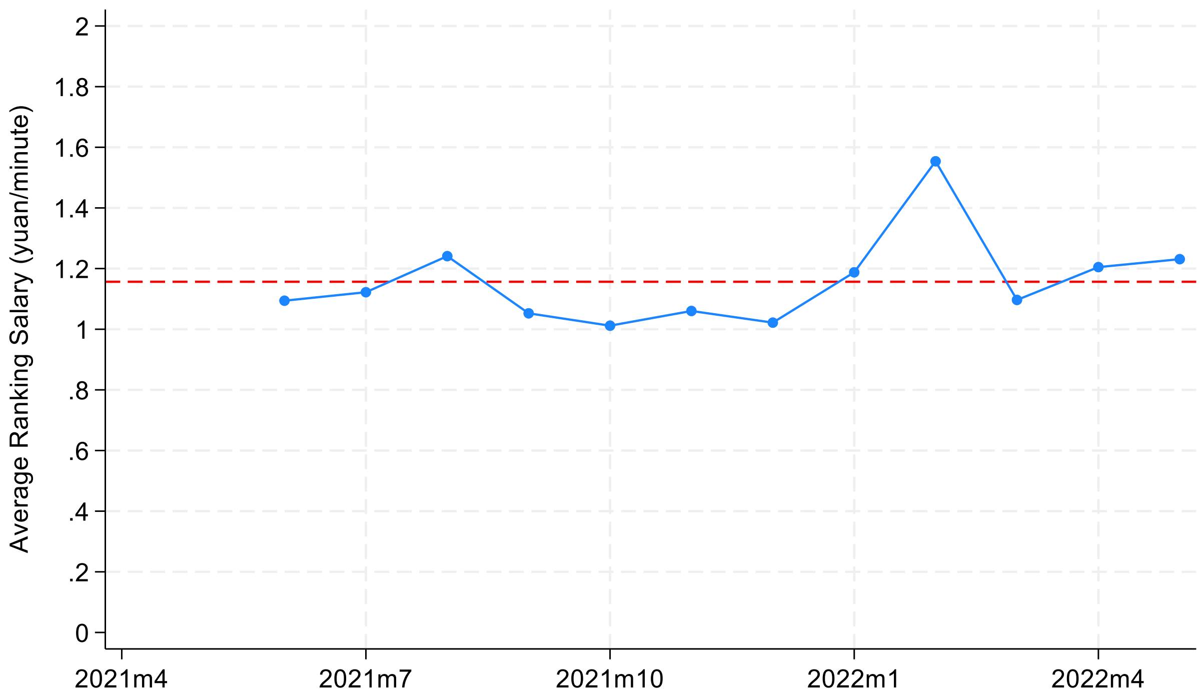


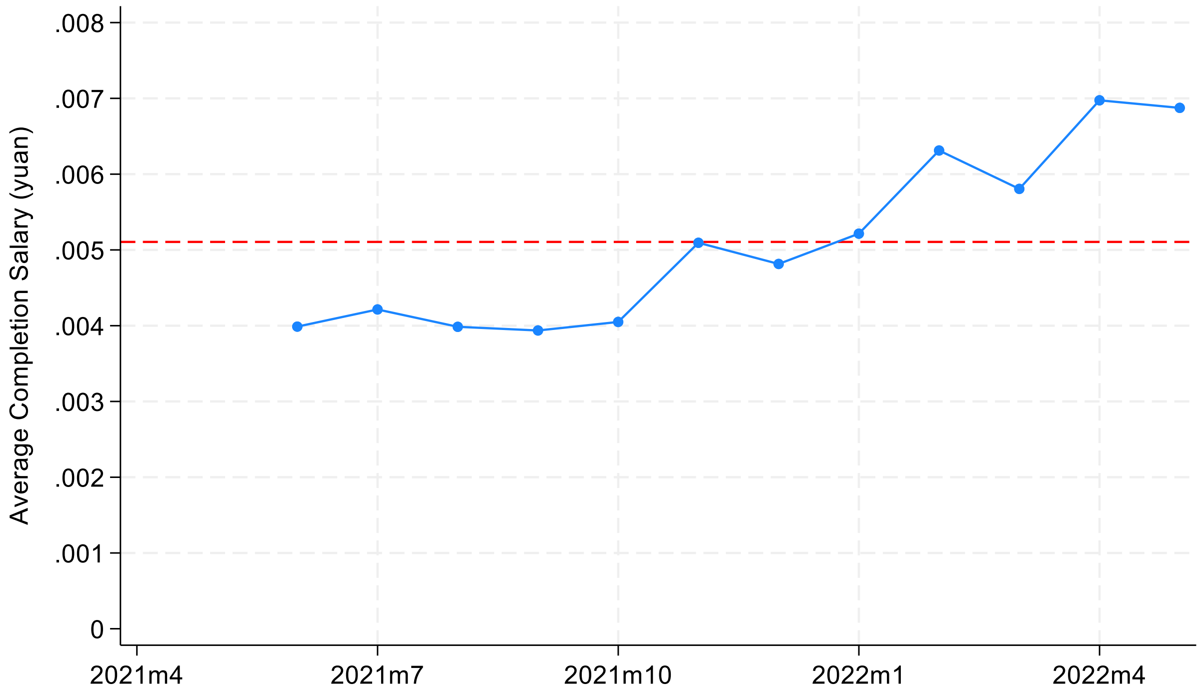
Figure B4. Time series of unit salary costs and unit workloads.

This figure shows the time series of unit salary costs and unit workloads among senior callers in the M1 Early stage. Panel (a) shows the ratio of total ranking salary to the total length of phone calls in every month, that is the average ranking salary paid by the company for handling one delinquent borrower for one minute, or equivalently, the unit fixed labor costs. Panel (b) reports the ratio between the total completion salary and the total money collected, that is, the unit variable labor costs.

(a) Ranking salary per minute of phone calls



(b) Completion salary per yuan of money collected



Online Appendix C. Additional Figures and Tables

Figure C1. Average phone call length per borrower-day by the number of days after delinquency.

This figure shows the total minutes human callers in our data spend talking to borrowers who are a given number of days past due, divided by the number of borrower-days in the data that correspond to that number of days past due. The spikes on days 6, 11, 26, and 60 occur when the stage of the debt collection process changes.

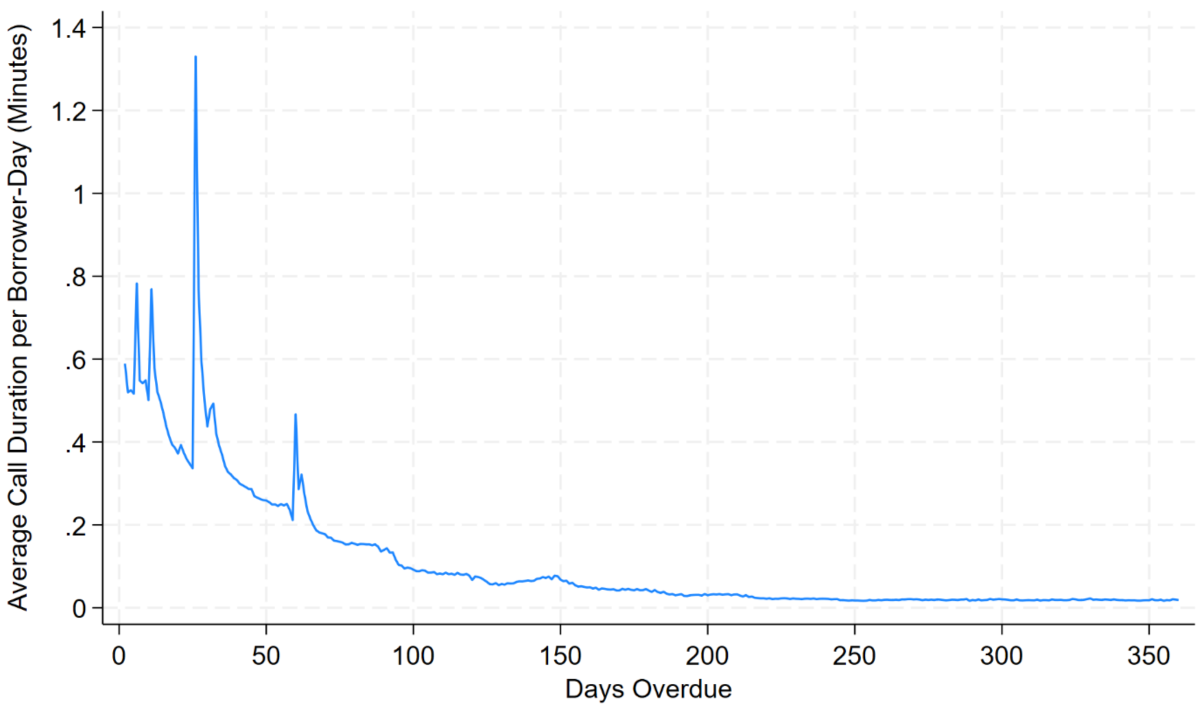


Figure C2. AI minus human caller cumulative collected undiscounted payments by horizon – small cases RD.

This figure shows the difference in the average cumulative undiscounted payments collected divided by balance due on day 2 between AI and human callers as a function of days past due. The differences are estimated by regression discontinuity around the 300-yuan remaining principal threshold for permanent AI treatment. As a benchmark, the graph also shows the difference in scaled NPVs collected, which are the same values shown in Figure 4 Panel (b). Differences are plotted every three days before day 60 and every 10 days after day 60.

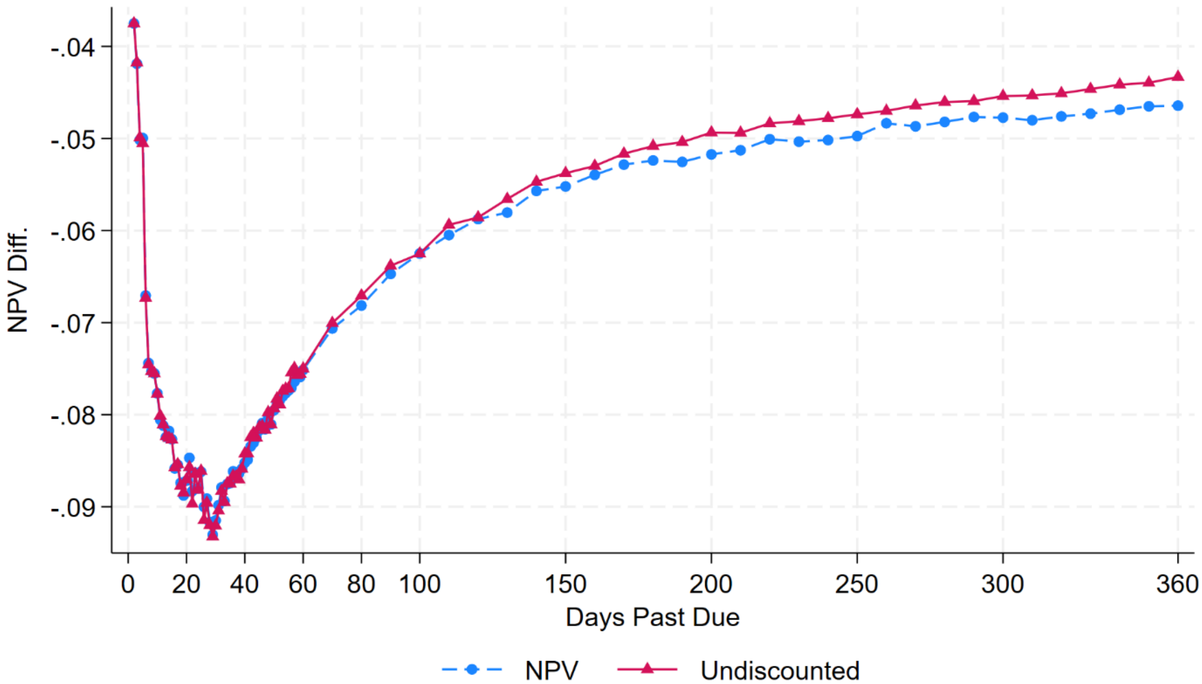


Figure C3. AI minus human caller cumulative collected undiscounted payments by horizon – Completely randomized subsample.

This figure shows the difference in average cumulative undiscounted payments collected divided by the balance due on day 2 between AI and human callers as a function of days past due, using the completely randomized subsample. The differences control for month-of-delinquency fixed effects. As a benchmark, the graph also shows the difference in scaled NPVs collected, which are the same values shown in Figure 6 Panel (a). Differences are plotted daily before day 30, and every 30 days afterwards.

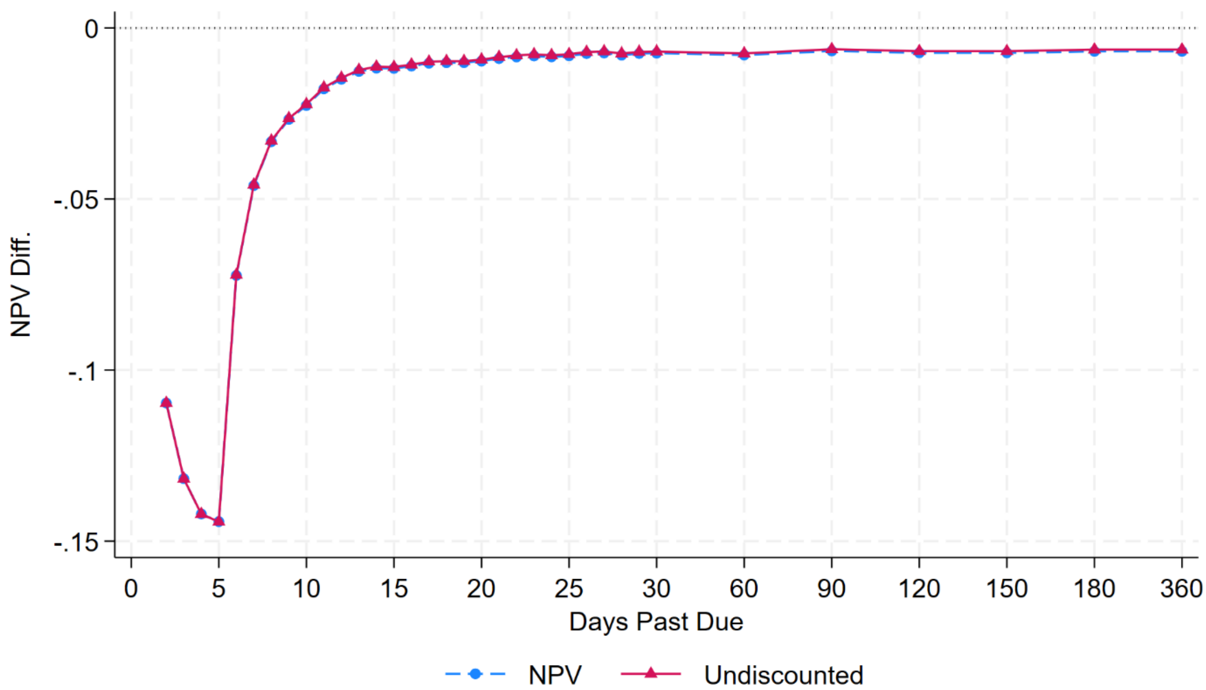


Figure C4. Distribution of phone call duration.

This figure is a histogram of phone call durations in seconds for the first answered phone call by each borrower on day 2 past due in the completely randomized subsample. Each bin is 10 seconds wide. Human caller observations are reweighted so that the distribution of the hour of day in which their calls were made is the same as that of AI calls.

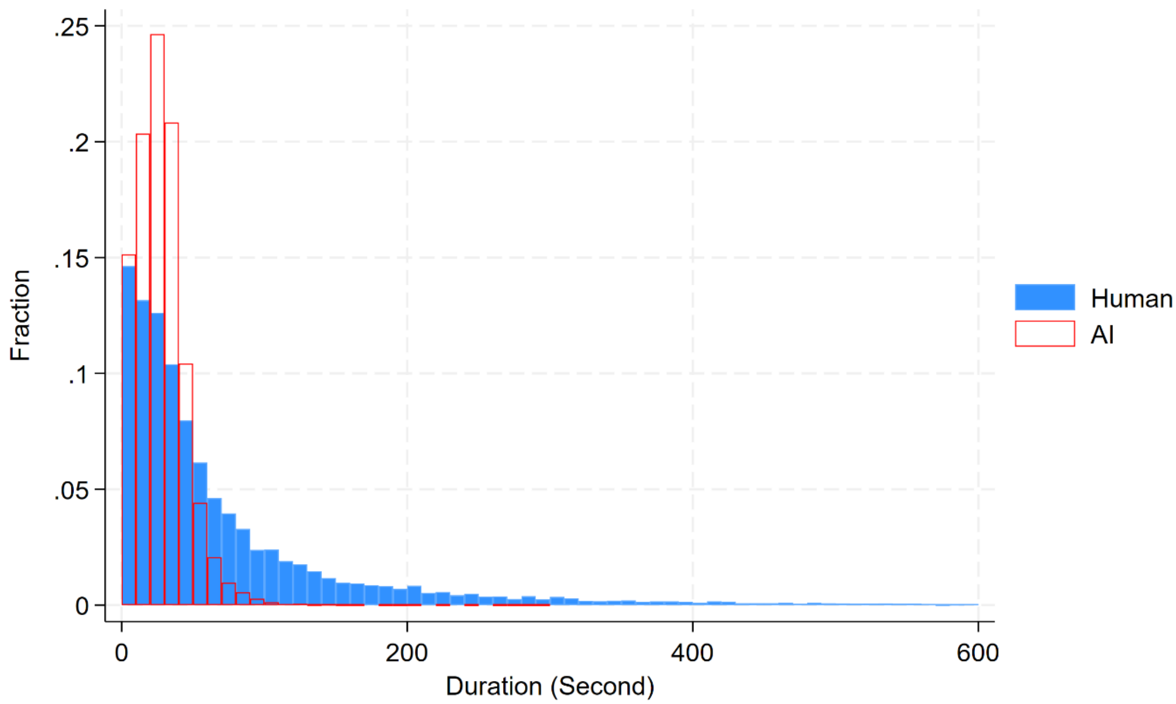


Table C1. Performance gaps between permanent AI and permanent human callers.

This table re-estimates Table 3 by restricting the cases on the right of the 300-yuan remaining principal threshold for a cleaner human caller benchmark. The table compares loan characteristics and performance of small cases assigned to AI callers almost permanently and to human callers by utilizing the 300-yuan remaining principal threshold using regression discontinuity design (RDD). The differences between AI and human callers are estimated with local linear regressions and the RD bias-corrected *z*-statistics are estimated with local quadratic regressions. Column 1 reiterates the main results from Table 3, column 2 restricts the cases on the right of the cutoff to the 10% completely randomized subsample, and column 3 further restricts the cases on the right of the cutoff to the 5% of cases in the completely randomized subsample that are always treated by humans starting from day 2. See the note in Table 3 for descriptions of other specifications. Significance level: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

	(1) Main results from Table 3 Panel B		(2) 10% completely randomized subsample		(3) 5% completely randomized subsample always treated by human	
Variable	Diff.	*z*-stat.	Diff.	*z*-stat.	Diff.	*z*-stat.
NPV 2d	-0.038***	-9.02	-0.046***	-4.22	-0.077***	-5.00
NPV 5d	-0.050***	-7.77	-0.064***	-4.14	-0.087***	-4.16
NPV 10d	-0.078***	-12.73	-0.083***	-5.91	-0.067***	-3.54
NPV 30d	-0.092***	-18.40	-0.107***	-10.48	-0.096***	-7.10
NPV 60d	-0.075***	-17.24	-0.092***	-10.38	-0.092***	-7.34
NPV 90d	-0.065***	-16.06	-0.083***	-9.82	-0.081***	-6.76
NPV 180d	-0.052***	-13.84	-0.064***	-7.76	-0.067***	-5.88
NPV 360d	-0.046***	-12.59	-0.057***	-7.01	-0.062***	-5.68

Table C2. Phone call outcomes of the first call between 9 and 9:30 A.M.

This table compares phone call outcomes of the first calls made between 9 and 9:30 A.M. on day 2 past due between two groups of cases: cases in the treatment group are handled by AI callers on day 2 to day 5 past due before being assigned to human callers on day 6 and cases in the control group are handled by human callers starting on day 2 past due using the 10% completely randomized subsample. Panel A reports results for all first calls. Panel B splits debts by their overdue payment amount at 1000 yuan, which is around the median size. The estimations are based on linear regressions of the variable of interest onto an AI-case indicator with calendar month fixed effects and the hourly time of call fixed effects. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Panel A. Outcomes of the first call.

		(1)	(2)	(3)	(4)
		Mean (AI)	Mean (Human)	Diff.: AI – Human	*t*-stat.
(1)	% Phone calls answered	0.244	0.255	-0.011	-1.53
Conditional on answering the first call:					
(2)	Ringing time to answer (sec)	25.95	25.78	0.17	0.28
(3)	Duration (sec)	31.41	23.48	7.93***	10.10
(4)	% Promise to repay	0.457	0.677	-0.220***	-14.16
(5)	Pay within 15min	0.045	0.032	0.013**	2.24
(6)	Pay within 30min	0.055	0.052	0.003	0.43
(7)	Pay within 1h	0.074	0.092	-0.018*	-1.92
(8)	Pay within 2h	0.110	0.155	-0.045***	-3.80
(9)	Pay within 5h	0.187	0.288	-0.101***	-6.90
(10)	Pay within 1d	0.324	0.489	-0.165***	-10.08

Panel B. Outcomes of the first call, by debt size in terms of overdue amount.

	Overdue amount:	≤ 1000 yuan		> 1000 yuan	
		Diff. AI – Human	*t*-stat.	Diff. AI – Human	*t*-stat.
(1)	% Phone calls answered	-0.008	-0.73	-0.011	-1.13
Conditional on answering the first call:					
(2)	Ringing time to answer (sec)	-1.10	-1.16	1.12	1.40
(3)	Duration (sec)	11.48***	9.58	5.82***	5.78
(4)	% Promise to repay	-0.159***	-6.63	-0.254***	-12.70
(5)	Pay within 15min	0.028***	3.14	0.002	0.22
(6)	Pay within 30min	0.019	1.64	-0.008	-0.88
(7)	Pay within 1h	0.010	0.66	-0.038***	-3.08
(8)	Pay within 2h	-0.013	-0.71	-0.067***	-4.37
(9)	Pay within 5h	-0.048**	-2.10	-0.138***	-7.21
(10)	Pay within 1d	-0.116***	-4.60	-0.198***	-9.36

Table C3. Repayment after the first call between 9 and 9:30 A.M., conditional on answering the calls and making a promise to repay.

This table reports the fraction of borrowers who repay their debts (fully or partially) within various periods after answering the first phone call from AI or human callers between 9 and 9:30 A.M. on day 2 after the due date, conditioning on whether or not the borrowers make a promise to repay their debts during the conversation. The analysis uses the 10% completely randomized subsample. The estimations are based on linear regressions of the variable of interest onto an AI-case indicator with calendar month fixed effects. The time of calling is accounted for by including time-of-day fixed effects for every hour. *t*-statistics are reported in parentheses for the differences in columns e-g. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

	With a Promise		Without a Promise				
	(a) AI	(b) Human	(c) AI	(d) Human	(e) Diff: (a) - (c)	(f) Diff: (b) - (d)	(g) Diff: (e) - (f)
15 minutes	0.069	0.042	0.033	0.016	0.036***	0.026***	0.010
					(3.69)	(7.17)	(1.00)
30 minutes	0.089	0.069	0.037	0.024	0.052***	0.045***	0.006
					(4.28)	(10.24)	(0.50)
1 hour	0.122	0.121	0.041	0.045	0.082***	0.076***	0.006
					(5.30)	(13.49)	(0.34)
2 hours	0.175	0.199	0.063	0.083	0.112***	0.117***	-0.004
					(5.90)	(16.75)	(-0.21)
5 hours	0.271	0.365	0.116	0.159	0.155***	0.206***	-0.051**
					(6.63)	(24.14)	(-2.07)
the same day	0.419	0.593	0.253	0.327	0.166***	0.266***	-0.100***
					(6.48)	(28.47)	(-3.69)

Table C4. Tests of long-run performance gaps, restricted subsample.

This table uses the 10% completely randomized subsample restricted to cases with more than 400 yuan overdue payment amount and internal credit score deciles between 4 and 7 and reproduces the tests of long-run performance gaps between AI and human callers as in **Table 8. Probability of delinquency on subsequent monthly payments.**

This table shows borrowers' probability of missing subsequent monthly payments after they are initially assigned to the treatment group (five days of AI followed by humans) or the control group (always human) in the completely randomized subsample. Subsequent monthly payments due are identified by the number of months between the payment in question and the month borrowers entering the completely randomized subsample. Columns 2 and 3 report the fractions of borrowers who miss the *p*th payment in the treatment and the control groups, respectively. Column 4 reports the difference (AI minus human) and column 5 the corresponding *t*-statistic. The sample for the *p*th payment only includes loans with at least *p* months of payments between the month of entry into the completely randomized subsample and the loan's maturity. The number of such loans are reported in the last column. *** p<0.01, ** p<0.05, * p<0.1.

(1) Subsequent Monthly Payment	(2) AI	(3) Human	(4) Diff.: AI - Human	(5) *t*-stat.	(6) # Loans
1	0.627	0.604	0.023***	8.99	143,482
2	0.601	0.579	0.022***	8.48	139,212
3	0.590	0.570	0.020***	7.23	133,979
4	0.575	0.556	0.019***	7.02	129,021
5	0.563	0.545	0.018***	6.15	122,279
6	0.558	0.539	0.019***	6.42	116,723
7	0.545	0.529	0.016***	5.45	111,998
8	0.535	0.517	0.018***	5.81	105,728
9	0.527	0.511	0.016***	5.11	98,268
10	0.513	0.499	0.014***	4.17	89,934
11	0.505	0.494	0.011***	2.99	80,074
12	0.506	0.492	0.014***	3.64	72,829

Table 9 and Table 10. Panel A reports results corresponding to **Table 8. Probability of delinquency on subsequent monthly payments.**

This table shows borrowers' probability of missing subsequent monthly payments after they are initially assigned to the treatment group (five days of AI followed by humans) or the control group (always human) in the completely randomized subsample. Subsequent monthly payments due are identified by the number of months between the payment in question and the month borrowers entering the completely randomized subsample. Columns 2 and 3 report the fractions of borrowers who miss the *p*th payment in the treatment and the control groups, respectively. Column 4 reports the difference (AI minus human) and column 5 the corresponding *t*-statistic. The sample for the *p*th payment only includes loans with at least *p* months of payments between the month of entry into the completely randomized subsample and the loan's maturity. The number of such loans are reported in the last column. *** p<0.01, ** p<0.05, * p<0.1.

(1) Subsequent Monthly Payment	(2) AI	(3) Human	(4) Diff.: AI - Human	(5) *t*-stat.	(6) # Loans
1	0.627	0.604	0.023***	8.99	143,482
2	0.601	0.579	0.022***	8.48	139,212
3	0.590	0.570	0.020***	7.23	133,979
4	0.575	0.556	0.019***	7.02	129,021
5	0.563	0.545	0.018***	6.15	122,279
6	0.558	0.539	0.019***	6.42	116,723
7	0.545	0.529	0.016***	5.45	111,998
8	0.535	0.517	0.018***	5.81	105,728
9	0.527	0.511	0.016***	5.11	98,268
10	0.513	0.499	0.014***	4.17	89,934
11	0.505	0.494	0.011***	2.99	80,074
12	0.506	0.492	0.014***	3.64	72,829

Table 9 in which cases are split by the day of the first contact and Panel B reports results corresponding to Table 10 in which cases are split by the caller's working experience in the first contact. See the notes in the original tables for more information about the specifications. Column 1 uses the full restricted subsample while columns 3-6 further split the cases by the median internal credit score. Calendar month fixed effects and loan characteristic control variables are included in all specifications. Even columns report the *t*-statistics. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Panel A: First calls on weekends minus first calls on weekdays.

	(1)	(2)	(3)	(4)	(5)	(6)
	All		Internal Credit Score 4-5		Internal Credit Score 6-7	
	Diff.	*t*-stat.	Diff.	*t*-stat.	Diff.	*t*-stat.
NPV2	-0.016***	-3.14	-0.016*	-2.09	-0.014	-1.50
NPV3	-0.019***	-3.93	-0.020**	-2.47	-0.013	-1.09
NPV4	-0.010*	-1.77	-0.014**	-2.15	-0.006	-0.54
NPV5	-0.007	-0.96	-0.013**	-2.24	-0.001	-0.11
NPV6	-0.002	-0.43	-0.013	-1.68	0.004	0.59
NPV10	-0.004	-0.78	0.009	0.90	0.007	1.03
NPV15	-0.001	-0.23	0.018*	1.88	0.009	1.28
NPV30	-0.003	-0.74	0.012	1.43	0.005	0.58
NPV60	-0.001	-0.30	0.011	1.30	0.003	0.43
NPV90	-0.002	-0.38	0.009	1.30	-0.003	-0.53
NPV180	-0.002	-0.42	0.014*	1.92	-0.009	-1.26
NPV360	-0.003	-0.68	0.012	1.64	-0.007	-0.96

Panel B: First calls by junior callers minus first calls by senior workers.

	(1)	(2)	(3)	(4)	(5)	(6)
	All		Internal Credit Score 4-5		Internal Credit Score 6-7	
	Diff.	*t*-stat.	Diff.	*t*-stat.	Diff.	*t*-stat.
NPV2	-0.014**	(-2.19)	-0.016*	(-2.09)	-0.014	(-1.50)
NPV3	-0.015*	(-1.84)	-0.020**	(-2.47)	-0.013	(-1.09)
NPV4	-0.009	(-1.30)	-0.014**	(-2.15)	-0.006	(-0.54)
NPV5	-0.006	(-0.96)	-0.013**	(-2.24)	-0.001	(-0.11)
NPV6	-0.002	(-0.41)	-0.013	(-1.68)	0.004	(0.59)
NPV10	0.007	(1.34)	0.009	(0.90)	0.007	(1.03)
NPV15	0.012**	(2.13)	0.018*	(1.88)	0.009	(1.28)
NPV30	0.007	(1.18)	0.012	(1.43)	0.005	(0.58)
NPV60	0.006	(0.95)	0.011	(1.30)	0.003	(0.43)
NPV90	0.001	(0.20)	0.009	(1.30)	-0.003	(-0.53)
NPV180	-0.000	(-0.02)	0.014*	(1.92)	-0.009	(-1.26)
NPV360	0.000	(0.04)	0.012	(1.64)	-0.007	(-0.96)

Table C5. Phone call outcomes of different versions of AI callers.

This table compares the phone call outcomes between different versions of AI callers. The five phone call-level outcomes are the fraction of calls that are answered by the borrowers, the duration of the calls (in seconds), the fraction of answered calls that end within 10 seconds, the fraction of borrowers who make a promise to repay during the call, and the fraction of calls labeled ineffective by AI. Ineffective calls include calls with only silence after picking up, voicemail or auto-reply voice messages, and conversations that AI cannot understand or classify into pre-specified scenarios. Panel A does not control for the time of calls, while Panel B controls for a set of hourly dummies for the time of calls. The numbers in parentheses are *t*-statistics clustered at the borrower level. *** p<0.01, ** p<0.05, * p<0.1.

Panel A. Not timing-adjusted.

Version Diff.	Phone Call Outcome				
	% Phone calls answered	Duration (sec)	% Calls ≤ 10 sec	% Promise to repay	% Ineffective calls
V2 - V1	0.0160**	0.5230	-0.0249	0.0106	-0.0623***
	(2.50)	(0.730)	(-1.57)	(0.64)	(-4.18)
V3 - V1	0.0029	-0.8290	-0.0544***	0.0874***	-0.1280***
	(0.64)	(-1.29)	(-5.24)	(7.18)	(-16.30)
V4 - V1	-0.0082	-2.0520***	-0.0674***	0.1070***	-0.1290***
	(-1.57)	(-4.02)	(-5.61)	(7.91)	(-12.21)
V5 - V4	-0.0029	2.578***	0.0038	-0.0846***	0.0013
	(-0.45)	(3.54)	(0.26)	(-4.54)	(0.11)
V6 - V5	0.0303***	-0.3130	-0.0053	-0.0009	-0.0106
	(6.18)	(-0.67)	(-0.55)	(-0.07)	(-1.39)

Panel B. Timing-adjusted.

Version Diff.	Phone Call Outcome				
	% Phone calls answered	Duration (sec)	% Calls ≤ 10 sec	% Promise to repay	% Ineffective calls
V2 - V1	0.0124*	0.7050	-0.0225	0.0155	-0.0587***
	(1.93)	(1.00)	(-1.42)	(0.96)	(-3.99)
V3 - V1	0.0038	-0.3960	-0.0530***	0.0974***	-0.1230***
	(0.86)	(-0.62)	(-5.11)	(8.25)	(-15.90)
V4 - V1	0.0044	-1.481***	-0.0694***	0.1240***	-0.1260***
	(0.85)	(-3.00)	(-5.76)	(9.31)	(-12.05)
V5 - V4	-0.0180***	1.455**	0.0016	-0.1090***	-0.0150
	(-2.84)	(1.98)	(0.11)	(-5.82)	(-1.19)
V6 - V5	0.0211***	-0.8160*	-0.0058	-0.0112	-0.0161**
	(4.27)	(-1.73)	(-0.6)	(-0.85)	(-2.10)

Table C6. Relationship between case assignment and performance ranking across human callers.

This table tests the orthogonality of case assignments across callers on day 6 with respect to callers' previous performance ranking terciles. The sample is the same as in Table 13. The tests are implemented with linear regressions where the dependent variables are observable information about the assigned cases, including an indicator for whether the cases were treated by AI V4 in the first five days, NPV collected in the first five days (NPV5), and loan characteristics as in Table 4. In the first two columns, the independent variables are dummies for the assigned caller's previous performance tercile in August 2021 as defined in Section 6.1. An F-test for the joint significance of the tercile dummies is reported. In columns 3 and 4, the independent variables are caller fixed effects, and the joint significance of the fixed effects is tested.

	Joint F-test of previous performance tercile dummies		Joint F-test of caller fixed effects		# Obs.
	F-stat.	p-value	F-stat.	p-value	
AI V4 indicator	1.984	0.138	0.879	0.972	3,328
NPV5	0.183	0.833	1.022	0.367	3,328
Overdue amount	2.001	0.135	0.858	0.988	3,328
Remaining principal	0.232	0.793	0.965	0.699	3,328
Internal credit score	1.021	0.360	0.924	0.879	3,328
Age	0.224	0.799	0.969	0.674	3,328
Male	1.892	0.151	0.995	0.523	3,328
Bachelor's degree or more indicator	0.822	0.440	1.083	0.111	3,328

Online Appendix D. Sample AI phone call transcripts

AI: Hello, may I speak to Mr. Zhang San? **User:** Hi, hello. **AI:** Hello, this is [Lending Company]. You have a bill of 642.01 yuan that is overdue by 3 days. Could you please take care of it within two hours? **User:** Okay, okay, okay, okay. **AI:** All right, please handle it promptly. If the payment is not made, we will contact you again. If you have any questions, you can call [Lending Company] customer service. Sorry to disturb you, goodbye!
AI: Hello, is this Ms. Zhang San? **User:** Yes, that's me. **AI:** This is [Lending Company] calling. Your bill of 1285.77 yuan is overdue by 2 days. Can you handle it within two hours? **User:** Oh, okay. **AI:** Then I will intercept the process for now. Can you confirm you can handle it within two hours? **User:** Yes. **AI:** All right, if the payment is not received within the stipulated time, overdue materials will be submitted on time, and you will bear the consequences of default. Goodbye!
AI: Hello, is this Ms. Zhang San? **User:** Who is this? **AI:** Hello, this is [Lending Company]. You have a bill of 764.79 yuan that has been overdue for 2 days. Could you please handle it within two hours? **User:** I have noted it down, is there anything else? **AI:** Sorry to bother you, goodbye!
AI: Hello, is this Ms. Zhang San? **User:** Ah, yes. **AI:** This is [Lending Company]. You have an overdue bill of 25,758.94 yuan with us for 3 days. Could you please handle it within two hours? **User:** Sure, okay, thank you. **AI:** All right, please handle it as soon as possible. If payment is not made, we will contact you again. For any issues you can call [Lending Company] customer service. Sorry for disturbing you, goodbye!
AI: Hello, is this Mr. Zhang San? **User:**

AI: Hello, is this Ms. Zhang San? **User:** **AI:** Hey, hello, can you hear me? This is [Lending Company]. Your bill with us is overdue. Please handle it immediately, okay? **User:** **AI:** To avoid any negative impact on your personal credit, please make sure to handle the [Lending Company] bill today. If it is not handled, I will contact you again. If you have any issues, you can also call [Lending Company] customer service. Goodbye!
AI: Hello, may I speak to Ms. Zhang San? **User:** Ah, yes. **AI:** This is [Lending Company]. You have a bill of 538.07 yuan with us that has been overdue for 2 days. Please handle it within two hours, okay? **User:** Ah, yes, yes. **AI:** All right, please handle it as soon as possible. If repayment is not made, we will contact you again. If you have any questions, you can call [Lending Company] customer service. Sorry for the disturbance, goodbye!
AI: Hello, is this Mr. Zhang San? **User:**
AI: Hello, may I speak to Mr. Zhang San? **User:** Yes. **AI:** Your [Lending Company] bill of 3062.55 yuan has been overdue for 4 days. Can you handle it within two hours? **User:** Okay, didn't you call yesterday? **AI:** The call is to remind you to repay to avoid negative impacts on your personal credit. Please take care of the debt within two hours, okay? **User:** I am thinking of a way. **AI:** If you don't have money, think of a way. Continuing to delay will result in overdue materials being reported and collection processes initiated. Can it be handled today? **User:** I am thinking of a way. **AI:** To avoid affecting your credit, I suggest you ask someone for a loan and pay it off before five o'clock. Goodbye!

AI: Hello, is this Mr. Zhang San? **User:** Yes. **AI:** This is [Lending Company]. You have a bill of 4762.83 yuan that is 5 days overdue. Could you please handle it within two hours? **User:** No money. **AI:** If you don't have money, you can figure out a way. Continuing to delay will result in your overdue details being reported and a collection process initiated. Can you handle it today? **User:** Unable. **AI:** This call is to remind you to repay the debt and avoid a negative impact on your credit. Please handle the arrears within two hours, okay? **User:** Hmm. **AI:** Okay, if we don't receive the payment within the stipulated time, the overdue materials will be submitted on time, and you will bear the consequences of the breach. Goodbye!
AI: Hello, is this Mr. Zhang San? **User:** Hello. **AI:** Hello, this is [Lending Company]. You have a bill of 2565.14 yuan that has been overdue for 3 days. Could you please handle it within two hours?
AI: Hello, is this Mr. Zhang San? **User:** Hmm. **AI:** Your [Lending Company] bill of 2129.20 yuan is overdue by 4 days. Can you handle it within two hours? **User:** Oh, there's a typhoon today and lots of people in Zhejiang are here. **AI:** The funder has been urging you about this overdue material. Can you confirm it will be handled in two hours? **User:** Hmm. **AI:** Then I will make a note of it for you. If it remains overdue after the specified time, the system will automatically upload your overdue materials to the central bank credit reporting. Goodbye!

AI: Hello, may I speak to Ms. Zhang San? **User:** **AI:** This is [Lending Company]. You have a bill of 604.15 yuan overdue for 3 days. Please handle it within two hours, okay? **User:** Good. **AI:** Continuing to be overdue will result in reporting the default materials to start the collection process. Are you sure you can handle it within two hours? **User:** This. **AI:** Hmm, go ahead. **User:** **AI:** Can you hear me? Continuing to be overdue will negatively impact your credit record. Please handle the [Lending Company] bill within two hours, okay? **User:** Yes. **AI:** This call is to remind you to make the payment, continuing to be overdue will negatively impact your credit record. Please handle the debt within two hours, okay? **User:** **AI:** Can you hear me? Continuing to be overdue will negatively impact your credit record. Please handle the [Lending Company] bill within two hours, okay? **User:** Okay. **AI:** Then hurry up. If the account is still overdue within the specified time, the process of recovering the overdue amount will be directly initiated. Goodbye!
AI: Hello, may I speak to Ms. Zhang San? **User:** Ah, yes. **AI:** This is [Lending Company]. You have a bill of 1261.98 yuan that is 2 days overdue. Please handle it within two hours, okay? **User:** Okay, okay. **AI:** Continuing to delay will result in reporting the overdue materials and initiating the collection process. You're sure you can handle it within two hours, right? **User:**
AI: Hello, may I speak to Mr. Zhang San? **User:** **AI:** This is [Lending Company], you have an overdue bill of 2193.76 yuan with us for 4 days, could you please handle it within two hours? **User:** Okay. **AI:** All right then, please go ahead and handle it. If you do not repay, we will contact you again. If you have any questions, you can call [Lending Company]'s customer service. Sorry to bother you, goodbye!

AI: Hello, may I speak with Mr. Zhang San? **User:** Yes, that's me. **AI:** This is [Lending Company]. You have a bill of 704.44 yuan that is 4 days overdue with us. Please handle it within the next two hours, okay? **User:** I'm busy right now, you'll have to wait a bit. **AI:** Okay, please handle it as soon as possible. If the payment is not made, we will contact you again. If you have any questions, you can call the [Lending Company] customer service. Sorry for the disturbance. Goodbye!
AI: Hello, is this Mr. Zhang San? **User:** **AI:** Hey, can you hear me? This is from [Lending Company], your bill with us is overdue. Please handle it as soon as possible, okay? **User:** **AI:** To avoid adverse effects on your personal credit, please make sure to handle the [Lending Company] bill today. If not handled, I will contact you again. You can also call [Lending Company] customer service if there are any issues. Goodbye!
AI: Hello, is this Ms. Zhang San? **User:** Oh, yes. **AI:** This is [Lending Company], you have a bill of 1294.49 yuan that has been overdue for 2 days, could you please handle it within two hours? **User:** Before the fourth. **AI:** You have taken out a credit loan, continuing to delay will result in reporting overdue materials to collections. Can you handle it today? **User:** Tomorrow. **AI:** Then I will note it down, not paying today will incur new overdue charges, and personal credit will also be affected, goodbye!
AI: Hello, is this Mr. Zhang San? **User:** Yes **AI:** This is [Lending Company]. You have a bill of 1115.82 yuan that has been overdue for 2 days. Please handle it within 2 hours, okay? **User:** I know. **AI:** I'll intercept the process for you. Can you confirm it can be handled within 2 hours? **User:** What process? **AI:** The call is to remind you for repayment to avoid negative impact on your personal credit. Please handle the debt within 2 hours, okay? **User:** Hmm, if something comes up, maybe tonight or tomorrow morning. **AI:** You have a credit loan. Continued overdue will report overdue materials and initiate collection process. Can it be handled today? **User:** I'll see, you just go ahead. **AI:** Okay, if it doesn't arrive within the stipulated time, the overdue materials will be submitted on time, and you will bear the consequences, goodbye!

AI: Hello, may I speak to Ms. Zhang San? **User:** Assistant, may I ask what you want to discuss with him? **AI:** Sorry to bother you, goodbye!
AI: Hello, may I speak to Mr. Zhang San? **User:** Hi, hello. **AI:** Hello, this is [Lending Company]. You have a bill of 642.01 yuan that is overdue by 3 days. Could you please take care of it within two hours? **User:** Okay, okay, okay, okay. **AI:** All right, please handle it promptly. If the payment is not made, we will contact you again. If you have any questions, you can call [Lending Company] customer service. Sorry to disturb you. Goodbye!
AI: Hello, is this Ms. Zhang San? **User:** Yes, that's me. **AI:** This is [Lending Company] calling. Your bill of 1285.77 yuan is overdue by 2 days. Can you handle it within two hours? **User:** Oh, okay. **AI:** Then I will intercept the process for now. Can you confirm you can handle it within two hours? **User:** Yes. **AI:** All right, if the payment is not received within the stipulated time, overdue materials will be submitted on time, and you will bear the consequences of default. Goodbye!
AI: Hello, is this Ms. Zhang San? **User:** Who is this? **AI:** Hello, this is [Lending Company]. You have a bill of 764.79 yuan that has been overdue for 2 days. Could you please handle it within two hours? **User:** I have noted it down, is there anything else? **AI:** Sorry to bother you. Goodbye!
AI: Hello, is this Ms. Zhang San? **User:** Ah, yes. **AI:** This is [Lending Company]. You have an overdue bill of 25,758.94 yuan with us for 3 days. Could you please handle it within two hours? **User:** Sure, okay, thank you. **AI:** All right, please handle it as soon as possible. If payment is not made, we will contact you again. For any issues you can call [Lending Company] customer service. Sorry for disturbing you. Goodbye!