Writing Quality and Soft Information in the GenAI Age: Evidence from Online Credit Markets*

Lin William Cong Yanhong Guo Xin Zhao Wenjun Zhou

Current Version: March 26, 2026
(First Version: September 1, 2024)

Abstract

We document systematically how various dimensions of writing quality matter for different borrower groups on a dominant online credit platform, and would be reshaped by the use of large language models (LLMs). Using human assessments, we find that LLMs (e.g., ChatGPT) significantly enhance the writing and perceived quality of loan applications. We further build proprietary BERT-based multimodal models of lenders' decision-making that accommodate lenders' constraints, and introduce "deep" Heckman correction for sample selection bias. We demonstrate that ChatGPT adoption decreases soft information conveyed and increases credit misallocation, likely due to convergence in writing; when lenders respond to borrowers' LLM adoption, they rely more on hard information, mitigating misallocation. We leverage generative modeling to characterize the counterfactual equilibrium with endogenous borrowers' LLM adoption and lenders' responses. Our findings provide insights into the evolving role of soft information and potential impacts of GenAI on lending.

Keywords: Generative AI, Heckman Correction, Lending, LLMs, Misallocation, Multimodal.

*Cong is at Nanyang Technological University, ABFER, CEPR, and NBER. Guo and Zhao are at Dalian University of Technology School of Economics and Management. Zhou is at Haslam College of Business, The University of Tennessee, Knoxville. We thank Benedict Guttman-Kenney, Simon Xu, Loïc Maréchal, Allen Hu, and conference and seminar participants at Boston Fed, Mitsui Life Symposium, Chinese University of Hong Kong, City University of Hong Kong, CUHK Shenzhen, Fudan University, 2024 Generative AI for Finance Conference (Montréal), 2025 MRS International Risk Conference (Boston), Lingnan University, 1st Annual MARC Conference at the University of Toronto, and National University of Singapore for constructive comments. Send correspondence to Cong at will.cong@ntu.edu.sg.

1 Introduction

It pays to write well in corporate disclosure, loan applications, and other financial settings (Iyer et al., 2016; Hwang and Kim, 2017; Gao et al., 2023). However, a systematic understanding of the various dimensions of "writing well" and how they link to the soft information conveyed is missing in the literature. Recent breakthroughs in generative AI (GenAI), especially large language models (LLMs), not only illuminate these issues but are also believed to dramatically change how people write in practice.[1] We investigate how LLMs affect writing quality in financial settings, the impact of borrower adoption and lenders response, and the implications for the transmission of soft information and allocative efficiency in financial intermediation.[2]

To this end, we use (i) historical data from Prosper, the first marketplace lending platform in the United States, and the most prominent one;[3] (ii) loan applications aided by ChatGPT, a representative LLM; and (iii) proprietary multimodal lender decision AI models. In this process, we introduce, for the first time in finance, a Heckman-like correction for deep-learning models and an empirical implementation of a new procedure for analyzing counterfactual equilibrium (Cong, 2025), both of which constitute methodological contributions.

Based on our data, we demonstrate that LLM adoption significantly alters credit allocation dynamics: while ChatGPT could have historically increased approval rates by up to 6.52 percentage points (pp) for some borrower groups, its actual adoption also raises the false positive rates (FPRs) by 8.31 pp (i.e., more than doubled), assuming lenders continue lending the way they did historically. When lenders endogenously react to widespread LLM adoption, such misallocation is mitigated, and whether allocative efficiency remains significantly below pre-GenAI levels depends on lender constraints (e.g., lack of access to powerful models or behavioral biases). As borrowers' adoption of LLMs increases and lenders respond to the reduction in soft information by focusing more on hard information, lenders' return on investment (ROI) drops from 23.9% historically

[1] While technological breakthroughs are often associated with improved tools for collecting higher-quality information, they can also diminish the availability or usefulness of soft information.

[2] On the one hand, better writing can more effectively convey soft information, which is often communicated through unstructured channels (texts, conversations, etc.). On the other hand, better writing could also be more persuasive for getting the recipient act in favor of the sender even with the same informational content.

[3] Although Prosper no longer uses borrower narratives today, since their dataset is the most complete with rich borrower and lender dynamics information, we continue to use its data to conduct our counterfactual investigation. Our goal is to demonstrate the impact of GenAI on financial information communication; thus, the findings are still informative for a wide range of other contexts, such as corporate disclosure versus investor reactions.

to 14.5% without lending adjustments, recovering to 20.6% with lenders' equilibrium responses. These findings underscore the likely nuanced impact of LLMs on online lending markets.

Specifically, following prior studies on online credit markets (e.g., Netzer et al., 2019; Gao et al., 2023), we use rich loan-level lending and performance information from the Prosper platform, which predates the introduction of any LLMs. To generate counterfactual loan applications when borrowers adopt LLMs, we rewrite historical loan descriptions using ChatGPT in a way similar to how borrowers would have done. Following OpenAI's tutorial, we design a prompt consisting of the role, task, and rule settings. Rule setting encompasses specific instructions on how the borrower wishes to change the original loan description, including tone, style, and formatting. Our main results are based on a prompt that aims to improve writing quality and loan fundability without advanced knowledge from research about what specifically would improve loan fundability—what borrowers would naturally ask—and are robust to rerunning randomized generations of each prompt version multiple times and employing alternative prompt versions.

We first verify how writing quality is historically associated with lending outcomes. We focus on the major categories of textual features, such as spelling errors and tone, as is standard in linguistic and financial studies. In general, text length, formality, and tone are positively associated with loan funding, whereas spelling and grammatical errors are significantly associated with rejection. However, the associations can be heterogeneous across individuals with different credit scores: more spelling errors by medium-credit borrowers and less formal writing by low-credit borrowers are associated with better funding outcomes.

To analyze the impacts of LLMs on writing quality, we first employ human-based evaluations to confirm that ChatGPT indeed improves the perceived writing quality and subjects' willingness to lend significantly. Concerning textual features, rewrites are more detailed, formal, sound, and contain fewer grammatical and spelling errors. By improving writing quality, ChatGPT seems to improve borrowers' chances of securing funding at all credit levels.

That said, LLMs may alter the soft information that lenders rely on to make lending decisions, especially for the online credit market, where loan descriptions constitute the primary channel soft information is conveyed to the lenders.[4] To quantify soft information in text, we assess lenders'

[4] Iyer et al. (2016) find that lenders can effectively gauge borrower quality using soft information in loan descriptions, such as word choice, writing clarity, and tone. This allows lenders to outperform traditional credit scores in predicting default. Similarly, Netzer et al. (2019) report that peer-to-peer lending borrowers' text descriptions offer insights into

decision quality when using both structured and textual features versus structured features alone. Decision quality is measured by classification accuracy[5] or return on investment (ROI). We employ a multimodal transformer model (Gu and Budhkar, 2021) that incorporates text embeddings processed by Bidirectional Encoder Representations from Transformers (BERT) and train a unified lender model that accommodates lender constraints in a reduced form.

The results reveal a significant decline in the predictive value of texts. Using cosine similarity to measure within-group homogeneity, we find that the loan descriptions of low-credit borrowers are more similar to each other, and those of high-credit borrowers are less so, despite the fact that homogeneity increases in both groups after ChatGPT rewriting. In other words, ChatGPT reduces the variability in writing, which may explain the diminished soft information. This convergence makes it harder to discern true loan quality. Moreover, this reduction in the effective use of soft information results in greater credit misallocation.

To mitigate the potential sample selection bias arising from solely relying on funded loans (for which loan quality and repayment outcomes are observed), we devise a "deep Heckman correction" for our multimodal transformer modal in the spirit of Heckman (1979) and building from the specification of Kahng et al. (2023). Compared with the pre-GenAI scenario, if borrowers have adopted LLMs while lenders lend as before, there would be a significant increase in credit misallocation, as measured by the false positive rate (FPR) and false negative rate (FNR) when juxtaposing the funding decision and the eventual loan outcome (repayment or default).[6]

The above analyses are at best under a "partial equilibrium" view, because we have not accounted for the fact that lenders can adapt their decisions in a marketplace anticipating LLM usage. Thus, we are interested in expanding the scope of our research by considering the lender response. Specifically, with large-scale LLM adoption by borrowers, lenders would update their decision model by training it with newer data, which includes descriptions produced by ChatGPT for some loans. In this counterfactual scenario, we retrain the multimodal lender model using generated loan descriptions to maximize some combination of conformity to historical lending practices and

past behavior and future intentions, enhancing loan outcome predictions.

[5] Following Netzer et al. (2019), we use the area under the curve (AUC) as a measure of the decision model's predictive or discriminative power. In particular, when a model is trained to predict whether a loan will default, we create a receiver-operating characteristic (ROC) curve based on the true positive rates and false positive rates at various cutoff values of the model's prediction. The AUC values range between 0 and 1, with 1 indicating perfect prediction.

[6] Precisely because of possible manipulation (Berg et al., 2020), we need the counterfactual response from the lenders. Using historical lender response would not be sufficient.

lenders' expected payoff out-of-sample. Misallocation is somewhat mitigated because soft information is partially recovered, and lenders also rely more on hard information.

In a counterfactual equilibrium (as defined in Cong, 2025), borrowers' LLM adoption level solves a fixed-point problem: lenders take LLM adoption and usage by borrowers as given, and use both historical data and applications generated under LLM adoption to train their model. Borrowers take the lenders equilibrium decision as given, and make optimal decisions on LLM adoption and usage. The aggregation of their adoption decisions, in turn, pins down the equilibrium adoption level. Guided by historical patterns, we conjecture that borrower adoption is correlated with credit rating, search for possible adoption levels, and verify whether each level is an equilibrium. With the small adoption costs of LLM, we find that full adoption of LLM is a unique equilibrium. We discuss soft information and misallocation in counterfactual equilibrium.

This study is the first to systematically verify the effectiveness of LLMs in improving the writing quality and fundability of loan applications. We show that although LLMs can help enhance fundability, they lead to convergence in writing, reduce soft information available to help lender decisions, and lead to greater credit misallocation. We are also the first empirical study to analyze long-run counterfactual equilibria under endogenous LLM adoption, while existing works focus on the impact of LLMs on decision makers (Bowen III et al., 2024) without considering responses by counterparties and equilibrium implications. While low-credit borrowers lose their historical advantage in conveying soft information, they still benefit from the ChatGPT's improvement in writing quality. In such a unique equilibrium, borrower loan descriptions converge in writing and lenders rely more on hard information.

We also make several methodological contributions to the literature. First, this is one of the few studies that build multimodal Transformer models in economics, and is the first one for the online credit market.[7] The multimodal modeling of lending decisions allows modality interactions and captures richer nonlinearity and information content from various types of data, achieving greater out-of-sample performance. Second, we extend the Heckman correction to account for selection biases in deep learning models. Third, our use of generated/synthetic data to study the counterfactual equilibrium also illustrates empirically how, in settings where historical data are

[7] Cong et al. (2019) is the first application of Transformer in Finance; Chen et al. (2022) also recently apply transformer-based language models to asset pricing.

insufficient because the data-generating process could have changed, generative modeling comes to rescue to enable counterfactual equilibrium analysis in a data-driven manner, complementing reduced-form theory and structural models.

2 Institutional Background and Data Description

We exploit historical data from a prominent online credit market to (i) understand how writing quality matters (before the advent of LLMs) and how LLMs improve writing quality, (ii) extract the underlying distribution of borrower types which is useful when considering counterfactual credit allocation, and (iii) generate loan applications in order to characterize equilibrium outcomes when adoption and usage of LLMs in the online credit market are endogenous.[8]

2.1 Online Credit Data from Prosper Marketplace

Prosper Marketplace, established in 2005 and began operations in 2006, is recognized as the first marketplace lending platform in the U.S.[9] On the platform, despite different loan origination dates, the lending process is consistent across all loans and follows several stages, as illustrated in Figure 1. In the pre-funding stage, borrowers submit loan requests and open listings for lenders to bid for. They specify the amount requested, maximum interest rate they are willing to accept, length of the bidding period (up to two weeks), minimum fulfillment rate (e.g., 80%) for the listing to become a loan, and structured and unstructured information about their credit profiles. Then in the bidding stage (Stage 1), lenders specify the bidding amount and the minimum interest rate they are willing to lend for this amount, all within the bidding period as specified in the listing.[10] Once the total requested amount has been fulfilled within the bidding period, the listing becomes a loan. Otherwise, the loan application is rejected. Stage 2 begins when a loan is funded and lasts until maturity. We consider a loan as having "defaulted" if its status in Prosper is "Charge-off," "Defaulted (Bankruptcy)," or "Defaulted (Delinquency)." Conversely, we categorize a loan as "paid"

[8] The implicit assumption is that the distribution of borrower quality is stationary; should this fail to hold, resampling from the historical distribution guided by stipulated evolution of borrower quality can be used.

[9] See https://www.prosper.com/about. An example listing can be found in Appendix D of Iyer et al. (2016).

[10] To effectively spread limited funds of individual lenders across different loans for diversification, the lending amount is usually much smaller than the requested amount, with a minimum of $25.

when it is marked as "Paid in full," "Settled in full," or simply "Paid." Since the text component of their loan requests is written in Stage 1, it contains information potentially useful in predicting both the likelihood of funding and the likelihood of default.

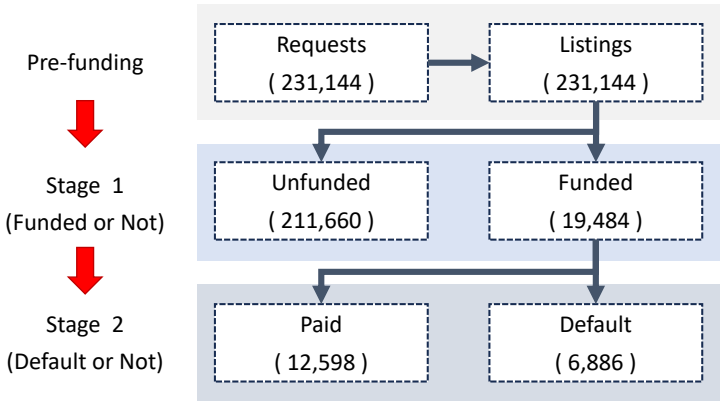


Figure 1: Life Cycle of Online Lending

We downloaded all loan requests posted between April 2007 and October 2008 with a total of 231,144 listings. We focus on this early episode of Prosper for multiple reasons: (i) data during this sample period were once publicly available and are richer and more diverse than data from later episodes after regulations tightened;[11] (ii) this period predates the widespread use of language models so that we can isolate the effect of how LLMs improve writing. Of these listings, 19,484 (8.4%) were successfully funded. The longest loan term was 36 months, and all loans were mature at the time of data collection with observable repayment outcomes. In summary, 6,886 (35.3%) of the funded loans defaulted and the remaining 12,598 (64.7%) were repaid.

2.2 Traditional Credit Features

Borrowers' financial and demographic information is available to lenders, mostly as structured data. These variables include loan request information, the borrower's credit profile, demographic information, and the Prosper's overall credit rating.

The borrowers' credit profile includes their debt-to-income ratios (*DebtToIncomeRatio*), number of delinquencies (*DelinquenciesLast7Years*), and number of inquiries in the past six months (*InquiresLast6Months*). These are commonly used in traditional risk models (Guo et al., 2016;

[11] This is consistent with other studies that use the Prosper dataset. For example, see Netzer et al. (2019).

Fitzpatrick and Mues, 2021). Their demographic information includes a categorical variable indicating current employment status (*EmploymentStatus*), tenure (in years) in current employment status (*LengthStatusMonths*), and a dummy variable indicating whether the borrower is a homeowner (*IsBorrowerHomeowner*).

Loan request information include the amount requested in U.S. dollars (*AmountRequested*) and the lenders' interest rate (*LenderRate*).[12] Similar to existing studies, we include a dummy variable for whether the user has a headshot photo in their user profile (Li and Xie, 2020; Ho et al., 2022). In addition, the credit grade of the borrower (*CreditGrade*) is an ordinal variable with seven possible levels ranging from AA (best credit) to HR (high risk). Appendix A provides more details of these features, including their definitions and descriptive statistics.

2.3 Loan Descriptions and Text Feature Extraction

Writing quality is a well-studied topic in linguistics, education, and cognitive psychology. Building on the existing literature, we consider the following key dimensions of textual features.

Text length. Texts should convey the right amount of information using concise and coherent language. Longer text may convey more detailed information, but it may reduce readability or even suggest deception (for example, Ott et al., 2012; Hancock et al., 2007). Meanwhile, shorter profiles often signal dishonesty on online dating sites because liars may avoid elaborating on certain subjects (Toma and Hancock, 2012). The relationship between loan length and quality in online lending remains unclear (Toma and Hancock, 2012; Hancock et al., 2007). That said, the free-text box for borrowers to enter a description at Prosper.com has a limit of up to 4,000 characters (Ryan et al., 2007), borrowers are prevented from writing longer texts and are often elected to write much shorter than this limit. To measure text length in loan applications, existing studies use the word count (Netzer et al., 2019; Siering, 2023) or character count (Dorfleitner et al., 2016; Larrimore et al., 2011) of the loan descriptions. We measure text length as the number of words in the loan description.

[12] Since the data was collected in a period when interest rates were determined by lenders bidding on the loan requests, lenders' interest rate reflects the extent to which the request is appealing from the perspective of investors.

Spelling and grammatical errors. Linguistic conformity reduces spelling and grammar errors. Few spelling errors directly enhance text readability by contributing to clarity and coherence (Loughran and McDonald, 2014). Dorfleitner et al. (2016) reports that spelling mistakes often act as indicators of traits associated with lower income. Consequently, Harkness (2016) found that spelling errors are associated with a lower probability of loan approval in traditional lending channels. In peer-to-peer lending, Netzer et al. (2019) accounts for the number of spelling errors in loan application texts. Gao et al. (2023) considers not only spelling errors but also grammatical errors as components of readability, with the count of each divided by the number of words in the loan request and then standardized to zero mean, unit variance. We measure (the lack of) conformity as the number of spelling and grammatical errors[13] divided by the number of words in the loan description.

Formality. According to Tausczik and Pennebaker (2010), this metric is typically employed to assess complex languages, educational levels, and social statuses. Nyhus and Webley (2001) found that individuals with higher education often have higher income and financial literacy, reducing their likelihood of loan default compared to those with less education. However, the use of sophisticated language can also be risky if readers perceive it as artificially or frivolously complex, suggesting that languages with higher complexity are likely to be used deceptively (Oppenheimer, 2006). Iyer et al. (2016) considers the extent of advanced vocabulary use by measuring the average number of characters in each word. Similar to Netzer et al. (2019), we measure sophistication using the percentage of words with six or more letters.

Readability. Although general writing quality can be assessed in various dimensions, recent research in finance has primarily focused on readability (Hwang and Kim, 2017; Cao et al., 2023; Gao et al., 2023), the ease with which readers can understand the text. Lower readability of firms' public disclosures has been linked to a significant discount in firm value, as it can reduce investor trust and quality perception (Hwang and Kim, 2017). Similarly, in the context of peer-to-peer lending, higher readability is linked to greater investor enthusiasm and lower interest rates (Gao et al., 2023). Several readability metrics have been used in finance literature to capture the years

[13] We used the "spellchecker" Python package to detect spelling errors, and the "language-tool-python" Python package to detect grammatical errors. See https://pypi.org/project/language-tool-python/.

of education required for a reader to understand the text at first glance (for example, Kim et al., 2019; Cao et al., 2024; Gao et al., 2023). We follow Netzer et al. (2019) and use a measure typically referred to as the "simple measure of gobbledygook," or the SMOG index (McLaughlin, 1969). It is interpreted as the number of years of formal education required to understand the text. Therefore, larger values indicate that the reader requires more education to understand it. To extract the SMOG readability score, we apply the "textstat" package.

Tone. Writing quality can also be measured through engagement and style. One of the engagement elements is related to emotional beat, often measured as tone or sentiment (Loughran and McDonald, 2011). Tone is an important aspect in communicating financial information, given that investors often respond favorably to a positive tone or sentiment (Tetlock, 2007; Tetlock et al., 2008), and entrepreneurs who demonstrate higher positivity have a greater likelihood of success (Zott and Huy, 2007). Similarly, in online credit markets, Gao et al. (2023) find that loan descriptions with positive sentiments are more likely to secure funding. We measure the tone of loan descriptions using sentiment scores provided by a pre-trained sentiment-BERT model,[14] with higher values indicating a more positive tone.

From Table A2, we can see that compared with unfunded loans, funded ones tend to be longer, with less spelling and grammatical errors, more formal, and has a slightly more positive tone.

3 Generating Loan Descriptions Using ChatGPT

3.1 GPT Prompt Design

To use LLMs to help with loan applications, borrowers would input prompts. As econometricians, we can mimic such behavior to generate loan application data for characterizing counterfactual equilibria that borrow endogenously adopt and use LLMs. OpenAI guidelines prescribe that an effective prompt should be specific and include role setting, task setting, and rule setting. Role settings enable ChatGPT to comprehend the context, task setting defines the task for ChatGPT,

[14] https://huggingface.co/michellejieli/emotion_text_classifier.

and rule setting clarifies the rules that it should follow. In the baseline prompt design, we asked ChatGPT to improve the writing and general fundability of the loan description.

A borrower using LLMs to write loan applications would want help from a professional. We thus describe the writer role as: "*You are a practitioner who has been in the lending business for many years and are well versed in matters of applying for loans*." The task component communicates the required actions. Since we want the ChatGPT to conduct a rewrite, we describe the task of ChatGPT as: "*Please help me rewrite the following narrative in a loan application*." Our template also includes a section of quoted text to capture the original text for the ChatGPT to rewrite. Figure 2 shows an example, in which the rule setting is: *Please improve the clarity and readability of the text, and emphasize anything that helpful for lenders' decision-making. Please make the reason compelling, and make myself sound trustworthy, so lenders willing to fund it.*

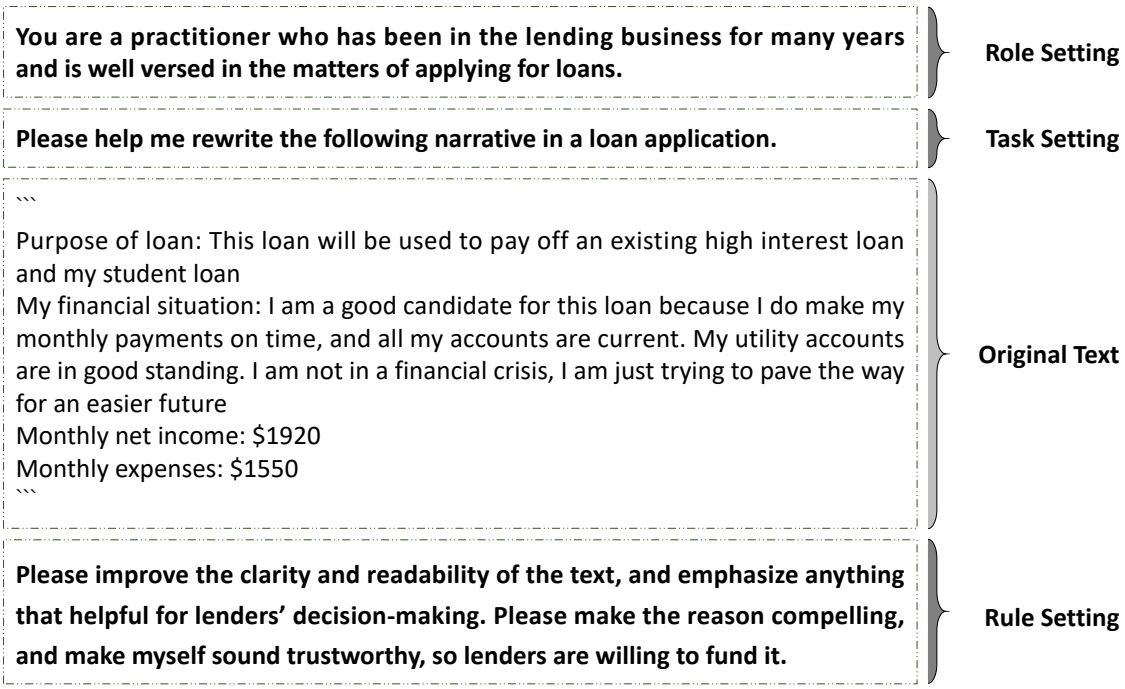


Figure 2: An Example of ChatGPT Prompt

Guided by this prompt template, loan descriptions can be generated based on input. For example, for the original text in Figure 2, a rewritten version is as follows: "*I am seeking this loan to consolidate my existing high interest loan and my student loan. To illustrate my financial situation, I am consistently making all of my monthly payments on time, and all of my accounts are up to*

date. My utility accounts are also in a good standing. I would like to clarify that I am not experiencing any financial difficulties; rather, I am considering this loan as a means to create a smoother future. As for my income, I earn a monthly net income of $1920. With regards to my expenses, they total $1550 per month." It is evident that there are significant differences in the description of financial status (monthly net income) before and after rewriting, with the rewritten ones having better writing quality and credibility. Appendix B offers more examples for illustration.

We used OpenAI's API in the GPT4o-mini model for text generation and leveraged its default parameters. In particular, we set the temperature parameter to one, which strikes a balance between randomness and consistency.[15] In addition, to mitigate possible bias in framing the rule settings in the prompt, before generating rewrites for the loan descriptions, we first generated variations in the sentences in the rule setting. We generated ten versions for each sentence. Then, when putting together the prompts to generate rewrites for the loan descriptions, we randomly select one of the versions of each sentence. In Appendix C, we document the variations in rule-setting sentences and report the results under alternative prompts, including those where borrowers possess advanced knowledge based on research on what improves fundability.

We first conduct a human perception test to determine whether the writing quality and fundability of loan descriptions are indeed improved by ChatGPT before we examine the specific textual features of the original loan descriptions and how they alter each of them.

3.2 Human-Rated Improvements

Recent literature has shown that LLMs such as ChatGPT surpass humans in producing high-quality writings (Herbold et al., 2023; Deveci et al., 2023). We first validate this claim in the context of lending using an online experiment involving human raters.

To conduct experiments in which humans assess the writing quality and fundability of ChatGPT rewrites relative to the original ones, we drew a stratified sample of our dataset based on funding status (unfunded, funded), credit score level (high, medium, low), and semantic similarity before and after rewriting (measured by the cosine coefficient between the embeddings of the original

[15] The temperature parameter ranges from 0 to 2. Larger values correspond to more random answers, whereas smaller values correspond to more consistent answers. See https://platform.openai.com/docs/api-reference/chat/create for a list of all parameters and their default values.

and rewritten text, divided into four quantiles). From the 24 possible combinations of these three factors, we randomly selected five loan applications per combination and obtained their rewritten versions. In other words, we have 120 loan application samples with their text descriptions before and after rewriting.

Our experiments were conducted on the Prolific platform with participants restricted to adults in the United States. To capture a broad perspective on perceived writing quality and willingness to lend, we did not impose specific requirements, such as investment experience or education level. Each participant was randomly assigned to one of two groups: one group assessed the perceived writing quality of the texts, whereas the other evaluated their willingness to fund based on the text.

Each participant evaluated 48 texts, including 24 original descriptions and their corresponding rewrites, with the sentence order randomly shuffled. Participants were not informed whether the text was original or ChatGPT-rewritten, and the correspondence between the original and rewritten text was not disclosed, as all 48 pieces of text were randomly ordered.

Does the ChatGPT-Rewritten Version Exhibit a Higher Writing Quality? One group of human raters was given the following instructions. "*In online peer-to-peer lending platform, individuals request loans from potential lenders. The borrower's request includes a loan description detailing why they need loans and their financial situation. As a lender, these are risky investments - historically, more than one-third of borrowers defaulted. After reading the loan request descriptions below, please rate their* ***their writing quality*** *on a scale of 1 to 10, where 1 represents a low quality and 10 represents a high quality.*"

Does the ChatGPT-Rewritten Version Enhance Fundability? The other group of human raters was given the following instructions. "*In online peer-to-peer lending platform, individuals request loans from potential lenders. The borrowers' request includes a loan description detailing why they need the loan and their financial situation. As a lender, these are risky investments - historically more than one-third of borrowers default. After reading the loan request descriptions, please respond to* ***your willingness to lend****. Rate your willingness on a scale of 1 to 10, with 1 indicating the lowest willingness and 10 indicating the highest willingness.*"

We asked human raters recruited on the Prolific platform to rate either the perceived writing

Table 1: Human Rater Results

	DV: Writing Quality	DV: Willingness to Lend
Rewrite	**2.342***	**1.450****
	(0.544)	(0.203)
Constant	5.283***	4.500***
	(0.272)	(0.102)
Listing Fixed Effect	Yes	Yes
Rater Fixed Effect	Yes	Yes
Observations	240	240
R-squared	0.730	0.717

Sig. levels: * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors (clustered at rater and listing levels) are shown in parentheses.

quality or their willingness to lend based on the text, all on a scale of 1-10. Results are summarized in Table 1. We can see that the human-perceived writing quality of loan applications improves when rewritten using ChatGPT. Similarly, the loan descriptions generated by ChatGPT by rewriting the original loans received higher funding interests. These results are consistent with our expectations.

3.3 LLM-Generated Changes in Textual Features

After verifying that the overall impression of writing quality is improved by LLM, the next question is in what ways the LLM has changed the writing. Therefore, we compare the textual features before and after the ChatGPT rewrite and show the results in Table 2. We note that ChatGPT significantly modifies all textual features, making the rewritten texts longer, with fewer errors and a slightly more positive tone. On average, following a ChatGPT rewriting, the number of words in loan descriptions increases by 17.7% on average, the percentage of spelling and grammatical errors decreases by 77.8% and 58.1%, respectively, and the tone increased by 1.8%. In addition, formality increases by 43.1% and readability increases by 48.5% on average. In summary, by using simple prompting, ChatGPT has moved the text features in directions linked to higher fundability.

Appendix D explores the relationship between text features and fundability. We find that historical funding decisions are associated with writing quality features, and the improved fundability from ChatGPT rewriting is explained by both baseline and changes in text features, especially text length. We also verified that the pattern of text feature change is robust with different versions of

Table 2: Textual Features Before and After Rewrite

	Original		Rewrite		Change	
	Mean	(SD)	Mean	(SD)	DIF[a]	%
Text Length	196.4	(127.8)	231.3	(81.20)	34.7***	+17.7%
Linguistic Conformity	0.110	(0.052)	0.039	(0.025)	-0.071***	-64.5%
Formality	0.239	(0.058)	0.342	(0.037)	0.103***	+43.1%
Readability	10.55	(4.043)	15.67	(2.092)	5.120***	+48.5%
Tone	0.731	(0.158)	0.744	(0.145)	0.013***	+1.8%

[a] Sig. levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

the prompt.

Results so far indicate that ChatGPT rewriting helps improve text features, which in turn could improves fundability. However, an important assumption here is that lenders make decisions as before and do not adjust their models. Our next research question is then, how will the online lending market look like when lenders adjust their decision models anticipating borrowers' use of ChatGPT.

4 Fundability and Soft Information in the GenAI Age

Lenders are decision-makers when funding loan applications. Consequently, they determine the allocation of credit and how soft information is received and utilized. Next, we describe our construction of the lender model using a multimodal transformer adapted with the Heckman correction. This model helps us examine the increase in fundability and loss of soft information when loan descriptions are generated by LLMs. Both the adoption of a multimodal setup and the Heckman correction for AI models in finance constitute novel technical contributions.

4.1 A "Lender Model" with Heckman Correction

To analyze counterfactuals in the GenAI age, we need a model that mimics lenders' decisions. It is possible to reproduce lender decision processes using a well-trained machine model. This strategy was adopted in previous studies (Bali et al., 2023; Cao et al., 2024). First, we recognize that lenders encounter both structured and textual data from loan applications. To jointly model text and structured features as predictors, we adopted a multimodal transformer (Gu and Budhkar,

2021) as the base architecture. Further technical details can be found in Appendix E.

Heckman correction in traditional empirical models. It is also important to note that our data are truncated; if a listing was not funded historically, we would not know whether the loan would have been repaid if funded. This poses a new challenge when using models trained on a selected sample. Traditionally, Heckman selection has been used to adjust for sample selection bias (Heckman, 1979). Specifically, suppose that at the population level, there is a prediction function f and a selection model g, specified as

$$y_i = f(\mathbf{x}_i) + U_{1i}, \tag{1}$$

$$z_i = g(\mathbf{x}_i) + U_{2i}, \tag{2}$$

respectively. Here, $\mathbf{x}_i$ represents sample i's feature vector, y_i represents its observed outcome (e.g., repaid or not), and z_i represents whether the sample is selected (e.g., being funded). U_{1i} and U_{2i} are the error terms, which are correlated at the sample level (and are independent across samples) with a covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

In the online lending context, prediction model f takes loan features $\mathbf{x}_i$ as input and predicts its quality y_i, represented as the propensity for repayment. The selection model g would also take loan features $\mathbf{x}_i$ as input and predict z_i, the selection status indicating whether the loan would be selected for funding. Only those selected for funding would allow us to observe the repayment status. In this case, if we attempt to use all listing samples to fit Eqn. (1), there will be missing values for y_i for all unfunded samples. Alternatively, if we use only funded loans to fit Eqn. (1), its model estimates may be subject to selection bias.

In simple regressions, Heckman (1979) showed that by assuming that U_{1i} and U_{2i} follow a joint bivariate normal distribution, it is possible to derive a sample-level Inverse Mills Ratio (IMR) from selection model g, which is then included as an extra term in prediction model f to account for sample selection bias. In particular, this IMR is computed as $\lambda_i = \frac{\phi(Z_i)}{\Phi(-Z_i)}$, where ϕ and Φ are the density and distribution functions of a standard normal variable, and $Z_i = -\frac{g(\mathbf{x}_i)}{\sqrt{\sigma_2^2}}$ represents sample i's standardized propensity of selection. In this case, Eqn. (1) would become $y_i = f(\mathbf{x}_i) + \lambda_i + U_{1i}$, whose estimates obtained for model f using the funded loans only would be unbiased due to the

IMR adjustments.

Heckman correction for deep learning models. We extend this classic Heckman correction procedure beyond regression models to the deep learning context, making the same assumption for the joint distribution of U_{1i} and U_{2i}. In particular, instead of a linear specification, functions f and g are nonlinear (e.g., multimodal transformers). Using a deep learning model for selection, we obtain the samples' selection propensity from its output and then follow the same logic as the classic Heckman model. To streamline this process, we employ a custom integrated learning algorithm inspired by a recent publication in the machine learning field (Kahng et al., 2023). More specifically, we assume that both the selection and prediction models need to be estimated. We propose a negative log-likelihood loss function that simultaneously fits models f and g, and provides estimates of the covariance matrix following a maximized likelihood estimation framework. The learning objective can be specified as

$$\arg\min_{f,g,\Sigma} \sum_i \left[s_i \Lambda(f(x_i), g(x_i); y_i, s_i; \Sigma) - \{ s_i \log \Phi(g(x_i)) + (1 - s_i) \log \Phi(-g(x_i)) \} \right], \qquad (3)$$

where f and g are two multimodal Transformer models (for prediction and selection, respectively), $\Lambda(f(x_i), g(x_i); y_i, s_i; \Sigma) = -\log p(y_i | s_i = 1, x_i)$ is the conditional negative log probability of y_i given $s_i = 1$. Note that in this formulation, the covariance structure Σ is also a learnable parameter and thus does not require users to set it.

We randomly partition the loans into training and test sets at the 4:1 ratio, developed a lender model based on a training set and assessed its performance on a test set. Using cross-validation to optimize hyperparameters, we arrive at a multimodal Transformer with "deep Heckman correction" as our lender model.[16] The lender model can be trained on any data the lender faces to maximize a given objective of the lenders (e.g., ROI), or it can be trained with historical actions as the correct labels, as we discuss later.[17]

[16] Appendix F documents how this model was trained for different contexts.

[17] Cong (2025) further discusses how behavioral and computational constraints can be incorporated.

4.2 Modeling Lender Responses

In the subsequent analysis, we consider three scenarios with different levels of lender responses to borrowers' usage of LLMs and different loan application data.

- No GenAI: We train the lender model using the training set from the historical data, and then report its performance on the test set from the historical data.
- Small-Scale Adoption: Some borrowers adopt ChatGPT to help generate loan descriptions, but lenders do not respond because they perceive the adoption as too sporadic to have a material effect. Therefore, the lender model is trained on the historical training set, but we report its performance on the counterfactual test set, which is essentially a historical test set with GPT rewrites.
- Large-Scale Adoption: Lenders believe there is large-scale or full adoption of LLMs by the borrowers. Therefore, they adjust the lending model by training it using a counterfactual training set, which is essentially a historical training set with GPT rewrites. We then report its performance on a counterfactual test set.

In sum, the lender model is trained using data from a period without GenAI tools (i.e., using the original text as the input), which applies to both "No GenAI" and "Small-Scale Adoption" scenarios. In reality, lenders can update their decision models, particularly when GenAI adoption is widespread. With the large-scale adoption of GenAI, lenders' decision-making processes are likely to be adjusted. In other words, the predictive model for simulating lender decisions should be trained primarily on ChatGPT-generated data. Figure 3 illustrates the three cases of adoption.

Since most existing studies focus on the "No GenAI" scenario, We will focus on Small-Scale Adoption" and "Large-Scale Adoption" compare them.

4.3 LLM-Generated Descriptions and Loan Fundability

As GenAI helps produce high-quality writing, it may help loan applicants enhance their fundability. To determine if the probability of obtaining funding increases after rewriting, we compute the predicted funding probability using the rewritten versus the original text as the input. The results are shown in Figure 4, in which each point represents a loan.

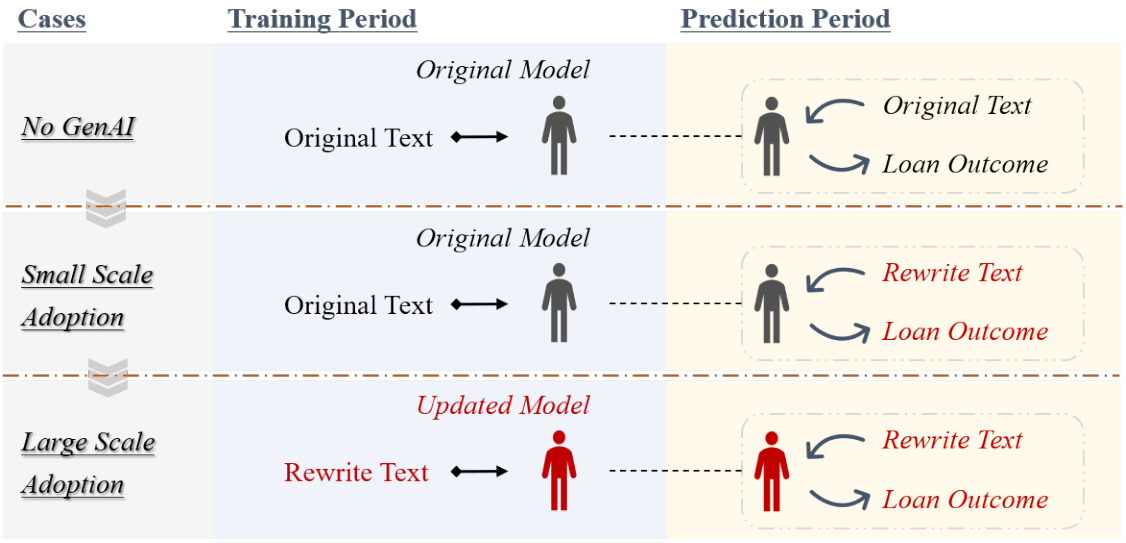
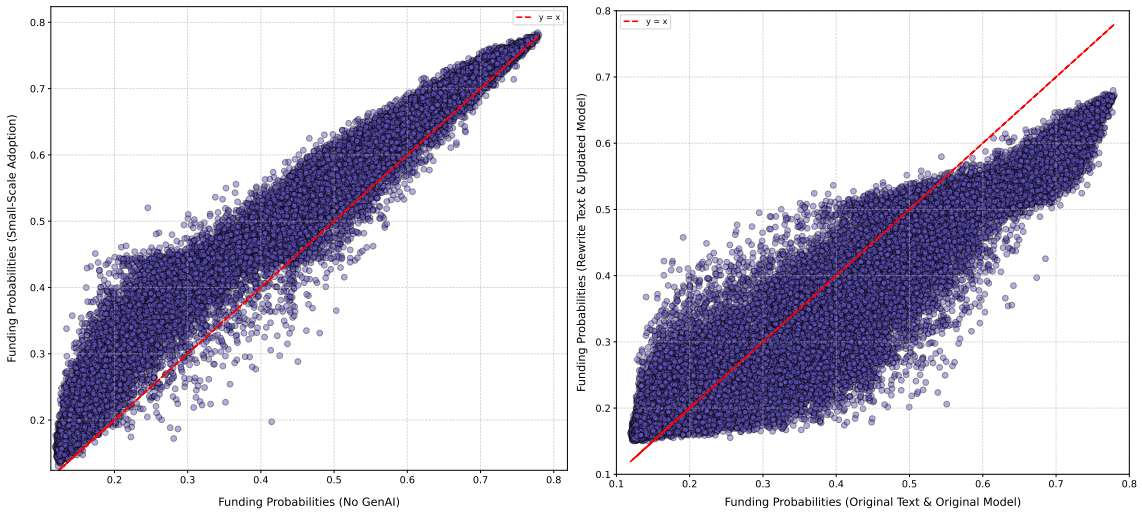


Figure 3: GenAI Adoption Scenarios



(a) Small-Scale Adoption (b) Large-Scale Adoption

Figure 4: Impact of ChatGPT on Fundability

Figure 4(a) shows the comparison of predicted probability of funding under the "Small-Scale Adoption" scenario (y-axis) compared with the "No GenAI" scenario (x-axis). We find that, using rewritten text as the input, our model generally produces a higher predicted probability of funding, as most data points are above the 45-degree line. Similarly, Figure 4(b) shows the comparison of predicted probability of funding under the "Large-Scale Adoption" scenario (y-axis) compared with the "No GenAI" scenario (x-axis). We observe that most points fall below the 45-degree line, indicating that the predicted probability of funding has decreased. This indicates that as widespread adoption improves the writing quality of loan applications for all borrowers, lenders adjust how they use the texts, and the funding probability drops back to be close to the pre-LLM era. The lenders' endogenous responses matter.

4.4 LLM-Generated Descriptions and Soft Information

A utility-based measurement of soft information. We measure soft information as the predictability of textual data on loan quality/outcomes, in addition to other non-textual information. In Eqn. (4), the utility function U is a performance metric, such as the AUC or ROI assessed on the test set using a model trained on the training set.

$$\text{soft-information} = U(\text{text}, \text{structured}) - U(\text{structured}) \tag{4}$$

The results are presented in Table 3. When using AUC as the measure of utility, The best-performing model using structured features only achieved an AUC of 74.1%.[18] When both text and structured data are used, our lender model reaches an AUC of 75.0% in the "No GenAI" scenarios, showing an increase of 0.9 pp (or a 1.2% relative increase). This increase may be considered soft information, quantified by the AUC.[19]

In addition to considering the risk-return trade-off between defaulted and repaid loans, we

[18] When using structured features only, a powerful traditional machine learning model often outperforms deep neural network models (Grinsztajn et al., 2022). Therefore, we consider several state-of-the-art machine learning models and their ensemble (instead of a multimodal transformer) as the "lender" model. See Appendix G for model benchmarking details, where the ensemble model achieves an AUC of 74.1% when using structured features only.

[19] Note that when using both structured features and text, the best performing traditional machine learning model achieves an AUC of 74.5% (see Appendix G). Our lender model, a multimodal model with Heckman correction, achieves a higher AUC. This indicates that our model outperforms traditional machine learning models and extracts a higher amount of soft information from text.

also examine soft information as measured by the ROI.[20] The best-performing model when using structured features only is the XGB model, achieving an ROI of 14.4%. When both text and structured data are used, our lender model reaches an ROI of 23.9% in the "No GenAI" scenario, showing an increase of 9.5 pp (or a 66% relative increase).

Table 3: Measuring Soft Information Using AUC and ROI

	AUC			ROI		
Scenario	Model 1A (Structured)[a]	Model 1B (Structured + Text)	Diff.[b] (Soft Info.)	Model 2A (Structured)[a]	Model 2B (Structured + Text)	Diff.[b] (Soft Info.)
No GenAI		75.0%	0.9%***		23.9%	9.5%***
Small-Scale Adoption	74.1%	74.2%	0.1%	14.4%	14.5%	0.1%
Large-Scale Adoption		75.0%	0.9%***		20.6%	6.2%***

[a] The performance metric does not change by LLM adoption since the model does not use text.
[b] Sig. levels: * $p < .05$, ** $p < .01$, *** $p < .001$. All AUCs and ROIs are averaged across 15 replications of the ten-fold mean.

The "No GenAI" scenario is straightforward and consistent with existing literature, which reports positive information value of texts in online lending (Netzer et al., 2019; Gao et al., 2023).

Soft information lost in LLM adoption. The second row of Table 3 shows results in the "Small-Scale Adoption" scenario. Specifically, our lender model reports a 74.2% AUC and 14.5% ROI. Compared with the utility without leveraging text, the quantified soft information would be reduced by 0.1 pp by AUC and 0.1 pp by ROI, both without statistical significance. In other words, comparing the "Small-Scale Adoption" scenario with the "No GenAI" scenario, we observe a reduction in soft information. Text has almost no information value to improving lender decisions.

The reduction in information value is somewhat mitigated once lenders are allowed to respond to large-scale LLM adoption. This can be observed in the third row of Table 3. Specifically, the AUC recovered to 75.0% and the ROI increased to 20.6%, reflecting increases in 0.9 pp and 6.2 pp, respectively. At first glance, this suggests that some "lost information" was recaptured. That is not the case. Lenders' response to shifting to using more hard information drives the results.

Convergence in writing and soft vs. hard information. The reduction in soft information may be partially attributed to the convergence of writing (Doshi and Hauser, 2023). To demonstrate this point, we compute a homogeneity index based on the text descriptions' pairwise similarity

[20] See Section F.1 for our calculation details.

using credit scores and compare them before and after the rewrites. [21] The results are presented in Figure 5, from which we can observe several interesting patterns. First, borrowers with better credit profiles show much less homogeneity than those with worse credit profiles do before the rewrites. This indicates that high-credit borrowers write in a more diverse manner than low-credit borrowers. Second, after rewriting, the homogeneity of each group increased substantially, with the high-credit groups increasing more than the low-credit groups, narrowing the gap in homogeneity across credit scores. This finding suggests that the application of ChatGPT increased the homogeneity of writing. Convergence in writing may explain the reduction in soft information when lenders attempt to distinguish high-quality loans from low-quality ones.

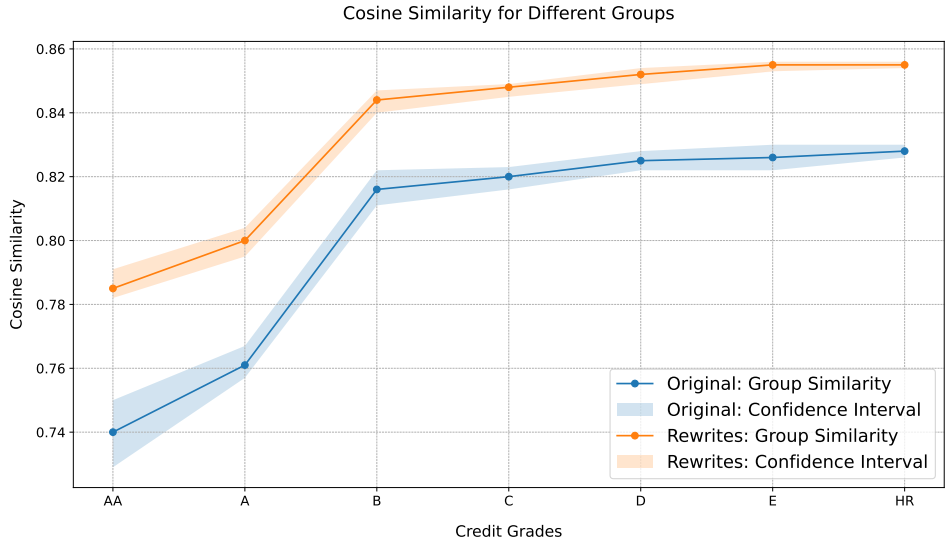


Figure 5: Text Similarity by Credit Grades Before and After Rewriting

Further, we show evidence that moving from the "Small-Scale Adoption" to the "Large-Scale Adoption" scenario, lenders increasingly rely on hard rather than soft information. Specifically, we measure the aggregated feature importance (AFI) among structured and textual features in the lender model.[22] The results are presented in Figure 6. We can see that the AFI score of structured

[21] The homogeneity index is computed as the average cosine similarity between any pair of texts in the group. Namely,

$$\text{Group Homogeneity} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Cosine Similarity}(\text{emb}(\mathbf{x}_i), \text{emb}(\mathbf{x}_j)), \tag{5}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ represent the textual descriptions of loans i and j, respectively, and n represents the total number of samples in this group. Owing to the substantial computational expense of calculating the similarity for all sample pairs, we employed a bootstrap sampling approach (Hernández et al., 2017) to estimate group homogeneity. Specifically, we randomly sample 20% of the data within each group for computation with 10 repetitions.

[22] We compute the AFI based on estimates of feature importance aggregated by feature group. In particular, we

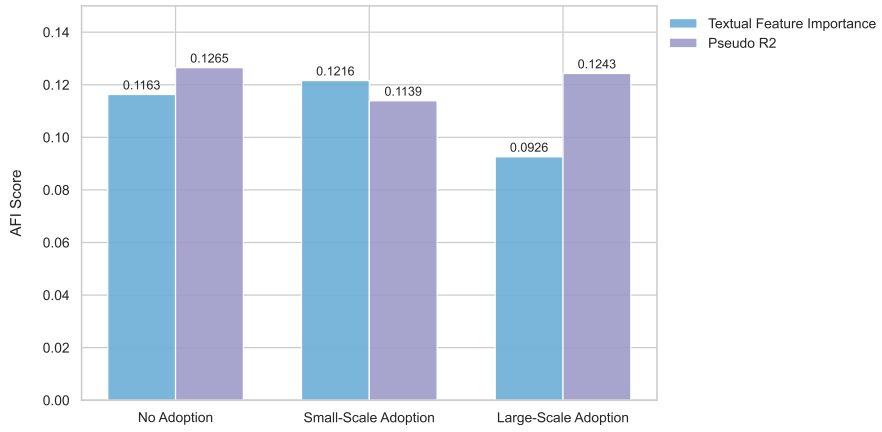


Figure 6: Aggregated Feature Importance (AFI) Scores in Different Scenarios

features is 0.84 in the "Small-Scale Adoption" scenario, which increased to 0.88 in the "Large-Scale Adoption" scenario.[23]

4.5 Credit Misallocation

To analyze the impact of GenAI adoption on lender misallocation, we evaluate misallocation using FPR and FNR as metrics.[24] The results are summarized in Table 4.

Without GenAI, the baseline FPR and FNR are 6.78% and 21.68%, respectively. In the con-

first compute the Shapley value (Štrumbelj and Kononenko, 2014) for each feature in the lender model. Then, the aggregated feature importance is computed as

$$\text{AFI}(\mathcal{S}) = \frac{\sum_{f \in \mathcal{S}} \text{SHAP}(f)}{\sum_{f \in \mathcal{A}} \text{SHAP}(f)}, \tag{6}$$

where $\mathcal{A}$ represents the set of all features in the model, $\mathcal{S} \subseteq \mathcal{A}$ represents a subset of features (e.g., all structured features) that can be aggregated, and $\text{SHAP}(f)$ represents the Shapley value of feature f. The AFI, which ranges from 0 to 1, can be interpreted as the relative weight of a group of features in a model. To make the Shapley value computation efficient, we used the XGBoost model (Chen and Guestrin, 2016) as a surrogate model of the multimodal transformer. Using a surrogate model is a common technique for assessing the Shapley values in deep learning models (Schoch et al., 2023).

[23] The pseudo R^2 reported in Figure 6 is the out-pf-sample McFadden Pseudo R^2, which is calculated as:

$$\text{McFadden Pseudo } R^2 = 1 - \frac{\text{log likelihood (full)}}{\text{log likelihood (null)}},$$

where log likelihood (full) represents the log-likelihood value of the full model and log likelihood (null) refers to the log-likelihood value of the model with only the intercept.

[24] Considering loan approval as the positive outcome, a false positive (FP) case is one in which the model classifies as subsequently repaid and it actually defaults, and a false negative (FN) case is one in which the model classifies as subsequently defaulted but it would actually be repaid. The false positive rate is calculated as the ratio between the number of false positive cases and the number of positive cases (i.e., FPR = FP/P). Similarly, the false negative rate is calculated as ratio between the number of false negative cases and the number of negative cases (i.e., FNR = FN/N).

text of "Small-Scale Adoption," FPR significantly increases to 15.09%, reflecting an increase in 8.31 pp (or a 123% relative increase). In the meantime, FNR significantly decreases to 18.94%, reflecting a decrease of 2.74 pp (or a 13% relative decrease). Since false positives (predicting a loan to repay while it actually went into default) are much more expensive than false negatives, this pattern indicates a substantial increase in misallocation if borrowers leverage GenAI to improve writing, whereas lenders have not updated their decision model. When lenders do update their decision model, as in the large-scale adoption scenario, FPR becomes 12.19%, which is still much higher than the "No GenAI" scenario but reflects a substantial reduction (2.9 pp, or 19% relative change) compared with the "Small-Scale Adoption" scenario. Meanwhile, FNR becomes 19.87%, which is reflects a small increase (0.93 pp, or 5% relative change) compared with the "Small-Scale Adoption" scenario. These trends indicate that, if lenders respond to borrowers' LLM adoption by updating their lending models, FPR partially recovers (at the expense of a small increase in FNR).

Table 4: Misallocation in Different GenAI Adoption Scenarios

Scenario	FPR	DIFF.[a]	FNR	DIFF.[a]
No GenAI	6.78%		21.68%	
Small-Scale Adoption	15.09%	**+8.31%*****	18.94%	- **2.74%*****
Large-Scale Adoption	12.19%	**+5.41%*****	19.87%	- **1.81%*****

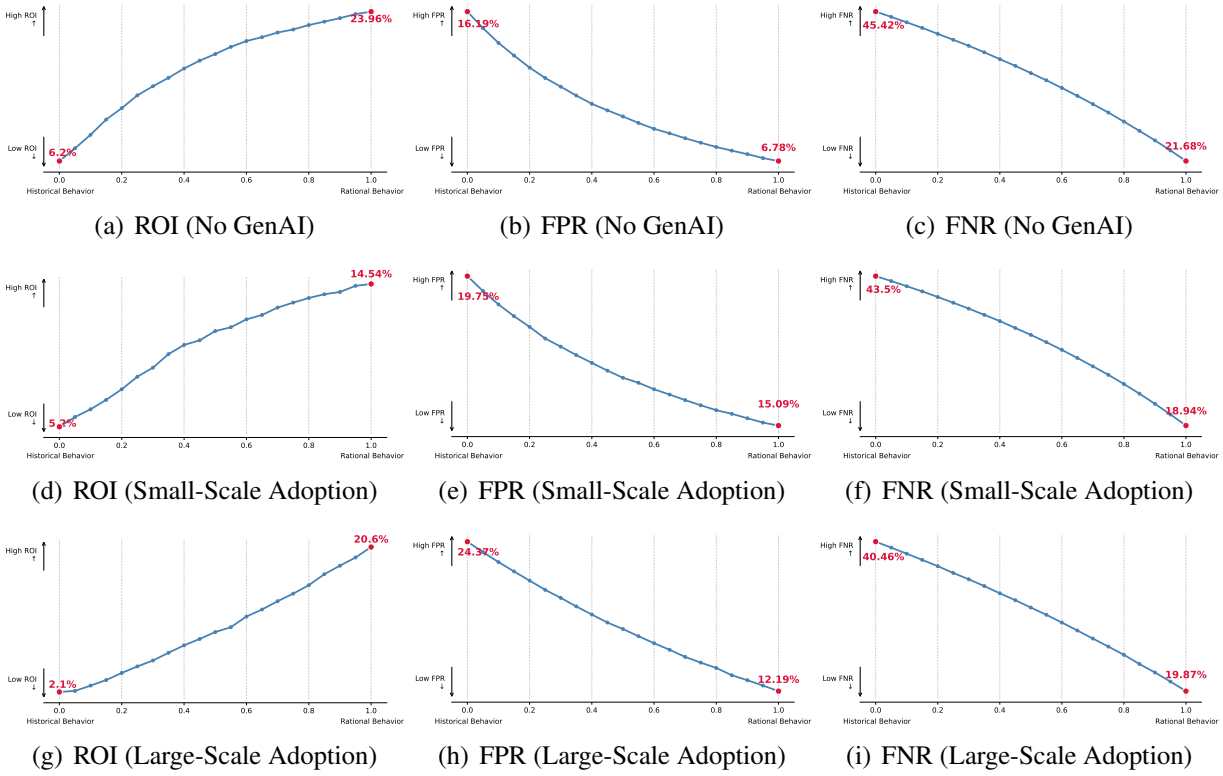
[a.] The differences are relative to the "No GenAI" scenario. Significance tests are based on 15 repetitions of the experiments.
[b.] Sig. levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

5 Constrained Lenders

We now discuss the aggregate implications of writing quality and LLM adoption in the economy by investigating credit allocation efficiency before and after the introduction of LLMs. This section focuses on how lenders' cognitive or computational constraints affect the results.

So far, we have built one of the most cutting-edge multimodal AI models for lenders to examine lenders' decision-making in a world where borrowers are adopting LLMs. However, in our sample period, the lenders did not have access to such powerful tools. In addition, they may face various cognitive constraints, leading to behavioral biases. To model lender decision-making in the online credit market more generally, we first define two decision-making modes: constrained

lender mode $\mathcal{M}_{cl}(\boldsymbol{x})$ and intelligent lender mode $\mathcal{M}_{il}(\boldsymbol{x})$. $\mathcal{M}_{cl}(\boldsymbol{x})$ is trained to mimic historical lending behavior, whereas $\mathcal{M}_{il}(\boldsymbol{x})$ is trained to predict whether the loan is repaid given the current information, and then use the prediction to maximize the ROI. Again, we consider all three scenarios: without GenAI, small-scale adoption of GenAI, and large-scale adoption of GenAI.



(a) ROI (No GenAI) (b) FPR (No GenAI) (c) FNR (No GenAI)

(d) ROI (Small-Scale Adoption) (e) FPR (Small-Scale Adoption) (f) FNR (Small-Scale Adoption)

(g) ROI (Large-Scale Adoption) (h) FPR (Large-Scale Adoption) (i) FNR (Large-Scale Adoption)

Figure 7: Misallocation Estimates in Different Counterfactual Scenarios

We stipulate that lenders' decisions in practice follow a combination of the two modes, which effectively captures that financial constraints, behavioral biases, technological limitations, and computational obstacles prevent the lenders from rationally maximizing financial returns with to-date the most powerful machine learning tool. We use a parameter $\alpha \in [0, 1]$ to indicate the proportion of intelligent lenders in the market, repeat the earlier analyses based on the general specification, and report the results in Figure 7. Each subfigure illustrates the pattern of a performance metric (ROI, FPR, and FNR, respectively, by column) in different market scenarios (by row) when α varies. Several interesting patterns are observed.

First, the ROI increases as the weight of the intelligent lender mode increases, and the FPR and FNR decrease with this weight. This implies that constrained lenders tend to leave money on the

table and have greater misallocation.

Second, the overall pattern through which the lender model adjustment mitigates misallocation is carried through, which is consistent with Tables 3 and 4 in Section 4. In fact, when lenders are constrained (i.e., do not have access to a powerful multimodal AI model to aid their lending decisions), lenders' adjustment in response to the data generated with LLM usage more or less restores the level of soft information usage and allocation efficiency to the pre-GenAI era.

6 A Generative Counterfactual Equilibrium

We now establish the existence and uniqueness of the counterfactual equilibrium in the GenAI age and characterize its aggregate properties. Recall that our ultimate goal is to analyze the counterfactual equilibrium in which borrowers endogenously adopt LLMs and lenders train their models based on the counterfactual data generated. Thus far, we have analyzed two extreme cases where lenders do not react to LLM adoption and adjust their lending decisions based on counterfactual loan applications, assuming borrowers' full adoption of LLMs. We now follow Cong (2025) in leveraging the power of GenAI to analyze a counterfactual general equilibrium.

Borrowers as counterfactual data generator. Each borrower $i \in G$ chooses a high-level action,

$$a_i^H \in \{\text{use ChatGPT}, \text{not use ChatGPT}\} \tag{7}$$

If ChatGPT is used, the loan application text a_i^L is drawn from a generative distribution as follows:

$$a_i^L \sim G_i(\cdot|a_i^H). \tag{8}$$

Borrowers maximize their expected funding probability:

$$\max_{a_i^H} \mathbb{E}\left[P_{\text{funding}}(a_i^H, a_i^L, \sigma_j^*).\right] \tag{9}$$

Lenders as data-driven decision-makers. Lenders $j \in D$ form beliefs about borrower creditworthiness (likelihood of repayment), θ, upon observing counterfactual data D^C:

$$\mu_j(\theta|D^C). \tag{10}$$

Given AI-driven changes in writing patterns, lenders have shifted toward relying more on hard information, such as credit scores and financial history.

The lenders update their decision function σ_j^* by solving:

$$\sigma_j^* = \arg\max_{\sigma_j} \mathbb{E}_{\mu_j}[u_j(\sigma_j(D^C), a^H, a^L, \omega)] \tag{11}$$

General search for equilibria. To solve for an equilibrium, one can follow several steps:

1. Initialize borrower strategies by conjecturing that a particular subset of borrowers use ChatGPT.

2. Simulate Loan Applications: Use generative AI to create counterfactual loan texts based on borrower strategies, and combine this with structural information about the borrowers to generate counterfactual loan applications.

3. Train Lender Models: Train lenders on both historical and generated data, updating their beliefs about loan qualities utilizing the Deep Heckman Correction we develop.

4. Update Borrower Decisions: Given lender responses, borrowers re-optimize their decision to use ChatGPT or not.

5. Iterate Until Convergence: Repeat until borrower adoption and lender responses reach a fixed point, which gives one equilibrium.

Rating-sorted equilibrium. The empirical results show that high-credit borrowers benefit more from AI-enhanced applications. This suggests a natural sorting structure, in which higher-score borrowers are more likely to adopt ChatGPT. Therefore, we focus on rating-sorted equilibria.

We conjecture an GenAI adoption criterion and level. Each borrower has a certain credit profile, takes the equilibrium GenAI adoption rate and thus the lender's model as given, and makes a

decision based on the utility from her use of GenAI. Aggregating all borrowers' adoption decisions and the loan applications generated, we train the proprietary lender's model. Finally, we verify that equilibrium adoption is indeed the conjectured value and borrowers' adoption decisions are rational. We repeat this process until all equilibria are identified.

Specifically, we define borrowers' utility as the improvement in fundability based on the rewrite when compared with fundability using the original text. We also conjecture that high-credit borrowers have the greatest incentive to adopt GenAI. We then check whether the top 1% credit-score borrowers adopt and evaluate the **marginal borrower condition**:

$$\mathbb{E}[\text{Fundability}(\text{ChatGPT}, \sigma^*)] - \mathbb{E}[\text{Fundability}(\text{No ChatGPT}, \sigma^*)] = C_{\text{ChatGPT}} \tag{12}$$

where C_{ChatGPT} is the cost of AI adoption, which we set to zero, given how cheap it is for individual borrowers to access LLMs these days. If the net benefit of adopting is strictly positive, we check the top 2%, and so on, until we find an adoption level that holds equality.

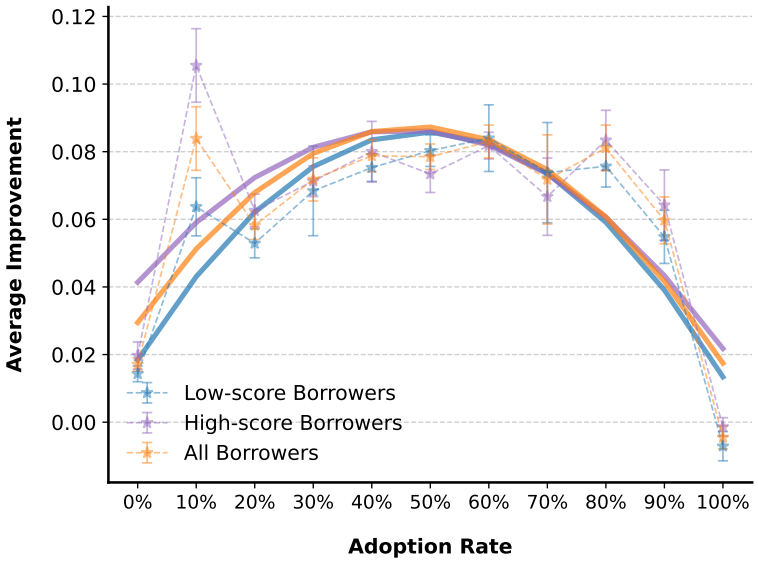


Figure 8: Market Equilibrium (Dashed lines are results computed at increments of 10% market adoption rate, and solid lines are smoothed curves.)

The results of the search by the marginal borrower are shown in Figure 8.[25] As the adoption rate increases from 0 to approximately 0.5, there is a rapid increase in fundability improvements

[25] For different market adoption rates (x-axis), we compute the change in funding probability (y-axis). The change in funding probability is defines as the probability of funding when adopting GenAI to produce a rewrite minus that of not adopting GenAI, as in Eqn. (12).

for all borrowers; however, the largest increase is for high-credit borrowers. As the adoption rate continues to increase to approximately 0.5, this improved funding rate begins to taper off, and is essentially similar across all credit groups. As the probability of adoption approaches 1, the increased gains for all borrowers reach a minimum, but at the same rate of adoption, high-credit borrowers still benefit the most. Despite some heterogeneity (e.g., higher credit borrowers receive higher utility, mid-level market adoption leads to higher personal utility), almost all borrowers, irrespective of their credit scores or assumed market adoption rates, receive positive utility from GenAI adoption. Therefore, if borrowers are rational and the cost of GenAI use is minimal, then all borrowers are motivated to adopt GenAI, resulting in a unique full-adoption long-run equilibrium.

In this unique credit-sorted equilibrium, every borrower adopts GenAI for online loan applications and lenders' decision models are data-driven and trained on synthetic rewrites. This is exactly the Large-Scale Adoption scenario. As we have shown earlier, the loss of soft information is somewhat mitigated relative to partial equilibrium outcomes, but the lender still relies more on hard information compared to the case without GenAI.

7 Conclusion

We utilize both commercially available LLM and proprietary multimodal AI models to investigate how the writing quality of online loan applications and the emergence of LLMs affect lending decisions and credit allocation. We simulate borrowers' use of LLMs to generate counterfactual loan applications and then model lender decisions with a multimodal transformer that accounts for sample observation bias using a novel Deep Heckman Correction and incorporates lenders' cognitive or computational constraints in reduced form. We also characterize a unique general equilibrium in which borrowers' adoption of LLMs is monotone in their credit rating scores.

We find that LLMs can significantly improve writing quality and loan fundability. Even with advanced lender models, ChatGPT-generated applications in aggregate reduce the predictive power of texts and, thus, the utility of soft information in the market, likely owing to the convergence in writing. Moreover, borrowers' strategic use of LLMs could lead to a systematic gap between perceived and actual loan risk, increasing the likelihood of credit misallocation. However, once lenders endogenously respond to applications generated by LLMs, they focus more on hard infor-

mation and misallocation is mitigated.

Our findings have the following implications: First, writing matters in online lending and borrowers, especially those with high credit scores, should improve the writing quality of their loan applications (e.g., using ChatGPT). Second, lenders and regulators should anticipate an increase in writing quality and decrease in information value from text-based loan descriptions. If borrower-generated text can no longer reliably signal behavioral traits, lenders will need to re-evaluate the weight they assign to textual information to focus more on hard information or rely on alternative forms of soft information, such as digital footprints or social network data. These platforms can also mandate the disclosure of LLM usage, allowing lenders to recalibrate trust in textual narratives.[26] More generally, our framework for analyzing counterfactual responses and general equilibria can be applied to other settings.

[26] Studies have developed methods to detect LLM usage. For example, see Cong et al. (2024).

References

Turan G. Bali, Bryan T. Kelly, Mathis Mörke, and Jamil Rahman. Machine forecast disagreement. *NBER Working Paper*, 31583, 2023.

Tobias Berg, Valentin Burg, Ana Gombović, and Manju Puri. On the rise of FinTechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7):2845–2897, 2020.

Donald E. Bowen III, S. McKay Price, Luke C.D. Stein, and Ke Yang. Measuring and mitigating racial disparities in large language model mortgage underwriting. *Social Science Research Network*, 4812158, 2024.

Sean Cao, Wei Jiang, Baozhong Yang, and Alan L Zhang. How to talk when a machine is listening: Corporate disclosure in the age of AI. *The Review of Financial Studies*, 36(9):3603–3642, 2023.

Sean Cao, Wei Jiang, Junbo Wang, and Baozhong Yang. From man vs. machine to man+machine: The art and AI of stock analyses. *Journal of Financial Economics*, 160:103910, 2024.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

Yifei Chen, Bryan T Kelly, and Dacheng Xiu. Expected returns and large language models. *Social Science Research Network*, 4416687, 2022.

Lin William Cong. Data-driven generative equilibrium, 2025. Working Paper.

Lin William Cong, Ke Tang, Jingyuan Wang, and Yang Zhang. AlphaPortfolio: Direct construction through reinforcement learning and interpretable AI. *Social Science Research Network*, 3554486, 2019.

Lin William Cong, Yao Lu, Hanqing Shi, and Wu Zhu. Automation-induced innovation shift. *Social Science Research Network*, 5049949, 2024.

Can Deniz Deveci, Jason Joe Baker, Binyamin Sikander, and Jacob Rosenberg. A comparison of cover letters written by ChatGPT-4 or humans. *Danish Medical Bulletin (Online)*, 70(11), 2023.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 4171–4186, 2019.

Gregor Dorfleitner, Christopher Priberny, Stephanie Schuster, Johannes Stoiber, Martina Weber, Ivan de Castro, and Julia Kammler. Description-text related soft information in peer-to-peer lending–evidence from two leading European platforms. *Journal of Banking and Finance*, 64: 169–187, 2016.

Anil R Doshi and Oliver Hauser. Generative artificial intelligence enhances creativity. *Social Science Research Network*, 4535536, 2023.

Trevor Fitzpatrick and Christophe Mues. How can lenders prosper? comparing machine learning approaches to identify profitable peer-to-peer loan investments. *European Journal of Operational Research*, 294(2):711–722, 2021.

Qiang Gao, Mingfeng Lin, and Richard Sias. Words matter: The role of readability, tone, and deception cues in online credit markets. *Journal of Financial and Quantitative Analysis*, 58(1): 1–28, 2023.

Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems*, 35: 507–520, 2022.

Ken Gu and Akshay Budhkar. A package for learning on tabular and text data with transformers. In *Proceedings of the 3rd Workshop on Multimodal Artificial Intelligence*, pages 69–73, Jun 2021.

Yanhong Guo, Wenjun Zhou, Chunyu Luo, Chuanren Liu, and Hui Xiong. Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, 249(2):417–426, 2016.

Jeffrey T Hancock, Lauren E Curry, Saurabh Goorha, and Michael Woodworth. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45(1):1–23, 2007.

Sarah K Harkness. Discrimination in lending markets: Status and the intersections of gender and race. *Social Psychology Quarterly*, 79(1):81–93, 2016.

James Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–161, 1979.

Steffen Herbold, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva, and Alexander Trautsch. A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13(1):18617, 2023.

Wilmar Hernández, Carlos Calderón-Córdova, Vicente González-Posada, Ángel Parra-Cerrada, José Luis Jiménez, José Enrique González-Garcia, José G Zato, and Oleg Yu Sergiyenko. Bootstrap-based frequency estimation method. *Measurement*, 95:193–200, 2017.

Juei-Ling Ho, Kuan-Ying Chen, Lan-Hsun Wang, Shih-Shuo Yeh, and Tzung-Cheng Huan. Exploring the impact of social media platform image on hotel customers' visit intention. *International Journal of Contemporary Hospitality Management*, 34(11):4206–4226, 2022.

Byoung-Hyoun Hwang and Hugh Hoikwang Kim. It pays to write well. *Journal of Financial Economics*, 124(2):373–394, 2017.

Rajkamal Iyer, Asim Ijaz Khwaja, Erzo FP Luttmer, and Kelly Shue. Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577, 2016.

Hyungu Kahng, Hyungrok Do, and Judy Zhong. Domain generalization via heckman-type selection models. In *The 11th International Conference on Learning Representations (ICLR)*, 2023.

Chansog Kim, Ke Wang, and Liandong Zhang. Readability of 10-k reports and stock price crash risk. *Contemporary Accounting Research*, 36(2):1184–1216, 2019.

Laura Larrimore, Li Jiang, Jeff Larrimore, David Markowitz, and Scott Gorski. Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, 39(1):19–37, 2011.

Yiyi Li and Ying Xie. Is a picture worth a thousand words? an empirical study of image content and social media engagement. *Journal of Marketing Research*, 57(1):1–19, 2020.

Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.

Tim Loughran and Bill McDonald. Measuring readability in financial disclosures. *The Journal of Finance*, 69(4):1643–1671, 2014.

G Harry McLaughlin. SMOG grading—a new readability formula. *Journal of Reading*, 12(8): 639–646, 1969.

Oded Netzer, Alain Lemaire, and Michal Herzenstein. When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, 56(6):960–980, 2019.

Fernando Nogueira. Bayesian Optimization: Open source constrained global optimization tool for Python, 2014. URL https://github.com/bayesian-optimization/BayesianOptimization.

Ellen K Nyhus and Paul Webley. The role of personality in household saving and borrowing behaviour. *European Journal of Personality*, 15(S1):S85–S103, 2001.

Daniel M Oppenheimer. Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 20(2):139–156, 2006.

Myle Ott, Claire Cardie, and Jeff Hancock. Estimating the prevalence of deception in online review communities. In *Proceedings of the 21st international conference on World Wide Web*, pages 201–210, 2012.

Joe Ryan, Katya Reuk, and Charles Wang. To fund or not to fund: Determinants of loan fundability in the prosper.com marketplace. Technical report, 2007. URL https://www.prosper.com/Downloads/Research/Prosper_Regression_Project-Fundability_Study.pdf.

Stephanie Schoch, Ritwick Mishra, and Yangfeng Ji. Data selection for fine-tuning large language models using transferred shapley values. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, volume 4: Student Research Workshop, pages 266–275, July 2023.

Michael Siering. Peer-to-peer (P2P) lending risk management: Assessing credit risk on social lending platforms using textual factors. *ACM Transactions on Management Information Systems*, 14 (3), Jun 2023. Article Number 25.

Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41:647–665, 2014.

Yla R Tausczik and James W Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.

Paul C Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168, 2007.

Paul C Tetlock, Maytal Saar-Tsechansky, and Sofus Macskassy. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467, 2008.

Catalina L Toma and Jeffrey T Hancock. What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication*, 62(1):78–97, 2012.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez,

Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, page 6000–6010, 2017.

Christoph Zott and Quy Nguyen Huy. How entrepreneurs use symbolic management to acquire resources. *Administrative Science Quarterly*, 52(1):70–105, 2007.

Appendix A Sample Descriptives and Baseline Features

We compile a list of features as detailed in Table A1. Table A2 reports descriptive statistics separately by loan status (funded vs. unfunded, and repaid vs. default) groups.

Table A1: Feature Definitions

ID	Feature	Description
Panel A: Loan Request and Verifiable Information		
1	AmountRequested	Natural logarithm of the loan amount applied by the borrower.
2	AmountDelinquent	Natural logarithm of one plus the total delinquency amount in dollars.
3	DelinquenciesLast7Years	Natural logarithm of one plus the total number of delinquencies recorded in the past seven years.
4	InquiriesLast6Months	Total number of credit inquiries made in the past six months.
5	PublicRecordsLast10Years	Total count of public records, including bankruptcies, civil judgments, and tax liens over the past ten years.
6	PublicRecordsLast12Months	The total count of public records, including bankruptcies, civil judgments, and tax liens from the past year.
7	CurrentCreditLines	Natural logarithm of one plus the borrower's current number of credit lines.
8	OpenCreditLines	Natural logarithm of one plus the number of open credit lines held by the borrower.
9	RevolvingCeditBalance	The natural logarithm of one plus the borrower's revolving credit balance in dollars.
10	BankCardUtilization	Ratio of the total balances owed by the borrower on open bankcards to the total credit limits of those cards.
11	CreditGrade	Letter representing the borrower's credit score grade at the time of application.
Panel B: Self-Disclosed Information		
12	DebtToIncomeRatio	The ratio of a borrower's total debt to their income, expressed as a percentage.
13	HasImage	Indicator showing whether the listing contains an image.
14	IsBorrowerHomeowner	Indicator representing home ownership status (owned vs. not owned).
15	EmploymentStatus	Indicators for employment status: unknown, working full-time, not employed, working part-time, retired, and self-employed.
16	LengthStatusMonths	Duration of the current employment status measured in months as of the listing's creation date.
17	IncomeLevel	Income level indicators, including below $25k, $25k to under $50k, $50k to under $75k, $75k to under $100k, above $100k, and missing information.
18	HasLoanDescription	Whether there is a non-empty description.
19	Text Length	Total word count in the loan description provided.
20	Linguistic Conformity	The number of spelling and grammatical errors divided by the total number of words.
21	Formality	Percentage of words with six or more letters when the loan description is nonempty.
22	Readability	The "simple measure of gobbledygook" (SMOG) readability index.
23	Tone	Sentiment scores of loan descriptions.
Panel C: Lending Outcomes (observed during or after the bidding stage)		
24	BorrowerMaximumRate	The maximum rate the borrower agrees to pay for the loan.
25	BorrowerRate	The rate the borrower pays if the loan closes at this point in time.
26	LenderRate	The rate that lenders would receive on the listing if the loan was to close at this point in time.
27	BidMaximumRate	The Maximum Rate in which a bidder will be able to lend money to create a winning bid.
28	BidCount	The total number of bids received by the listing.
29	FundLabel	An indicator that equals one if a listing is fully funded and becomes a loan.
30	RepayLabel	An indicator that equals one if the loan is fully repaid.

Table A2: Summary Statistics (All Listings)

		Stage 1 Outcomes				Stage 2 Outcomes			
		Unfunded (U) (n=211,660)		Funded (F) (n=19,484)		Default (D) (n=6,886)		Repaid (R) (n=12,598)	
ID	Numeric Features	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
1	AmountRequested	8.661	0.825	8.434	0.853	8.393	0.845	8.509	0.862
2	AmountDelinquent	4.613	3.911	1.715	3.135	2.317	3.473	1.385	2.882
3	DelinquenciesLast7Years	1.569	1.414	0.820	1.204	0.994	1.287	0.725	1.146
4	InquiriesLast6Months	4.048	4.730	2.580	3.551	3.584	4.372	2.031	2.863
5	PublicRecordsLast10Years	0.628	1.345	0.368	0.885	0.483	1.084	0.305	0.746
6	PublicRecordsLast12Months	0.077	0.338	0.038	0.235	0.049	0.279	0.032	0.207
7	CurrentCreditLines	1.995	0.708	2.195	0.617	2.176	0.662	2.205	0.591
8	OpenCreditLines	1.890	0.701	2.056	0.617	2.038	0.661	2.066	0.591
9	RevolvingCreditBalance	6.990	3.452	7.880	2.925	7.686	3.176	7.986	2.773
10	BankCardUtilization	0.640	0.433	0.552	0.373	0.604	0.388	0.524	0.361
12	DebtToIncomeRatio	0.515	1.283	0.329	0.891	0.381	1.046	0.300	0.793
16	LengthStatusMonths	23.31	54.58	23.70	54.80	27.19	57.71	21.80	53.04
19	Text Length	194.1	126.4	222.2	139.4	235.1	146.7	215.1	134.7
20	Linguistic Conformity	0.111	0.052	0.095	0.048	0.098	0.050	0.093	0.047
21	Formality	0.240	0.058	0.236	0.055	0.237	0.056	0.235	0.055
22	Readability	10.52	4.096	10.88	3.405	10.78	3.485	10.94	3.359
23	Tone	0.729	0.158	0.751	0.158	0.744	0.158	0.755	0.159
24	BorrowerMaximumRate	0.209	0.092	0.218	0.080	0.246	0.072	0.202	0.080
25	BorrowerRate	0.208	0.092	0.182	0.077	0.213	0.074	0.166	0.073
26	LenderRate	0.207	0.093	0.181	0.077	0.211	0.074	0.165	0.073
27	BidMaximumRate	0.207	0.093	0.181	0.077	0.210	0.074	0.164	0.073
28	BidCount	7.756	38.39	154.5	148.8	151.4	153.9	156.2	145.9
	Categorical Features	n	(%)	n	(%)	n	(%)	n	(%)
11	CreditGrade								
	CreditGrade = AA	5,704	2.7%	2,571	13.2%	448	6.5%	2,123	16.9%
	CreditGrade = A	8,349	3.9%	2,529	13.0%	654	9.5%	1,875	14.9%
	CreditGrade = B	13,145	6.2%	3,373	17.3%	1,132	16.4%	2,241	17.8%
	CreditGrade = C	24,943	11.8%	4,166	21.4%	1,523	22.1%	2,643	21.0%
	CreditGrade = D	36,172	17.1%	3,568	18.3%	1,436	20.9%	2,132	16.9%
	CreditGrade = E	37,700	17.8%	1,657	8.5%	791	11.5%	866	6.9%
	CreditGrade = HR	85,647	40.5%	1,620	8.3%	902	13.1%	718	5.7%
13	HasImage	113,492	53.6%	13,114	67.3%	4,594	66.7%	8,520	67.6%
14	IsBorrowerHomeowner	74,181	35.0%	9,205	47.2%	3,429	49.8%	5,776	45.8%
15	EmploymentStatus								
	EmploymentStatus = FT	172,077	81.3%	16,722	85.8%	5,843	84.9%	10,879	86.4%
	EmploymentStatus = SE	20,108	9.5%	1,474	7.6%	644	9.4%	830	6.6%
	EmploymentStatus = PT	8,397	4.0%	754	3.9%	209	3.0%	545	4.3%
	EmploymentStatus = R	5,927	2.8%	377	1.9%	148	2.1%	229	1.8%
	EmploymentStatus = NE	5,151	2.4%	157	0.8%	42	0.6%	115	0.9%
17	IncomeLevel								
	Income = $0 or unavailable	17,687	8.4%	767	3.9%	277	4.0%	490	3.9%
	Income = $1-24,999	35,386	16.7%	2,376	12.2%	866	12.6%	1,510	12.0%
	Income = $25,000-49,999	86,764	41.0%	7,272	37.3%	2,646	38.4%	4,626	36.7%
	Income = $50,000-74,999	43,983	20.8%	4,940	25.4%	1,708	24.8%	3,232	25.7%
	Income = $75,000-99,999	15,547	7.3%	2,214	11.4%	758	11.0%	1,456	11.6%
	Income = $100,000+	12,293	5.8%	1,915	9.8%	631	9.2%	1,284	10.2%

Appendix B Sample Rewrites

In this section, we present some examples of LLM-generated loan descriptions. As shown in Table B3, ChatGPT rewrites significantly improve the readability and formality of loan descriptions.

Table B3: Examples of Loan Description Rewrites

Original	Rewrite
My financial situation: I am a good candidate for this loan because I am preparing for a future business so it is imperative that I get reed of these high intrest debts and stop making the credit card companies rich. I will never miss a payment one because Im using direct payments two because I want to use this site later on for my business loan.	Financial Situation: I present myself as an ideal candidate for this loan, as my unwavering commitment to fiscal responsibility is evident. With the vigor of an entrepreneur striving to secure a prosperous future, I am determined to eliminate these burdensome debts. By employing direct payment methods, I guarantee the punctuality of my monthly installments, reinforcing my reliability. Furthermore, I seek to utilize this esteemed lending platform for my future business aspirations, further exemplifying my commitment to honor my financial obligations. Monthly Monetary Overview: Amidst the discerning evaluation of my income and expenses, my financial prudence is unquestionably apparent.
Monthly net income: $ 3,700	My monthly net income stands at a commendable $3,700, allowing for ample liquidity to address the pertinent matter at hand.
Purpose of loan: The purpose of the loan id get back from a payed vacation that i get won 7 day eastern caribbean curise all payed for by cal worthington (go see cal go see cal) jsut want to have enough money to cover all house hold expense.	Loan Purpose: The purpose of this loan is to efficiently manage my household expenses after returning from an all-expenses-repaid 7-day Eastern Caribbean cruise, graciously won through the renowned Cal Worthington (Go see Cal, go see Cal!). By securing this loan, I intend to ensure that my financial obligations are met seamlessly.
I am using these funds to re-invest in select Prosper loans. I am currently a Banker, a Real Estate Developer and a Citrus Grower. I will be retiring from the Banking industry (at age 44) towards the end of June and will have an income well excess of $100k annually even after leaving and going full time in pursuit of development and citrus. I am borrowing a small amount and probably won't keep for the full term but will keep for at least 9-12 months. At this point, until I see the results of my lending efforts I am not willing to take any undue risks. Small time arbitrage!!Personal debt obligations are minimal and the majority of what shows under revolving is American Express paid off monthly and my Home Equity LOC in the amount of $40K now (was $70K) at time of pull. By the way, $25K of my LOC will be re-paid within 6 months as those were funds loaned to my development entity and the funds will be available within 6 months from that entity as lots are sold...more information than you needed probably. Be happy to answer any questions you may have.	I seek financial support to further invest in carefully chosen Prosper loans. As a seasoned Banker, Real Estate Developer, and Citrus Grower, I have diverse expertise that enhances my investment acumen. Despite my decision to retire from banking at the young age of 44, my earnings will exceed $100,000 annually post-retirement as I focus entirely on real estate development and citrus cultivation. While my loan amount is modest, I plan to retain it for a minimum of 9-12 months, allowing me to assess the outcomes of my borrowing ventures. Prudent decision-making is my priority at this juncture, as I refrain from undertaking unnecessary risks, adopting a strategy akin to small-scale arbitrage. My personal debt is minimal, largely comprising a monthly-paid American Express balance and a Home Equity Line of Credit of $40,000 (previously $70,000). Notably, $25,000 of the LOC is earmarked for repayment within six months, sourced from a loan to my development enterprise, to be recouped upon lot sales. Should you require further information, I am more than willing to address any queries. Your trust in my financial vision is valued and your support is integral to my continued success. Thank you for considering my loan application.

Appendix C Prompt Details

C.1 Prompt Versions

We consider several versions of prompt templates, as summarized in Table C4 and illustrated in Figure C1. These versions correspond to several common scenarios for econometricians.

1. We examine a special case without rule setting, in which borrowers directly use GenAI tools to rewrite their loan applications.

2. We assess the scenario in which LLMs are adopted to enhance readability and emphasize content that could assist lenders in decision-making, termed "Specification 1" (S-1).

3. We assess a scenario in which LLMs are adopted to compel and improve the trustworthiness of the loan description, termed "Specification 2" (S-2).

4. Building on existing research, we designed settings from the perspective of textual features that have been proven to increase fundability, termed "Specification 3" (S-3).

Notably, each setting aims to improve the writing quality to enhance fundability.

Table C4: Prompt Templates Based on Different Specifications

Prompt Template	Role Setting	Task Setting	Rule Setting	Purpose
Version 1	✓	✓	None	Robustness
Version 2	✓	✓	S-1	Robustness
Version 3	✓	✓	S-1 + S-2	Baseline
Version 4	✓	✓	S-1 + S-2 + S-3	Robustness

Version 3 corresponds to our baseline prompt and the results are reported in the main manuscript. For robustness tests, we opt to progressively broaden the rule settings stems from the fact that, with a deeper understanding of ChatGPT, users' techniques and abilities will consistently advance. The results from the other prompt template versions are compared with those from Version 3 in Section D.3.

C.2 Different Rule-Setting Specifications

Rule settings are relatively flexible and likely exhibit variation across borrowers, due to differences in their ability to use LLMs and their cognitive levels.

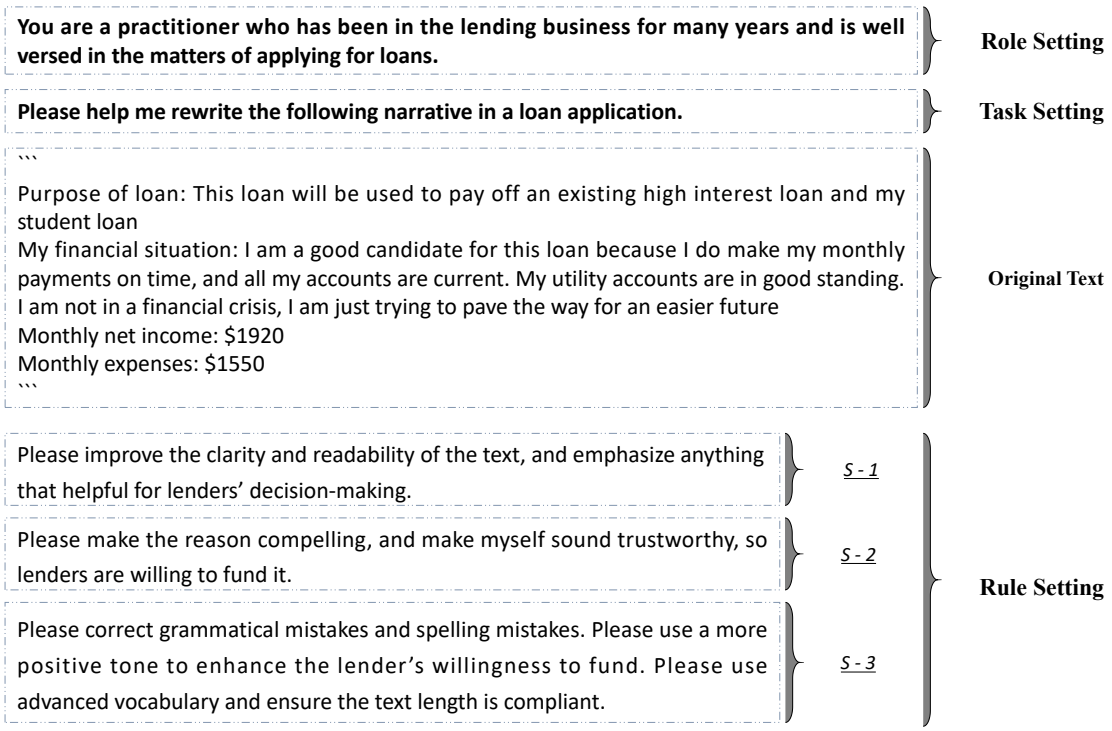


Figure C1: An Example ChatGPT Prompt Showing Different Levels of Specification.

First, a borrower might adopt ChatGPT to enhance readability and emphasize content, which could assist lenders in decision-making. To do so, they might use the sentence textitPlease improve the clarity and readability of the text and emphasize anything that is helpful for lenders' decision-making." in the rule setting of prompts. We term this rule specification as "Specification 1" (S-1). To ensure that the rule setting is not influenced by the exact wording in the sentence, we used ChatGPT to generate ten other ways to express this specification, and the variations are documented in Table C5. When putting together the prompts, we randomly picked one of these ten variations and put it in the full rewrite prompt.

Second, the borrower might want to adopt LLMs to help compel their reason and improve their funding success. This rule setting, termed "Specification 2" (S-2), may be expressed with the sentence "*Please make the reason compelling, and make myself sound trustworthy, so lenders*

Table C5: Variations of **S-1** (The original sentence is "*Please improve the clarity and readability of the text, and emphasize anything that helpful for lenders' decision-making.*")

1	"Please enhance the clarity and readability of the text, highlighting any details that could assist lenders in their decision-making."
2	"Improve the text's clarity and readability, ensuring that important points for lenders' decisions stand out."
3	"Enhance the text to make it clearer and easier to read, with a focus on highlighting information that helps lenders in their decision process."
4	"Please improve the readability of this text, making sure to emphasize key points that aid lenders in their decision-making."
5	"Rewrite the text for improved clarity and readability, drawing attention to any details that could influence lenders' decisions."
6	"Please refine the text to boost clarity and readability, while emphasizing critical information helpful to lenders."
7	"Make the text clearer and easier to read, and highlight any details that would assist lenders in deciding."
8	"Improve the text's readability and clarity, focusing on highlighting points that are crucial for lenders' decision-making process."
9	"Improve the text to ensure it's easy to read and clear, while stressing any information useful for lenders' decisions."
10	"Rewrite the text for clarity and ease of understanding, while highlighting points that would be helpful for lenders in their decision-making."

willing to fund it. " Similar to the above, we generate 10 variations as listed in Table C6, and randomly select one when dynamically generating a full rewrite prompt.

Table C6: Variations of **S-2** (The original sentence is "*Please make the reason compelling, and make myself sound trustworthy, so lenders willing to fund it.*")

1	"Please make the reason persuasive and present myself as reliable, so lenders are eager to provide funding."
2	"Ensure the reason is convincing and I come across as trustworthy, encouraging lenders to fund the loan."
3	"Craft the reason in a compelling way, making me appear dependable, so lenders feel confident in offering support."
4	"Please make the reason persuasive, portraying me as trustworthy, so lenders are inclined to fund it."
5	"Help make the reason strong and convincing, building trust so lenders will want to provide funding."
6	"Ensure the reason is appealing and I appear trustworthy, so lenders feel comfortable funding the loan."
7	"Make the reason convincing and portray me as dependable, so lenders feel confident in offering their support."
8	"Write the reason in a way that inspires confidence and makes me appear trustworthy, so lenders want to fund the loan."
9	"Make sure the reason is strong and I come across as reliable, encouraging lenders to offer funding."
10	"Make the reason compelling and reflect my reliability, so lenders feel comfortable providing funds."

Finally, suppose that the borrower has been exposed to insights from research regarding which listings are more likely to be funded. For example, Netzer et al. (2019) and Gao et al. (2023) reported many characteristics of listings that were more successful in securing funding. Referring to existing research, we put together another sentence termed "Specification 3" (S-3), which says "*Please correct grammatical and spelling mistakes. Please use a more positive tone to enhance*

Table C7: Variations of **S-3** (The original sentence is "*Please correct grammatical mistakes and spelling mistakes. Please use a more positive tone to enhance the lender's willingness to fund. Please use advanced vocabulary and ensure the text length is compliant.*")

1	"Please correct any grammatical or spelling errors. Ensure the tone is more optimistic to encourage the lender's willingness to fund, and use advanced vocabulary while keeping the text length appropriate."
2	"Correct any grammar or spelling issues and adjust the tone to be more positive, making it appealing to lenders. Use sophisticated language and keep the text length within limits."
3	"Please revise grammatical and spelling errors, adopt a more encouraging tone to increase lender interest, and use advanced terminology while ensuring text length is appropriate."
4	"Correct any errors in grammar and spelling, using a more positive tone to encourage lender engagement. Employ advanced vocabulary and maintain a suitable length for the text."
5	"Please edit grammatical and spelling mistakes, ensuring a positive tone to heighten lender interest. Use an elevated vocabulary and ensure the text adheres to the required length."
6	"Correct all grammar and spelling errors, adjusting the tone to be more uplifting and persuasive for the lender. Apply advanced wording and keep the text concise."
7	"Correct any grammatical and spelling issues, ensuring the tone is upbeat to appeal to lenders. Use high-level vocabulary and make sure the text length is suitable."
8	"Revise the text for any grammar or spelling issues, using a more positive tone to encourage funding. Employ advanced language while ensuring the text stays within length requirements."
9	"Correct all grammatical and spelling errors, and ensure a positive tone to increase the lender's confidence. Use elevated vocabulary and ensure the text remains within the required length."
10	"Correct any grammar or spelling mistakes, adopt a more encouraging tone to appeal to lenders, and use higher-level language while ensuring the text stays within the desired length."

the lender's willingness to fund. Please use advanced vocabulary and ensure the text length is compliant." Table C7 lists its ten variants.

Appendix D Fundability and Text Features

Textual features potentially reflect writing quality and exhibit a strong connection to fundability, as documented in many existing studies.

D.1 Text Features and Historical Fundability

How do each of these features help explain fundability, especially when considering their heterogeneous effects on borrowers of different credit grades and when controlling for other textual features? We employ a logistic regression model and control for known loan characteristics. The model is specified as follows.

$$\begin{aligned}\text{logit}(Outcome_i) =& \beta_0 + \beta_1 \times IsHighCredit_i + \beta_2 \times IsLowCredit_i + \beta_3 \times TF_i \\ &+ \beta_4 \times TF_i \times IsHighCredit_i \\ &+ \beta_5 \times TF_i \times IsLowCredit_i + \boldsymbol{\gamma} \cdot \boldsymbol{c}_i + \varepsilon_i, \end{aligned} \tag{13}$$

where $Outcome_i$ represents whether loan i is funded, TF represents a textual feature, $\boldsymbol{c}_i$ represents the vector of control variables, $\boldsymbol{\gamma}$ is the corresponding coefficient vector, and ε_i represents the random error. With this model specification, the estimated effect of the textual feature is $\hat{\beta}_3$ for medium-credit borrowers, $\hat{\beta}_3+\hat{\beta}_4$ for high-credit borrowers, and $\hat{\beta}_3+\hat{\beta}_5$ for low-credit borrowers.[27]

The results are shown in Table D8, where each column considers a model with a textual feature, as indicated in the model name.

[27] We employed a linear model to estimate the credit score using the structured features. The credit scores were then divided into tertiles, with values above the 67% percentiles designated as "high credit", those below the 33% percentiles as "low credit", and the rest as "medium credit."

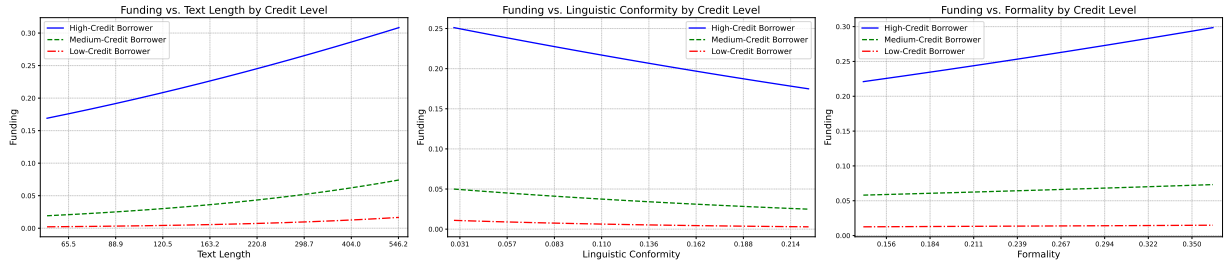
Table D8: Relationship Between Textual Features and Funding Outcome

	DV: Funding (Funded: 1, Unfunded: 0)				
TF ▷	Model 1 (Text Length)	Model 2 (Linguistic Conformity)	Model 3 (Formality)	Model 4 (Readability)	Model 5 (Tone)
TF × High-Credit Borrower	-0.168***	0.070***	0.039*	-0.015	-0.073***
	(0.019)	(0.021)	(0.020)	(0.023)	(0.019)
TF × Low-Credit Borrower	0.172***	-0.163***	-0.019	0.060	0.063*
	(0.031)	(0.035)	(0.032)	(0.040)	(0.031)
Text Length	0.378***	0.276***	0.279***	0.281***	0.280***
	(0.018)	(0.011)	(0.011)	(0.011)	(0.011)
Linguistic Conformity	-0.152***	-0.192***	-0.160***	-0.159***	-0.159***
	(0.010)	(0.019)	(0.010)	(0.010)	(0.010)
Formality	0.089***	0.085***	0.062***	0.087***	0.087***
	(0.009)	(0.009)	(0.018)	(0.009)	(0.009)
Readability	0.030**	0.023*	0.022*	0.027	0.023*
	(0.010)	(0.010)	(0.010)	(0.020)	(0.010)
Tone	0.033***	0.033***	0.033***	0.033***	0.074***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.016)
Amount Requested	-0.594***	-0.594***	-0.594***	-0.594***	-0.596***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Amount Delinquent	-0.698***	-0.698***	-0.697***	-0.697***	-0.697***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Delinquencies Last 7 Years	-0.254***	-0.256***	-0.255***	-0.255***	-0.256***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Inquiries Last 6 Months	-0.256***	-0.259***	-0.259***	-0.259***	-0.261***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Public Records Last 10 Years	-0.042***	-0.041***	-0.040***	-0.040***	-0.040***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Public Records Last 12 Months	-0.028*	-0.027*	-0.027*	-0.028*	-0.027*
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Current Credit Lines	0.176***	0.176***	0.177***	0.177***	0.177***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Open Credit Lines	-0.292***	-0.293***	-0.295***	-0.294***	-0.294***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Revolving Credit Balance	0.275***	0.278***	0.278***	0.277***	0.278***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Bankcard Utilization	-0.425***	-0.431***	-0.429***	-0.429***	-0.430***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Debt To Income Ratio	-0.114***	-0.114***	-0.113***	-0.113***	-0.113***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Length Status Months	0.092***	0.093***	0.094***	0.094***	0.094***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Is Borrower Homeowner	0.630***	0.634***	0.634***	0.634***	0.634***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Image Dummy	0.284***	0.289***	0.290***	0.290***	0.290***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Credit Level (High)	0.612***	0.587***	0.575***	0.574***	0.578***
	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)
Credit Level (Low)	-0.611***	-0.569***	-0.527***	-0.533***	-0.529***
	(0.036)	(0.035)	(0.033)	(0.033)	(0.033)
Employment Status (Full Time)	0.726***	0.729***	0.736***	0.736***	0.738***
	(0.093)	(0.093)	(0.094)	(0.094)	(0.094)
Employment Status (Self Employed)	0.528***	0.534***	0.548***	0.543***	0.543***
	(0.095)	(0.096)	(0.096)	(0.096)	(0.096)
Employment Status (Part Time)	0.518***	0.519***	0.525***	0.524***	0.527***
	(0.101)	(0.101)	(0.101)	(0.101)	(0.101)
Employment Status (Retired)	0.735***	0.739***	0.740***	0.739***	0.743***
	(0.109)	(0.109)	(0.109)	(0.109)	(0.109)
Income L1	-1.326***	-1.333***	-1.330***	-1.330***	-1.330***
	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)
Income L2	-0.699***	-0.707***	-0.705***	-0.704***	-0.706***
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Income L3	-0.400***	-0.408***	-0.406***	-0.406***	-0.407***
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Income L4	-0.187***	-0.193***	-0.191***	-0.191***	-0.191***
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Income L5	-0.017	-0.020	-0.019	-0.019	-0.020
	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
Constant	-3.873***	-3.851***	-3.856***	-3.854***	-3.859***
	(0.099)	(0.099)	(0.099)	(0.099)	(0.099)
Observations	231,144	231,144	231,144	231,144	231,144
Pseudo R2	0.212	0.211	0.210	0.210	0.211

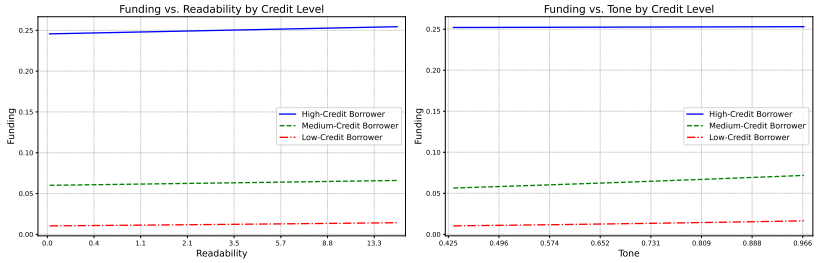
Note: Sig. levels: * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors in parentheses.

From Table D8, we can see that each text feature shows a significant relationship with fundability, at least for certain subgroups with different credit grades. For example, spelling errors have a more significant association with fundability for high-credit borrowers, whereas grammatical errors have a more significant association with fundability for low-credit borrowers. Additionally, tone has a significant positive association with fundability for low-credit borrowers, but this effect is not as prominent for high-credit borrowers, suggesting that low-credit borrowers may rely more on a positive tone to secure loans. In contrast, the readability score has a significant positive association with fundability for high-credit borrowers, indicating that making the description more advanced increases the chances of receiving funding only for high-credit borrowers.

The estimated effects by credit level are shown in Figure D2. We can see a positive relationship between fundability and text length, formality, readability, and tone, and a negative relationship with linguistic conformity. These correlations are closely related to writing quality, as high-quality writing typically features more content, fewer errors, a more positive tone, and higher readability. As depicted in Figure D2, while increased readability seems helpful for high-credit borrowers, Increased formality could decrease the fundability of low-credit borrowers.



(a) Text Length (b) Linguistic Conformity (c) Formality



(d) Readability (e) Tone

Figure D2: Visualizing the Relationship Between Textual Features and Historical Fundability Based on Regression Results in Table D8.

D.2 Attribution Test for Improved Fundability

Our next question is, which feature(s) led to greater improvements in fundability when the loan description text was rewritten by ChatGPT? To investigate this question, we ran an attribution test using the feature-importance approach.

Specifically, we first compute the increased fundability as the difference between fundability using the rewritten text and that using the original text. Then, we fit a predictive model using the increased fundability as the target variable, textual features of the original text, change in textual features between the original and the rewritten text, and the structured features as predictors.[28] Based on this model, we evaluated the feature importance, and a list of the most important features is visualized in Figure D3.

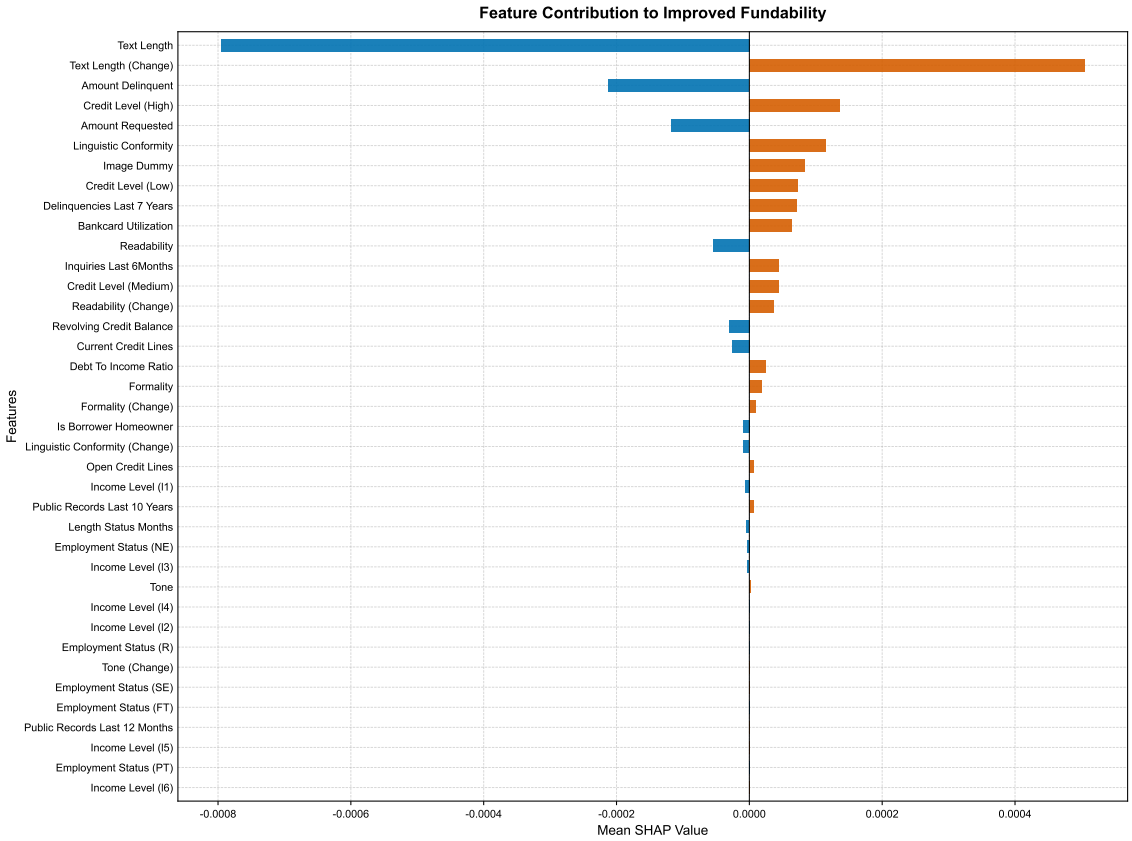


Figure D3: SHAP Value Feature Importance for Predicting Increased Fundability. Features are shown with their positive and negative contributions to the prediction.

We can see that the text length is among the most important variables for predicting improve-

[28] We used the XGBoost model (Chen and Guestrin, 2016) as it is widely known as one of the best performing models.

ment in fundability. In particular, the baseline text length (i.e., the text length of the original loan description) is negatively associated with the improvement in fundability, and the text length change (i.e., the increase of text length via ChatGPT rewriting) is positively associated with the improvement in fundability. Baseline linguistic conformity, and baseline and increase in formality also have positive effects on the improvement in fundability.

D.3 Robustness Results of Prompt Design and Textual Features

Based on the previous setup, we tested the changes in the textual features across different prompt templates, as presented in Table D9.

Table D9: Textual Features Before and After Rewrite

	Original	Version 1		Version 2		Version 3		Version 4	
	Mean (SD)	Mean (SD)	DIF[a] (%)	Mean (SD)	DIF[a] (%)	Mean (SD)	DIF[a] (%)	Mean (SD)	DIF[a] (%)
Text Length	196.4 (127.8)	184.1 (98.9)	**-12.3***** (-6.3%)	198.8 (88.2)	**2.4***** (+1.2%)	231.3 (81.2)	**34.9***** (+17.8%)	237.8 (89.0)	**41.4***** (+21.1%)
Linguistic Conformity	0.110 (0.052)	0.058 (0.041)	**-0.052***** (-47.3%)	0.052 (0.033)	**-0.058***** (52.7%)	0.039 (0.025)	**-0.071***** (-64.5%)	0.042 (0.025)	**-0.068***** (-61.8%)
Formality	0.239 (0.058)	0.328 (0.047)	**0.089***** (+37.2%)	0.336 (0.044)	**0.097***** (+40.6%)	0.342 (0.037)	**0.103***** (+43.1%)	0.363 (0.037)	**0.124***** (+51.9%)
Readability	10.55 (4.043)	14.04 (2.910)	**3.49***** (+33.1%)	14.97 (3.041)	**4.42***** (+41.2%)	15.67 (2.092)	**5.12***** (+48.5%)	16.27 (2.254)	**5.72***** (+54.2%)
Tone	0.731 (0.158)	0.776 (0.146)	**0.045***** (+6.2%)	0.784 (0.143)	**0.053***** 7.3%	0.744 (0.145)	**0.013***** (+1.8%)	0.748 0.144	**0.017***** (+2.3%)

[a] Sig. levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table D9 indicates that while ChatGPT's impact on textual features varies across different prompt templates, the improvements remains consistent. Corroborating the findings in Section 3.3, different prompt versions (as nuanced variations of ChatGPT adoption) led to mostly consistent results: they enhanced length compliance, tone positivity, formality, and readability while reducing spelling and grammatical errors.

Appendix E Technical Information

E.1 Architectures for Text Embeddings

Bidirectional Encoder Representations from Transformers (BERT) were first introduced in Devlin et al. (2019). As a transformer model, BERT is based on transformer architecture (Vaswani et al., 2017), specifically its encoder part. The transformer employs a self-attention mechanism to process the input data, enabling the model to consider the relationships between all the elements in the input sequence.

The basic principles of BERT are as follows: Bidirectional Contextual Understanding and traditional language models are unidirectional and unable to consider both the preceding and following contexts of a word. BERT innovatively uses a bidirectional context, considering the context on both sides of each word to obtain richer language representations. Before training and Fine-tuning, BERT employs a two-stage training process. First, it was pretrained on large-scale text data to learn a general language representation. Then, it was fine-tuned on task-specific data to adapt the model for specific NLP tasks.

The pre-training process of BERT relies on two key techniques: the Masked Language Model (MLM) and next sentence prediction (NSP). MLM sets the objective of maximizing the likelihood of masked tokens. For a masked token, w_m, the objective can be formulated as

$$L_{MLM} = -\sum_{i \in M} \log P(w_m | context) \tag{14}$$

where $context$ denotes unmasked tokens. $P(w_m|context)$ is the probability of the masked word given the context provided by other words. The objective of the NSP is to model the probability of the second sentence (B) in the next sentence. The formula can be represented as

$$L_{NSP} = -(y \log(P)) + (1 - y) \log(1 - P) \tag{15}$$

where L_{NSP} denotes the loss for the NSP task. P is the model's predicted probability that sentence B follows sentence A and y is the actual label (1: if B follows A; 0: otherwise).

SBERT modifies the BERT architecture to produce embeddings that are directly comparable

using cosine similarity, which is more efficient for tasks such as semantic search or similarity assessment. SBERT uses Siamese and triplet network structures to derive semantically meaningful sentence embeddings, which can be compared using cosine similarity. These designs ensure that sentences with similar meanings are close in vector space, whereas those with different meanings are further apart. Additionally, in the training process, triplet loss was used to improve the learning efficiency. The formula for triplet loss is:

$$L = \sum_{i=1}^{N} \max(0, \alpha + d(a_i, p_i) - d(a_i, n_i)), \tag{16}$$

where $d(x, y)$ is a distance metric function to calculate the similarity between the embeddings of sentences x and y, α is the margin, a_i, p_i, and n_i are the anchor, positive, and negative sentences in the ith triplet, respectively.

E.2 Multimodal Transformer

To make predictions using both structured features and textual features, we leverage a multimodal transformer model (Gu and Budhkar, 2021), illustrated in Figure E4.

Essentially, different types of features are first preprocessed in their own domains. While traditional features, such as categorical and numeric features, can be represented with a set of nodes, text features pass through a text embedding model, for which we use the commonly used BERT embedding. More specifically, we initialize the text embeddings with a pretrained BERT model from Huggingface[29], which encodes text into a set of nodes. These traditional feature nodes and text-embedding nodes are then concatenated to create a sample-level numeric vector representation. We then add a multi-layer perceptron (MLP) layer with cross-entropy loss[30] to map the numeric vector representation of the sample onto the outcome class label. The multimodal transformer model is implemented with Pytorch.[31]

In this multimodal transformer model, we let both the text transformer and final MLP layer

[29] https://huggingface.co/

[30] Depending on the data and label being used, this model may be used to classify funding status (among all listings) or repayment status among those funded. Because the target variables for both the selection model and the outcome model are binary, the cross-entropy loss is well suited to our needs. However, in other contexts, the outcome variable may be continuous, for which we may use MSE loss instead.

[31] https://pytorch.org/

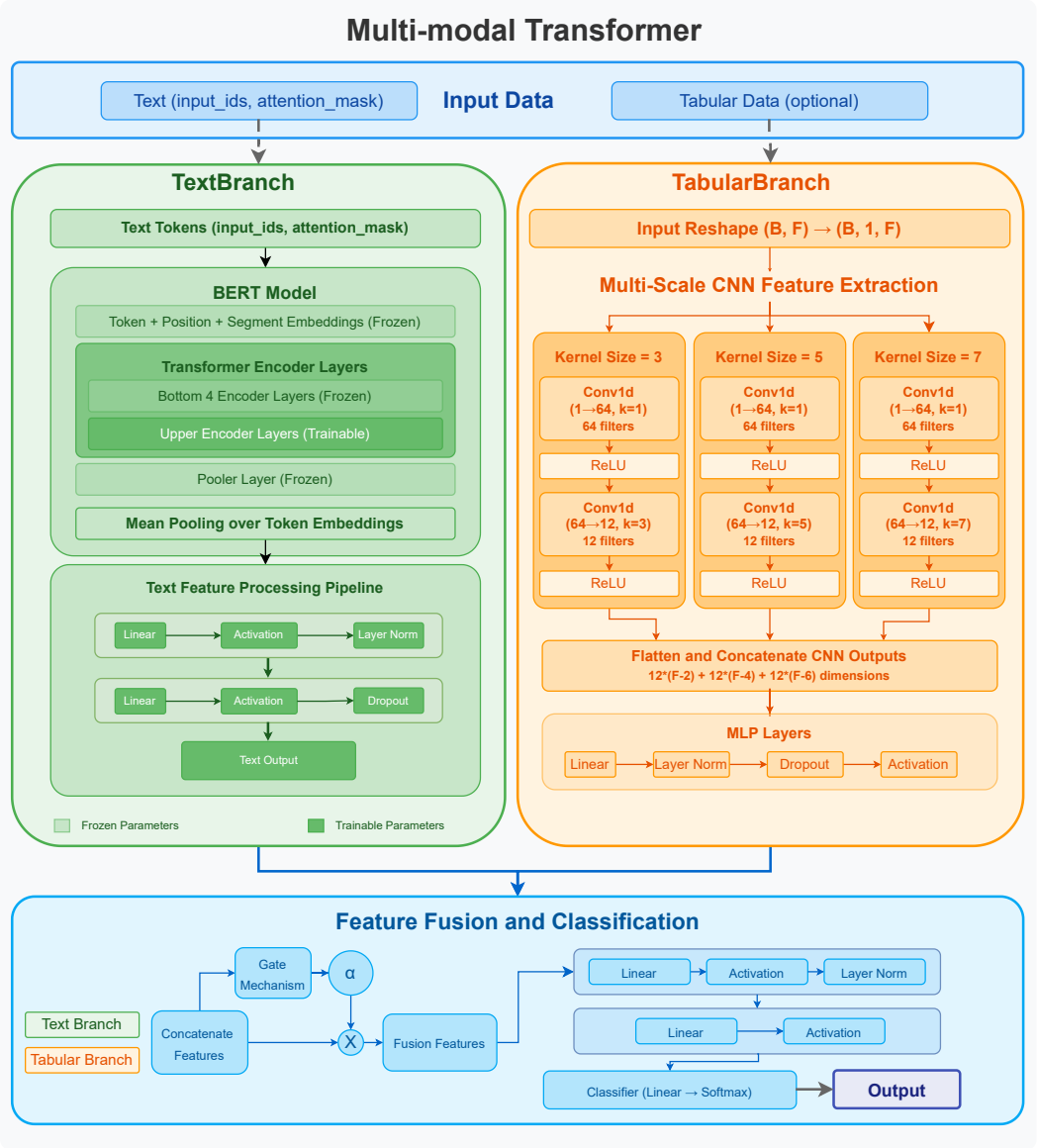


Figure E4: Architecture of the Multimodal Transformer

be learnable. The hyperparameters are listed in Table E10. We used a grid search to determine the best hyperparameters. This multimodal transformer model had approximately 111 million learnable parameters. The tabular branch has 845,344 parameters and the linear MLP step has 395,008 parameters. The other parameters are from the text branch; using BERT embedding, there are 109,744,896 parameters.

Table E10: Hyperparameter Settings for the Multimodal Transformer

Parameter	Description	Range
learning_rate_text	The learning rate of the pre-trained language models.	[3e-6, 3e-5]
learning_rate_tabu	The learning rate of the MLP model, which is employed to learn the features of structured features if activated.	[0.01, 0.015]
weight_decay_text	The regularization part of the language model.	[5e-3, 5e-4, 5e-5]
weight_decay_tabu	The regularization part of the MLP model.	[1e-1, 1e-2, 1e-3]
epoch	Number of epochs to train the model in each round.	[6, 7, 8]
round	Number of rounds to train the model for each group of parameters.	[15]
batch_size	Batch size.	[15, 30]
dropout_rate	The dropout rate of models.	[0.45]
activate_func	Activate functions.	['relu','leaky_relu']

E.3 Deep Heckman Model Estimation

Our Heckman-based lender model consists of two versions of this multimodal transformer, one for the selection model f, and the other for the prediction model g (see Section 4.1 for a high-level description of its formulation, rationale, and the loss function). To manage the class imbalance problem, we use the focal loss. This is applicable to both the selection and the prediction models.

A key innovation in our work is to allow the selection model f and the prediction model g to be multimodal transformers so that we can incorporate predictions that encompass both structured and textual features. In their empirical testing, Kahng et al. (2023) reported that while it is possible to use different types of models (e.g., tree or deep neural networks) for selection and prediction, using the same type of model seems to work the best. Additionally, Kahng et al. (2023) showed that, while using a one-step procedure to optimize the loss function would be ideal, a two-step procedure is more stable and produces an equally good result. Therefore, we adopt a two-step optimization approach.

Given samples of $(\mathbf{x}_i, y_i, s_i), \forall i = 1, 2, \ldots, N$, where $\mathbf{x}_i$ is the feature vector, y_i is the outcome label (whether the loan defaults in our case), and s_i is the selection status (whether the listing is funded in our case) of sample i, along with some hyperparameters (such as the batch size and learning rate), the pseudocode is provided in Algorithm 1.

Algorithm 1: Two-Step Optimization for DeepHeckman

Data: Data $\mathcal{D} = \{(\mathbf{x}_i, y_i, s_i) \in X \times Y \times S, i = 1, \ldots, N\}$, Batch size b, Learning rate γ.

Result: $\hat{f}$, $\hat{g}$, $\hat{\Sigma}$

Randomly initialize $\hat{f}$, $\hat{g}$, $\hat{\Sigma}$;

`/* Step 1: Learn the Selection Model` $\hat{g}$ `*/`

while $\hat{g}$ *not converged* **do**

  $\mathcal{B} \leftarrow BatchSampler(\mathcal{D}, b)$;

  $\hat{g} \leftarrow \hat{g} + \frac{\gamma}{b} \nabla \Sigma_{(\mathbf{x}_i, s_i) \in \mathcal{B}} \left[s_i \log \Phi(\hat{g}(\mathbf{x}_i)) + (1 - s_i) \log \Phi(-\hat{g}(\mathbf{x}_i)) \right]$,

end

`/* Step 2: Learn the Outcome Model` $(\hat{f}, \hat{\Sigma})$ `*/`

while $\hat{f}$ *not converged* **do**

  $\mathcal{B} \leftarrow BatchSampler(\mathcal{D}, b)$;

  $\hat{f} \leftarrow \hat{f} - \frac{\gamma}{b} \nabla \Sigma_{(\mathbf{x}_i, y_i, s_i) \in \mathcal{B}} \left[s_i \Lambda \left(\hat{f}(\mathbf{x}_i), \hat{g}(\mathbf{x}_i); y_i, s_i; \hat{\Sigma} \right) \right]$,

  $\hat{\Sigma} \leftarrow \hat{\Sigma} - \frac{\gamma}{b} \nabla \Sigma_{(\mathbf{x}_i, y_i, s_i) \in \mathcal{B}} \left[s_i \Lambda \left(\hat{f}(\mathbf{x}_i), \hat{g}(\mathbf{x}_i); y_i, s_i; \hat{\Sigma} \right) \right]$,

end

Owing to random initialization, when conducting our context-specific experiments (e.g., misallocation and equilibrium), we fit the same model setting 15 times and report their average performance based on the test set.

Appendix F Modeling Details

F.1 ROI Calculation

We adopt the method described in Netzer et al. (2019) to calculate the ROI, which incorporates interest rates into its calculation, making it possible to account for the associated risk. Specifically, we first compute the expected profit of each loan as

$$E(profit)_j = [1 - P_j] \times IE_j + [P_j] \times [-0.75 \times AG_j + 0.25 \times IE_j],$$

where $E(profit)_j$ is the expected profit of loan j, P_j is loan j's probability of default based on the corresponding predictive model, IE_j is the interest rate on the loan's term, AG_j is loan j's principal amount, which we assume $\$100$ for each loan.

Following Netzer et al. (2019), we also assume that when a loan becomes default, an average of 25% of the principal is repaid, based on data published by crowdfunding consulting agencies. We then rank the loans based on their expected profits and select the top 1,000 loans with the highest expected profits to create an investment portfolio. Assuming that \$100 is invested in these 1,000 loans, we calculate the actual profit (or loss) of each loan by using its actual default status.[32]

Assuming a total investment of \$100,000 (1,000 loans at \$100 per loan), the ROI is first computed as the ratio between the net profit (actual total profit minus the total investment) and the total investment, and then converted to an annual ROI based on the loan's term.

F.2 Lender Models

In this study, we trained several different versions of the lender model to investigate different lender types, market scenarios, and adoption rates. For clarity, this section summarizes the models

[32] Note that we do not know the actual default status if the loan was historically not funded. Therefore, if a loan was not actually funded in the historical data, we inferred its repayment status using a gold-standard deep Heckman model, and the historical repayment status if the loan was selected for funding. The gold-standard deep Heckman model was fit using all samples (i.e., without train-test split) because the goal was to maximize the data-model fit instead of out-of-sample generalization. The parameters were selected using cross-validation, and we adopted the optimal hyperparameters to retrain the model using all data as the final model. Since converting all 231,144 text descriptions into BERT embeddings is extremely memory and computing intensive, we randomly down-sampled the full sample at a 20% ratio to make the analysis feasible. Such down sampling completely at random ensures that the subsample we used for training is representative.

we trained. To make the evaluations truthful and fair, the first step was to randomly partition the data (n=231,144) by an 8:2 ratio into a training set (n=184,915) and a testing set (n=46,229). While the training set may be further partitioned for hyperparameter optimization, we kept the test set completely separate from the training process and only use it for evaluation.

Under each setting (for a version of the lender model, as detailed below), parameters were selected using cross-validation, and the optimal hyperparameters were used to retrain the model using all training samples to obtain the final model. Since the training sample size remains quite large, and converting all 184,915 text descriptions into BERT embeddings is still extremely memory and computing intensive. Therefore, we randomly downsampled the full sample at a 20% ratio to make the analysis feasible. Such down sampling completely at random ensures that the subsample we used for training is representative of the full sample.

F.2.1 Lender Constrainedness and Market Scenarios

Section 5 investigated the robustness of ROI and misallocation results when the market consists of a mix of constrained and intelligent lenders. The constrained lender model M_{cl} was to imitate lenders who may not necessarily have the ability or resources to learn and adapt their investment strategies. Therefore, we may consider them likely to "repeat the past." To implement this type of lenders, we train a set of constrained lender models using historical funding decisions as the target. In contrast, the intelligent lender model M_{cl} was trained using a two-step model via Heckman multimodal. The first-stage outcome was whether it was funded, and the second-stage outcome was whether it was repaid. This has been verified as the best model in prediction, and thus, could represent the behavior of intelligent lenders.

Table F11: Lender Models for Different Constrainedness and Market Scenarios

Lender Type	Market Scenario	Target Y	Predictors for Training X	Predictors for Testing X'
Constrained Lender M_{cl}	No GenAI	is_funded	original text	original text
	Small-Scale Adoption	is_funded	original text	rewritten text
	Large-Scale Adoption	is_funded	rewritten text	rewritten text
Intelligent Lender M_{il}	No GenAI	is_repaid[a]	original text	original text
	Small-Scale Adoption	is_repaid[a]	original text	rewritten text
	Large-Scale Adoption	is_repaid[a]	rewritten text	rewritten text

[a] Using a two-step model with Heckman correction.

To produce results in Figure 7, we trained both the constrained and the intelligent lender models under each market scenario. Given two types of lenders and three market scenarios, we trained six (i.e., 2×3) lender model versions, as summarized in Table F11. Each subplot in Figure 7 is produced by varying α and the weight of intelligent lenders in the market. Practically, we assume that α loans are allotted to intelligent lenders to make funding decisions and the rest are allotted to constrained lenders.

F.2.2 Different Market Adoption Rates

When investigating the market equilibrium in Section 6, we focus on the intelligent lender under varying level of market adoption. To make the computing tractable, we set different GenAI adoption rates at every 10% increments, and train additional versions of the lender model, as summarized in Table F12.

Table F12: Lender Models for Different Market Adoption Rates

Adoption Rate	Target Y[a]	Predictors for Training X and Testing X'
0%	is_repaid	all original text
10%	is_repaid	90% original text, 10% rewritten text
20%	is_repaid	80% original text, 20% rewritten text
30%	is_repaid	70% original text, 30% rewritten text
40%	is_repaid	60% original text, 40% rewritten text
50%	is_repaid	50% original text, 50% rewritten text
60%	is_repaid	40% original text, 60% rewritten text
70%	is_repaid	30% original text, 70% rewritten text
80%	is_repaid	20% original text, 80% rewritten text
90%	is_repaid	10% original text, 90% rewritten text
100%	is_repaid	all rewritten text

[a] Using a two-step model using Heckman adjustments.

In particular, when the adoption rate is 0%, the lender model is equivalent to lenders in the "No GenAI" market scenario. The other extreme is when the adoption rate is 100%, in which the lender model is equivalent to those in the "Large-Scale Adoption" scenario. Everything in between shows a mix of borrowers in terms of adopting ChatGPT to rewrite the text component of their loan applications. The intelligent lender would always use the data they see (i.e., a mix of loan applications with original and rewritten text in the training set) to train a two-step model with Heckman correction and then assess their decisions on the test set, which was also randomly selected for rewriting at the same market adoption rates. Each data point in Figure 8 corresponds

to the change in funding probability (after minus before text rewriting; y-axis) of loans under each level of GenAI adoption (x-axis), aggregated at each credit level.

Appendix G Traditional Machine-Learning Classifiers

In this section, we document the performance of several strong machine learning models that are not based on deep neural networks. Following Netzer et al. (2019), we consider regularized logistic regressions (GLM), random forest (RF), extremely randomized trees (ERT), and an ensemble model consisting of these base models. For stronger benchmarking, we also included a light gradient-boosting machine (LGBM) and eXtreme Gradient Boosting (XGB) as the additional base models. The ensemble model was constructed by stacking the aforementioned models. Specifically, the prediction of the ensemble model is a linear combination of the base models.

All these models were implemented using sklearn.[33] To handle class imbalance, we use class weights proportional to inverted sample size (class_weight="balanced"). The other hyperparameters are listed in Table G13. The final estimator was selected through cross-validation. We employed a Python implementation of global optimization with Gaussian processes for the optimization of hyperparameters (Nogueira, 2014).

Table G13: Hyperparameter Settings for Classifiers

Model	Parameter	Description	Range
GLM	C	Inverse of regularization strength: must be a positive float. Smaller values indicate stronger regularization.	(0.01, 10)
	penalty	Specify the norm of the penalty.	['l1', 'l2']
	solver	Algorithm to use in the optimization problem.	[0, 1]
RF, ERT	n_estimators	Number of trees in forest.	(10, 1000)
	max_depth	Maximum tree depth	(5, 30)
	max_features	Number of features to consider when looking for the best split.	(0.1, 0.999)
	min_samples_leaf	Minimum number of samples required to be at a leaf node.	(1, 25)
	min_samples_split	The minimum number of samples required to split an internal node.	(2, 25)
LGBM, XGB	n_estimators	Number of trees,	(10, 1000)
	max_depth	Maximum tree depth	(5, 30)
	learning_rate	The rate of learning of the model.	(0.01, 0.3)
	min_child_samples	Minimum number of data needed in a leaf.	(5, 50)
	num_leaves	Number of leaves per tree.	(24, 45)

To account for the two-step nature of lending, we incorporate the Inverse Mills Ratio (IMR) to account for selection bias in each of these models. For simplicity, and similar to our multimodal

[33] https://scikit-learn.org/stable/index.html

Heckman procedure, we fit the two models (selection model f and outcome model g) separately using the following logic:

1. Train a selection model f using all training samples to predict funding (1=funded, 0=not funded).

2. Compute the sample-level IMR as true label minus model predicted probability of funded.

3. Add this IMR to the second-stage model as an additional input. Train this outcome model g using only loans that were funded.

The combined model is applied to the test set to report performance.

Performance of ML Models. The results are presented in Table G14. Model 1 is based solely on structured features, Model 2 includes both structured features and text as predictors. When deriving features from text to fit the machine learning classifiers, we consider both textual features (as described in Section 2.3) and bigram features in a setting similar to Netzer et al. (2019). As formulated in Eqn. (4), soft information was quantified by the performance improvement from Model 1 to Model 2. Every model reports a positive value for soft information.

Table G14: Performance Benchmarking by Predicting Default Among Funded Loans

	AUC			ROI		
	Model 1 (Structured)	Model 2 (Structured + Text)	Soft Information	Model 1 (Structured)	Model 2 (Structured + Text)	Soft Information
GLM Only	72.7%	72.8%	0.1%	13.1%	15.0%	1.9%
RF Only	73.4%	73.7%	0.3%	13.1%	14.1%	1.0%
ERT Only	72.5%	73.2%	0.7%	12.0%	13.4%	1.4%
LGBM Only	72.5%	72.9%	0.4%	7.6%	12.4%	4.8%
XGB Only	73.7%	74.1%	0.2%	**14.4%**	**15.9%**	1.5%
Ensemble	**74.1%**	**74.5%**	0.4%	13.8%	15.7%	1.9%

Note: Each set of model results (AUC or ROI) are averaged across 10 replications of the ten-fold mean.

Unsurprisingly, the ensemble model yielded the best AUC among all the models considered, regardless of the set of features used. In particular, when using structured features only, the ensemble model reached an AUC of 74.1% (corresponding to an ROI of 13.8%), and when using both structured and textual features, the ensemble model reached an AUC of 74.5% (corresponding to

an ROI of 15.7%). It is also worth noting that our proprietary multimodal model outperformed all the ML models, indicating a strong performance gain.