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

Lin William Cong Yanhong Guo Xin Zhao Wenjun Zhou

Abstract

Writing matters in finance for both informational and non-informational reasons. Applying ChatGPT, a generative AI (GenAI) tool, to data from a dominant online credit platform, we study how the various dimensions of writing contribute to lending outcomes and are reshaped by GenAI. We also introduce a composite Writing Quality Index (WQI) that is easy to compute and effectively predicts. While GenAI significantly enhances the writing and perceived quality of loan applications, the effects are heterogeneous, with high-credit-score borrowers benefiting more from ChatGPT rewrites because they treat good credit and quality writing as substitutes in their applications whereas these are empirically complements for receiving funding. Furthermore, we build proprietary language models and introduce a novel Heckman Correction Procedure for deep learning, to demonstrate that GenAI decreases soft information conveyed, likely due to the convergence in writing. When lenders adjust lending decisions considering borrowers' GenAI adoption, they recapture some soft information, mitigating potential credit misallocation. We characterize the hypothetical long-run general equilibrium in which GenAI adoption and responses in credit decision-making are fully endogenous. The findings provide insights into the evolving role of soft information and potential impacts of GenAI in lending.

Keywords: AI, BERT/RoBERTa, ChatGPT, Heckman Correction, LLMs

*Cong is at Cornell University SC Johnson College of Business (Johnson) 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 Simon Xu, Allen Hu, and conference and seminar participants at 2024 Generative AI for Finance Conference (Montréal) for constructive comments. Send correspondence to Cong at will.cong@cornell.edu.

1 Introduction

It pays to write well (greater readability, persuasiveness, clarity, etc.) in corporate disclosure, loan applications, and other financial settings (Iyer et al., 2016; Hwang and Kim, 2017; Gao et al., 2023). However, a systematic and joint consideration of the various dimensions of "writing well" is missing from previous studies, so is the link between writing well and the soft information the writing conveys. Meanwhile, the recent breakthroughs in generative AI (GenAI), especially large language models (LLMs), are believed to dramatically change how people generate textual content. We are yet to understand how GenAI affects writing quality in finance, GenAI adoption and decision-making in response to the adoption, and what all these imply regarding the transmission and usage of soft information in financial intermediation.[1]

To these ends, we us (i) data from Prosper, the first marketplace lending platform in the United States, and still the most prominent one, (ii) ChatGPT, an interface for engage with LLMs, and (iii) proprietary multi-model large models to capture lender's decision-making process, to understand not only how the various dimensions of writing relate to credit outcomes but also how writing quality and GenAI affect lending decisions, allocative efficiency of credits, and soft information conveyed. In so doing, we also, as methodology contributions, introduce a Heckman-like correction for deep-learning models in social sciences, and an algorithm for implementing a generative general equilibrium.

Similar to existing studies (Netzer et al., 2019; Gao et al., 2023), we use rich loan-level lending and borrower performance information from the Prosper platform covering more than a year. This observation window is also prior to the introduction of any GenAI tools, allowing us to isolate human produced and AI produced writing. We rewrite loan descriptions using ChatGPT. Following OpenAI's tutorial, we design a prompt consisting of the role, task, and rule settings. The 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 is informed by existing research regarding what helps fundability. More analyses are conducted to assess the robustness of findings when rerunning the same prompt multiple times and when

[1] 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.

employing additional variations in prompting.

To analyze the impacts of GenAI on writing quality, we first examine changes in text features. In particular, we find that the rewrites are more detailed, formal, and contain fewer grammatical and spelling errors. Employing a human-based evaluation, we confirm that GenAI significantly improves the perceived writing quality and willingness to fund. The difference between the original description and the rewrite represents GenAI's impact on the credit market at the loan level. Using GenAI rewrites as "optimal" benchmarks, we propose a holistic measure, the Writing Quality Index (WQI), using cosine similarity between the text embeddings of the original and the rewritten text. We find that low-score borrowers exhibit a higher WQI than high-score borrowers. WQI also shows predictive power to loan outcomes, especially fundability. Furthermore, the WQI's effect towards fundability is heterogeneous at the individual level: high-score borrowers could benefit more from increasing WQI.

While GenAI may historically have helped borrowers improve their writing and, in turn, increased their chances of securing funding, it may also have reduced the soft information that lenders rely on to make lending decisions. For lenders, text data provide a vital supplement to traditional credit data by conveying this soft information.[2] To quantify soft information in text, we assess lenders' decision quality when using both tabular features and text versus tabular features alone. Decision quality is measured by classification accuracy[3] or return on investment (ROI). We employ advanced language models, including Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Pretraining Approach (RoBERTa), to represent the lenders' decision-making. To integrate text embeddings and tabular features, we use a multimodal transformer to develop a unified model. Our results show a significant decline in the predictive value of texts. Using cosine similarity to measure within-group homogeneity, we find that loan descriptions of low-score borrowers are more similar to each other, while the those of high-score borrowers are less so. After GenAI rewriting, homogeneity increased in both groups. In other words, GenAI reduces variability in writing, which may explain the diminished soft information.

[2] 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 defaults. Similarly, Netzer et al. (2019) report that peer-to-peer lending borrowers' text descriptions offer insights into past behavior and future intentions, enhancing loan outcome predictions.

[3] Following Netzer et al. (2019), we use the area under the receiver-operating characteristic curve (AUC) as a measure of decision model accuracy.

This convergence makes it harder to discern true loan quality compared to traditional methods.

Nonetheless, the above analyses are under a partial equilibrium setting without accounting for the fact that lenders can adapt their decisions in a marketplace anticipating and observing GenAI usage. We are thus interested in expanding the scope of our research by considering lender response. Specifically, in a large-scale adoption scenario, lenders would update their decision model by training it with newer data, which include descriptions produced by GenAI for some loans. In this counterfactual sense, we wish to assess the extent of misallocation, requiring knowledge of the repayment status for unfunded loans. To mitigate the potential sample selection bias arising from solely relying on funded data where loan quality is observed, we devise a loan repayment prediction model following the principles of Heckman correction (Heckman, 1979) for our multimodal transformer modal. We find that, compared with the pre-GenAI scenario, if borrowers have adopted GenAI while lenders have not updated their decision model, there is a significant increases in misallocation, as measured by the false positive rate (FPR) and the false negative rate (FNR). In contrast, if lenders update their decision model, the misallocation rates slightly improve, suggesting partial soft information recovery.

Borrowers' GenAI adoption and lenders' response will reach an equilibrium. Assuming that the cost of model update is negligible, rational lenders would always benefit from updating their model with newer data that reflect the market status (i.e., GenAI adoption). Searching over equilibria with various levels of GenAI adoption by borrowers and the corresponding responses by lenders, we find that using GenAI consistently improves individual borrower's fundability compared with not using GenAI, especially for high-score borrowers and when the market adoption is around 50%. This pattern implies a unique general equilibrium in which all borrowers would adopt GenAI. This counterfactual analysis thus informs us how GenAI plausibly affects online lending in the long run.

Our work contributes to the literature in several key directions. First, we are the earliest study to verify the effectiveness of GenAI in improving the writing quality and fundability of loan applications. We find that although GenAI can help enhance fundability, it concurrently reduces soft information available to help lender decisions. Second, we propose the WQI and test its predictive power on fundability. WQI not only provides easy computation and interpretability in summarizing the extent of change of each loan's description by GenAI, but also allows us an effective measure of writing quality in human-generated texts. Thirdly, we explore the response of lenders

and the market equilibrium implications. We find that all borrowers, irrespective of their credit levels or assumed market adoption rates, receive a positive utility from GenAI adoption. Rational borrowers should adopt GenAI as long as the costs of using GenAI is low. To the best of our knowledge, ours is the first study to investigate the impact of GenAI adoption on the online credit market in a counterfactual long-run general equilibrium, offering crucial insights for online credit market participation in the GenAI era. Last but not least, to simultaneously model tabular and text features, we leverage the multimodal transformer to represent lenders' decision to assessing loan quality. In the process, We extend Heckman correction to account for selection biases in deep learning models, which is a methodological contribution itself.

2 Institutional Background, Data, and Prompt Design

2.1 Online Credit Data from Prosper Marketplace

Prosper Marketplace, established in 2005 and began operations in 2006, was recognized as the first P2P lending platform in the U.S.[4] On the Prosper platform, despite different loan origination dates, the lending process is consistent across all loans and follow several stages, , as illustrated in Figure 1. In the pre-funding stage, borrowers submit loan requests with structured and unstructured information, specify the amount and the maximum interest rate they are willing to accept, and open the listings for lenders to bid. Then in the bidding stage (Stage 1) which typically lasts for two weeks for each listing, lenders specify the amount (usually smaller than the requested amount) and the minimum interest rate they are willing to lend, all within a fixed bidding period (e.g., two weeks). Once the total requested amount is fulfilled within the bidding period, the listing becomes a loan; otherwise, the loan application is considered rejected. Stage 2 starts when a loan is funded and lasts to its 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" when it is marked as "Paid in full," "Settled in full," or simply "Paid." Since the text component of their loan requests was written in Stage 1, it contains information potentially useful in predicting both the likelihood of funding and the likelihood of default.

[4] https://www.prosper.com/about

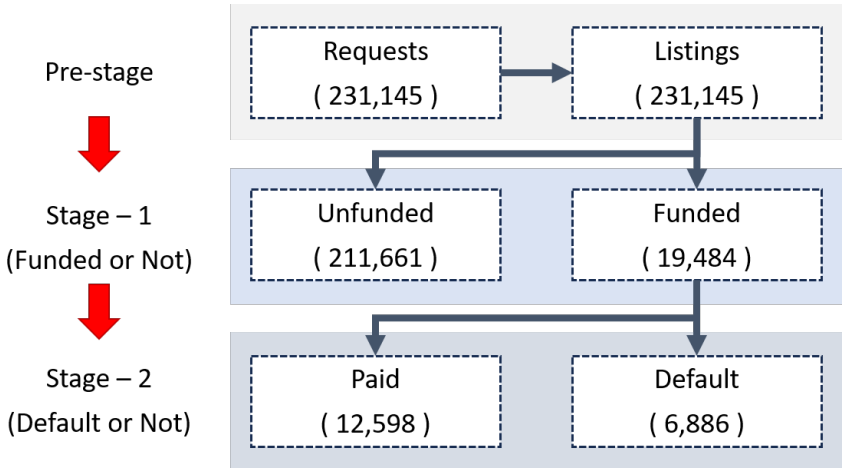


Figure 1: Loan Service Procedures.

We downloaded all loan requests posted between April 2007 and October 2008, with a total of 231,145 listings. We focus on this early episode of Prosper for multiple reasons: (i) Data during this sample period is publicly available, and richer and more diverse than later episodes(Netzer et al., 2019) before regulations tightened. (ii) The period predate the widespread use of language models, so that we can isolate the effect of how LLMs improve writing. Among 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 other 12,598 (64.7%) were repaid. For rejected listings, we extracted an equal number of samples based on the credit distribution of funded loans, and refer to them as unfunded loans. Consequently, in our subsequently analysis, the total dataset includes 38968 observations, comprising 19,484 funded loans and 19484 unfunded loans.[5]

2.2 Tabular Features

Borrowers' financial and demographic information is available to lenders mostly as structured data (i.e., hard or verifiable information Caldieraro et al., 2018). These variables include the loan request information, the borrower's credit profile, demographic information, and an overall credit rating by Prosper. Borrowers' credit profile include their debt-to-income ratios (*Debt-To-Income Ratio*), the number of delinquencies (*Current Delinquencies*), and the number of inquires in the

[5] We do not take the full unfunded sample because rewriting with ChatGPT is costly.

past six months (*Recent Inquires*). They are commonly used in traditional risk models (Guo et al., 2016; Fitzpatrick and Mues, 2021). Their demographics information include a categorical variable indicating the current employment status (*Employment Status*), the length of time (in years) in the current employment status (*Status Time*), and a dummy variable indicating whether the borrower is a homeowner (*Is Homeowner*).

Loan request information include the amount requested in U.S. dollars (*Amount Requested*) and the lenders' interest rate (*Lender Interest Rate*). Since the data were 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. Similar to existing studies, we include a dummy variable of whether the user has a head shot photo in their user profile, since previous studies reported the use of images on various platforms (Li and Xie, 2020; Ho et al., 2022). In addition, the credit grade of the borrower is an ordinal variable with seven possible levels, ranging from AA (best credit) to HR (high risk). Appendix A contains more details of these tabular features, including definitions, descriptive statistics, and cross-group comparison.

2.3 Textual Features

Writing quality is a well-studied topic in fields such as linguistics, education, and cognitive psychology. Although general writing quality can be assessed in many different dimensions, recent work in finance mostly focuses on readability. Readability refers to the ease with which a reader can understand the text, often influenced by factors like grammatical and syntactic accuracy, word choice, sentence structure, and overall coherence.Informed by Gao et al. (2023); Netzer et al. (2019), we carefully examine six widely recognized dimensions of quality writing, in addition to using ChatGPT to construct a novel composite metric. We describe the six linguistic dimensions next.

Text length. Text length should not be too short or too long. It should convey the right amount of information with concise and coherent language. Longer text not only reduces readability, but also suggests deception. Existing research does not find that longer or shorter texts being more likely linked to higher quality loans either (Toma and Hancock, 2012; Hancock et al., 2007). Given that the free text box has a limit of up to 4,000 characters in application text length (Ryan et al., 2007),

borrowers are prevented from writing too long texts but may write texts that are too short. Similar to Netzer et al. (2019), we measure text length as the number of characters in the text box.

Basic linguistic conformance. For better readability, the text must not only be error-free but also conforms to expected linguistic norms and conventions, especially proper grammar, spelling, and punctuation. Few spelling errors directly enhance text readability by contributing to its clarity and coherence (Ghose and Ipeirotis, 2010; Loughran and McDonald, 2014). Harkness (2016) reported that in traditional lending channels, spelling errors, often indicative of traits associated with lower income, are associated with a lower probability of loan approval. This is because such mistakes often act as indicators of traits associated with lower income (Dorfleitner et al., 2016). Netzer et al. (2019) used this characteristic to qualify the differences in loan application texts within the P2P market. They implemented it as the count of spelling mistakes by the Enchant spellchecker. Gao et al. (2023) considers the ratio of the number of grammatical errors to the number of words in the loan request, standardized to zero mean, unit variance. We measure (the lack of) conformance as the percentage of spelling errors and the percentage of grammatical errors in loan descriptions.

Formality. A rich, accurate vocabulary contributes to writing quality by enabling the writer to express ideas more precisely and vividly. According to Tausczik and Pennebaker (2010), this metric is typically employed to assess complex language, educational level, and social status. 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 complex language can also be risky if readers perceive it to be artificially or frivolously complex, suggesting that higher complexity language was likely used deceptively (Oppenheimer, 2006). Similarly to Netzer et al. (2019), we measure the sophistication of the vocabulary by the percentage of words with six or more letters.

Readability. Several readability metrics have been used in the finance literature. For instance, Kim et al. (2019) and Cao et al. (2024) use Gunning Fog Readability Index (FOG) in corporate filing. Gao et al. (2023) also measured readability with lexical complexity and proposed a ReAb readability score, which combines standardized spelling error rates, grammatical error rates, and standardized FOG. Considering that a key factor whether an advanced vocabulary is used at the

appropriate level is to see if it matches the audience's level, a measure typically referred to as the "simple measure of gobbledygook" (SMOG) (McLaughlin, 1969). The readability evaluates writing quality based on the years of education required for a reader to understand the text at first glance. Similarly to Netzer et al. (2019), we consider the readability index.

Tone. Another factor of writing quality is measured by engagement and style. Engaging writing keeps the reader interested, often by varying sentence structure, tone, and vocabulary. Existing research often refers to the emotional content of text as tone or sentiment (Loughran and McDonald, 2011). Investors often respond favorably to 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). In online credit markets, Gao et al. (2023) similarity discovered that loan descriptions with a positive sentiment were more likely to secure funding. Based on these research, we measure the tone of loan descriptions using sentiment scores provided by a pre-trained sentiment-BERT model.

Table 1: Textual Features for Writing Measurement

Nums	Feature	Definition
1.	Text Length	The number of words in loan descriptions.
2.	Spelling Error	The percentage of spelling errors within loan descriptions.
3.	Grammatical Error	The percentage of grammatical errors within loan descriptions.
4.	Formality	The percentage of words with six or more letters.
5.	Readability	The "simple measure of gobbledygook" (SMOG) readability index.
6.	Tone	The sentiment of the loan descriptions.

To summarize, we selected textual features across six dimensions (see Table 1) to assess the writing of loan applications. These features were computed using various Python packages. Specifically, for text length and formality, we used fundamental string operations in Python; for spelling error, we used the "spellchecker" package; for grammatical error, we utilized the "language_tool_python" package; for readability, we applied the "textstat" package; and for tone, we employed a pre-trained sentiment-BERT model from HuggingFace.[6]

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

2.4 GPT Prompt Design

OpenAI guidelines prescribe that an effective prompt should be specific and can include role setting, task setting, and rule setting. Role settings enable ChatGPT to comprehend the context, task setting defines the task for ChatGPT, and rule setting clarifies the rules it should follow.

A borrower using LLMs to write loan applications would want to get helpful from a professional. We thus describe its 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 action needed. 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 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. 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.*

Guided by this prompt template, the loan descriptions have been rewritten. For example, for the original text in Figure 2, the re-written version might look somewhat like the following: "*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 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 the rewrite, with the rewritten ones better in writing quality and credibility. Table A6 offers more examples.

3 Textual Features, Writing Quality, and Online Lending

We first examine the text features of the loan description, their linkage to loan fundability, and how LLMs changes the text features for better writing quality. We then propose a writing quality index

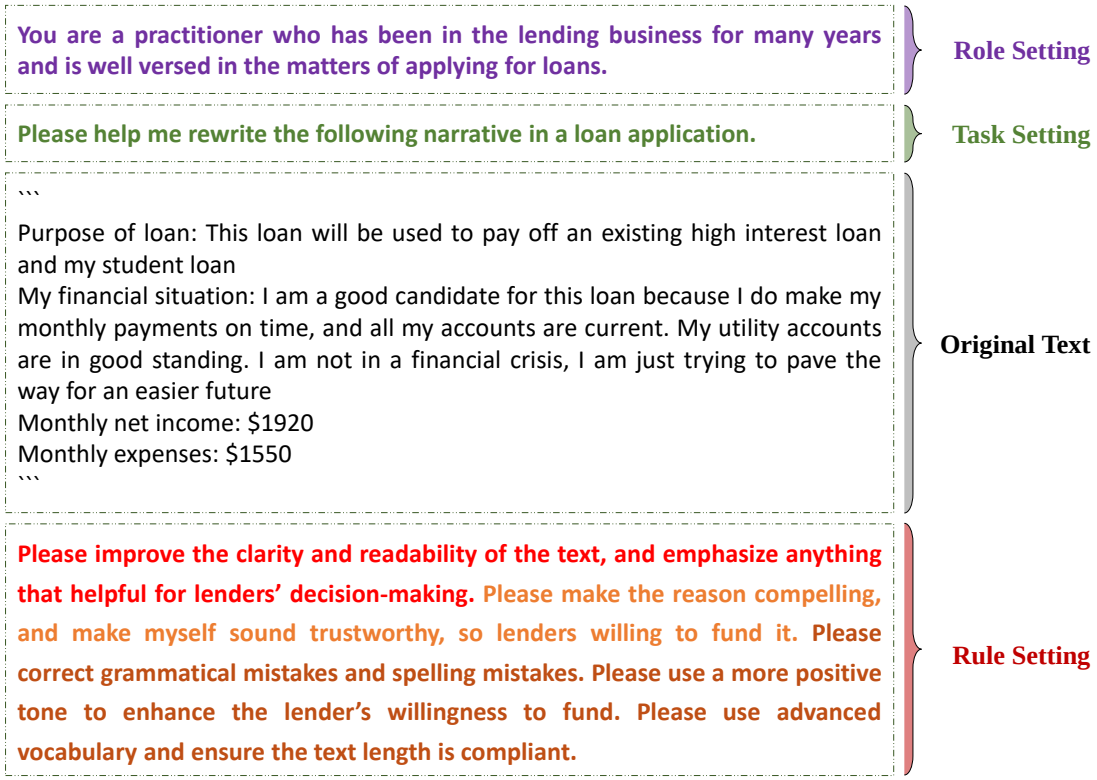


Figure 2: An Example ChatGPT Prompt.

(WQI) as a comprehensive measure of writing quality and the potential improvement LLMs can bring, before demonstrating its relationship with credit level and funding status.

3.1 Text Features and Changes by LLMs

Text features potentially reflect writing quality and exhibit a strong connection with fundability. To demonstrate this link, we fit a regression model to explain fundability using the text features, while interacting each text feature with the credit level.[7] The regression table can be found in Table A5, and Figure 3 visualizes the relationship between each text feature and fundability. We observe a positive relationship between fundability and text length, readability, and tone, alongside a negative relationship with spelling errors, grammatical errors, and formality. These correlations closely related to writing quality, as high-quality writing typically features appropriate text length, fewer errors, a positive tone, and higher readability. The effect is stronger for certain groups. For example, increasing percentage of grammatical errors and increasing level of formality reduces the loan fundability more for low-score borrowers. Similarly, while increasing readability scores increases fundability for each group, fundability of the high-score group increases more for each unit of increase in readability.

Employing ChatGPT and prompt templates, we analyzed the impact of GenAI adoption on writing from the perspective of text features. Table 2 illustrates the changes in textual features of loan descriptions before and after GenAI rewriting.

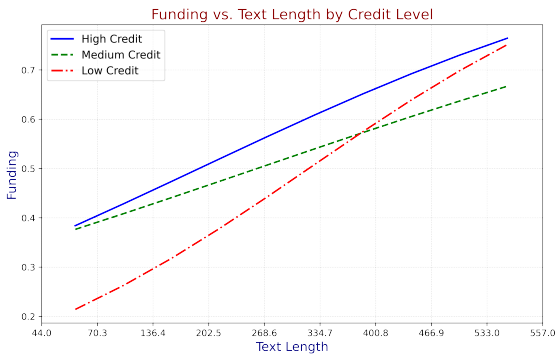
Table 2: Textual Features Before and After Rewrite

	Original		Rewrite		Change	
	Mean	(SD)	Mean	(SD)	DIF[a]	(%)
Text Length	202.5	(132.2)	281.1	(99.44)	78.6***	(+38.8%)
Spelling Error	0.035	(0.021)	0.009	(0.007)	-0.026***	(-74.3%)
Grammatical Error	0.064	(0.041)	0.028	(0.023)	-0.036***	(-56.3%)
Formality	0.239	(0.058)	0.367	(0.035)	0.128***	(+53.6%)
Readability	10.45	(3.775)	15.75	(1.684)	5.30***	(+50.7%)
Tone	0.745	(0.157)	0.754	(0.144)	0.009***	(+ 1.2%)

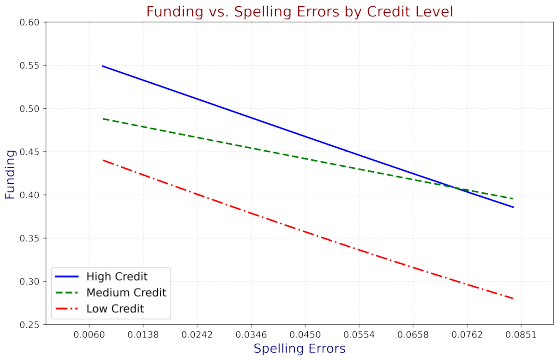
[a] Sig. levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 2 demonstrates that GenAI adoption significantly improved the text features, making the

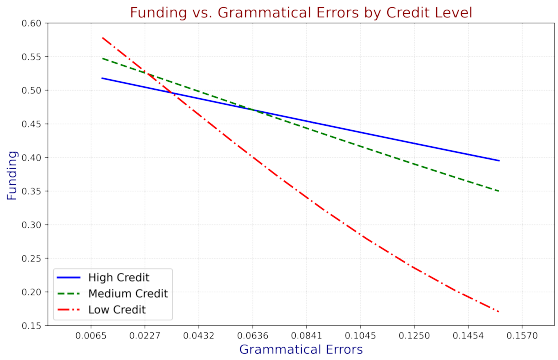
[7] Following Netzer et al. (2019), we classify credit grades into three levels: Low, Medium, and High. Low = low credit grades: E, HR; Medium = medium credit grades: B, C, D; High = high credit grades: AA, A.



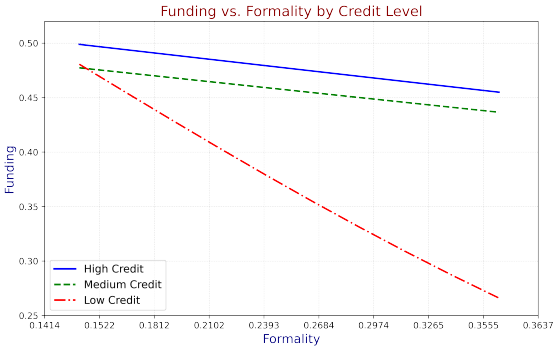
(a) Text Length



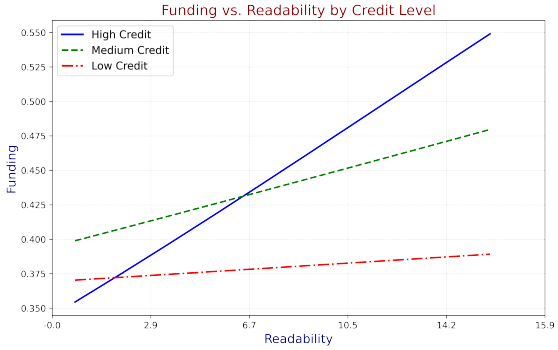
(b) Spelling Errors



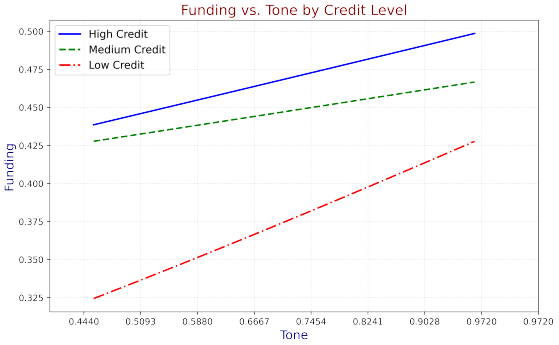
(c) Grammatical Errors



(d) Formality



(e) Readability



(f) Tone

Figure 3: Text Features and Fundability.

rewritten texts longer, with fewer errors, and with a more positive tone. Specifically, on average, following GenAI rewriting, the number of words in loan descriptions increased by 38.8%, the percentage of spelling and grammatical errors decreased by 74.3% and 56.3%, respectively; and tone increased by 1.2%. In general, these changes improve writing quality by aligning loan applications more closely with those that are successfully funded.

In addition, formality increased by 53.6% and readability increased by 50.7%, on average. This indicates that the GenAI rewrite version tends to use a higher proportion of complex words (i.e., those with 6 characters or more), and it may require a higher level of education to read and easily understand the text. As depicted in Figure 3, while increased readability seems helpful for high-score borrowers, increased formality could decreases the fundability of low-score borrowers.

3.2 GenAI Measurement of Writing Quality

The change in each text feature reflects the influence of GenAI rewriting on the text in a specific dimension as discussing in the literature on writing quality. We also introduce a holistic metric, the Writing Quality Index (WQI), to capture the changes before and after GenAI rewriting. Using state-of-the-art deep learning techniques, we can convert any text into a numeric vector, known as an embedding, and use the cosine coefficient to measure the semantic similarity between them. This cosine similarity can provide a method for measuring change in a holistic way. Meanwhile, this metric can also help evaluate the extent of convergence.

Measuring semantic similarity. Given two numeric vectors $\mathbf{x}_1 = (x_{11}, x_{12}, \ldots, x_{1n})$ and $\mathbf{x}_2 = (x_{21}, x_{22}, \ldots, x_{2n})$, the cosine similarity is calculated as:

$$CosineSimilarity(\mathbf{x}_1, \mathbf{x}_2) = \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|} \tag{1}$$

where $\mathbf{x}_1 \cdot \mathbf{x}_2 = \sum_{j=1}^{n} x_{1j}x_{2j}$ is the dot product of the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, and $\|\mathbf{x}_i\| = \sqrt{\sum_{j=1}^{n} x_{ij}^2}$ is the norm of vector $\mathbf{x}_i$. We employ the Siamese-BERT (SBERT) model to encode loan descriptions, which is more efficient and effective than BERT to assess the similarity between pairs of sentences (Reimers and Gurevych, 2019). We then calculate the cosine similarity between these encoded descriptions to measure their content similarity.

Writing Quality Index (WQI). We define WQI as:

$$WQI(\mathbf{x}) = CosineSimilariy(emb(\mathbf{x}), emb(\mathbf{x}')), \tag{2}$$

where $\mathbf{x}$ represents the original text, $\mathbf{x}'$ represents its rewrite, and $emb()$ represents the conversion from raw text to its numeric embedding. The assumption is that by default, GPT writes well and the prompt produces an improved version, which is supported by the literature (Herbold et al., 2023; Deveci et al., 2023) and our earlier analyses. We also validate this assumption in the context of lending using an online experiment involving human raters (see Appendix D).

WQI measures the similarity of loan descriptions before and after GenAI writes. When we consider the GenAI-rewritten version as the optimal version, it assesses the writing quality of the original text by measuring how close it is to the optimal from a holistic view. Higher values represent higher writing quality, as the original version is more similar to the rewrite version, and lower values indicate lower writing quality.

3.3 Individual Heterogeneity in Writing Quality

How is WQI related to credit level and funding status? We conduct a two-way ANOVA test to check the group differences. We find that credit levels, funding outcome, and their interactions are significantly associated with WQI. Mean WQIs by group are summarized in Figure 4.

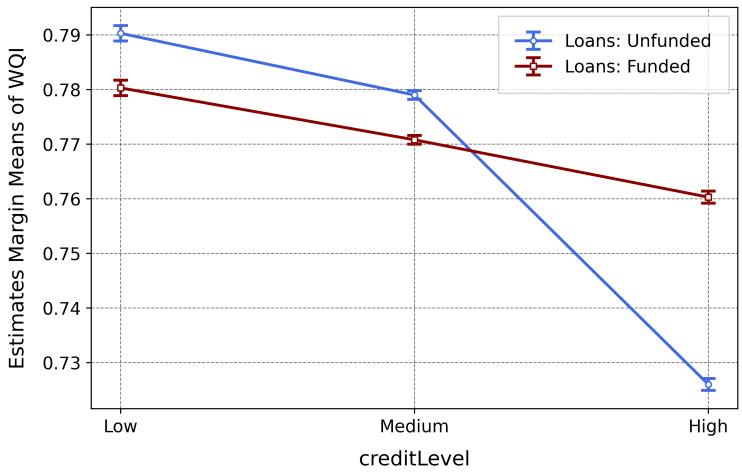


Figure 4: WQI by Credit Level and Funding Outcome.

From Figure 4 we can see that low-score borrowers produced significantly higher WQI, which indicates that low-score borrowers' loan descriptions exhibit better writing quality. Take the funded loans for instance, the WQI for low-, medium-, and high-score borrowers are expected to be 0.780, 0.7710, 0.760, respectively. Among the unfunded loans, the WQI shows the same ordering, with a mean WQI of 0.790, 0.779, 0.726 for low-, medium-, and high-score borrowers, respectively. Low-score borrowers, due to their poor credit profiles (i.e., hard information), often face difficulties when applying for conventional loans. When they apply for loans in the online credit markets, they have the opportunity of disclosing helpful soft information via the textual explanations. Unlike credit scores, loan descriptions are entirely crafted by the borrowers and serve as a communication tool with lenders. Consequently, these low-score borrowers typically focus more on good writing to enhance their chances of obtaining funding. In contrast, high-score borrowers are less concerned about explanations given their good credit profiles, and thus provide writing of lower quality.

The interaction between credit levels and the funding status reveals another interesting phenomenon. Specifically, for low and medium credit borrowers, the WQI is slightly higher for unfunded loans (0.790 and 0.779) than funded loans (0.780 and 0.771), whereas this pattern is reversed for high-score borrowers—the WQI for unfunded loans (0.726) is much lower than funded loans (0.760). This indicates that good credit profiles did compensate for the lack of good writing to an extent. High-score borrowers need to write poorer to not get funded.

3.4 Predicting Fundability with WQI

To check the relationship between writing quality and getting funded, we employ a logistic regression model and select previously mentioned tabular features as control variables to conduct the significance test. The model may be specified as follows:

$$\begin{aligned}\text{logit}(Outcome_i) =& \beta_0 + \beta_1 \times IsHighCredit_i + \beta_2 \times IsLowCredit_i + \beta_3 \times WQI_i \\ &+ \beta_4 \times WQI_i \times IsHighCredit_i \\ &+ \beta_5 \times WQI_i \times IsLowCredit_i + \boldsymbol{\gamma} \cdot \boldsymbol{c}_i + \varepsilon_i, \end{aligned} \tag{3}$$

where $Outcome_i$ represents whether loan i is funded, $\boldsymbol{c}_i$ 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 WQI would be $\hat{\beta}_3$ for medium credit borrowers, $\hat{\beta}_3 + \hat{\beta}_4$ for high-score borrowers, and $\hat{\beta}_3 + \hat{\beta}_5$ for low-score borrowers.

Table 3: Regression Analysis of Writing Quality and Funding Outcome

	DV: Funding (Funded: 1, Unfunded: 0)			
	Model 1	Model 2	Model 3	Model 4
WQI	**0.916*****	-0.352	-0.289	**-1.789*****
	(0.146)	(0.194)	(0.154)	(0.203)
WQI × High-Score Borrower		**5.256*****		**5.709*****
		(0.350)		(0.354)
WQI × Low-Score Borrower		**-1.682*****		**-1.414*****
		(0.387)		(0.394)
Text Length			**0.363*****	**0.370*****
			(0.015)	(0.015)
SEP			-0.000	0.017
			(0.014)	(0.014)
GEP			**-0.112*****	**-0.129*****
			(0.014)	(0.014)
Formality			**0.053*****	**0.050*****
			(0.013)	(0.013)
Readability			**0.057*****	**0.048*****
			(0.012)	(0.012)
Tone			0.022	0.018
			(0.012)	(0.012)
Tabular Features Controlled	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes
Observations	38,968	38,968	38,968	38,968
Pseudo R^2	15.3%	15.9%	17.3%	18.0%

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

Results are reported in Table 3. All models include an intercept and controls for the tabular features. In Models 1 and 2, we use WQI alone to predict the funding outcome. In Models 3 and 4, we also control for text features while using WQI to predict fundability. Models 2 and 4 includes interaction terms between the WQI and the credit levels to help us investigate the heterogeneous effect of WQI for borrowers with different credit levels. We can see that overall the WQI has a significant positive association with fundability (Model 1), indicating that loans with higher WQI are more likely to be funded. When considering heterogeneous effects by credit levels (Model 2), we can see that WQI is positively associated with fundability for high-score borrowers, and negatively associated for low-score borrowers.

When controlling for text features, the effect of WQI becomes insignificant (Model 3). This

seems to suggest that the writing quality captured in WQI is explained away by the text features considered. Additionally, while controlling for text features, the heterogeneous effects are statistically significant (Model 4). There is a negative association between WQI and fundability for medium ($\hat{\beta}_3 = -1.789$, $p < 0.001$) and low-score borrowers ($\hat{\beta}_3 + \hat{\beta}_5 = -3.203$, $p < 0.001$), with a significant difference between them. High-score borrowers still show a positive association between WQI and fundability ($\hat{\beta}_3 + \hat{\beta}_4 = 3.920$, $p < 0.001$). Figure 5 visualizes the relationship between WQI and fundability based on Model 4 in Table 3. This result seems to suggest that writing in a more formal and standard way (i.e., with a higher WQI) is beneficial for high-score borrowers, but counterproductive for medium and low-score borrowers.

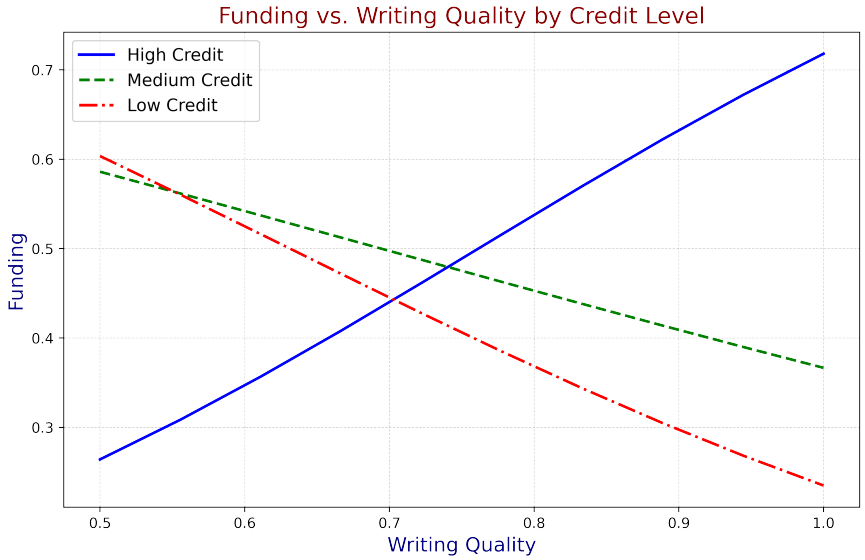


Figure 5: Heterogeneous Effect of WQI on Fundability.

4 Fundability and Soft Information in the GenAI Age

We describe our methodology in developing the lender model using a multimodal transformer adapted with Heckman correction. This model helps us examine the increase in fundability and the loss of soft information when loan descriptions are rewritten using GenAI.

4.1 A "Lender Model" with Heckman Correction

To analyze counterfactuals in the GenAI age, we need a model to mimic lenders' decisions. It is possible to reproduce lenders' decision process using a well trained machine model. This strategy has been adopted in existing studies (Bali et al., 2023; Cao et al., 2024). To jointly model text and tabular features as predictors, we adopt the multimodal transformer (Gu and Budhkar, 2021) as the base architecture (see more details in Section F.2).

It is important to note that our data are truncated: if a listing is not funded, we would not know whether the loan would have been repaid if funded. This poses a new challenge when using models trained on a biased sample. Traditionally, Heckman selection could be used to help 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

$$y_i = f(\mathbf{x}_i) + U_{1i}, \tag{4}$$

$$z_i = g(\mathbf{x}_i) + U_{2i}, \tag{5}$$

respectively. In these model specifications, $\mathbf{x}_i$ 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 our lending context, the prediction model f takes loan features $\mathbf{x}_i$ as input, and predicts its quality y_i, represented as the propensity of repayment. The selection model g would also take loan features $\mathbf{x}_i$ as input, and predicts 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. (4), there would be missing values for y_i for all the listing that were unfunded. Alternatively, if we use only the funded loans to fit Eqn. (4), its model estimates may be subject to a selection bias.

In the regression modeling context, 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 the selection model g, which would then be included as an extra term in the

prediction model f to account for the 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. (4) 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.

While the Heckman model was derived in the context of regression models, we may extend it for a deep learning model specification for f and g. In particular, instead of a linear specification, functions f and g will become nonlinear (a multimodal transformer in our case). Using a deep learning model for selection, we obtain samples' selection propensity from its output, and then follow the same logic as the classic Heckman model. To make this process more streamlined, we employ a custom, integrated learning algorithm inspired by a recent publication in the machine learning field (Kahng et al., 2023). In particular, suppose that both the selection and prediction models need to be estimated, we propose a negative log-likelihood loss function, which simultaneously fits the models f and g, and provide estimates of the covariance matrix following a maximized likelihood estimation framework. The learning objective can be specified as

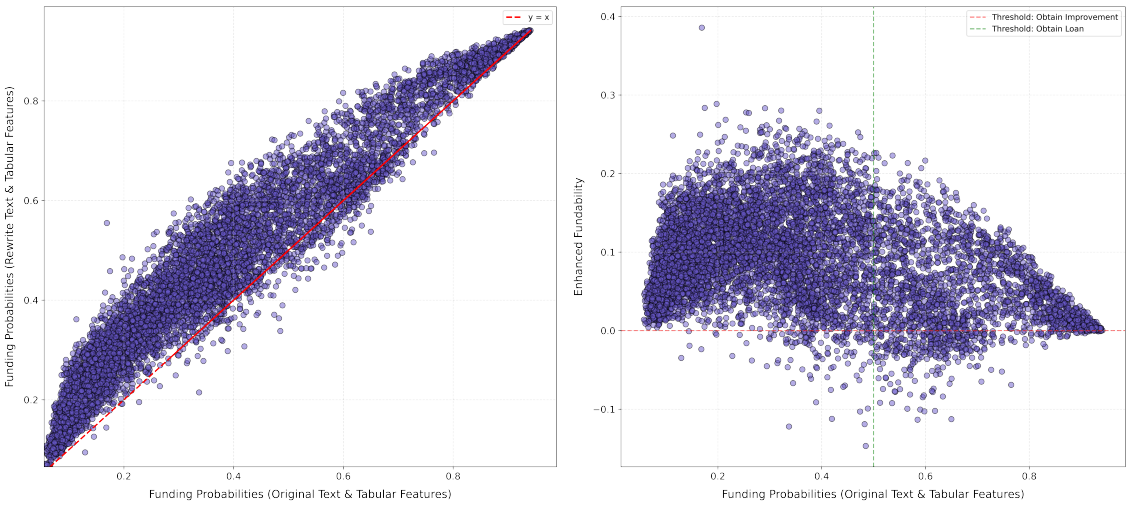
$$\arg\min_{f,g,\Sigma} \sum_i [s_i \Lambda(f(x_i), g(x_i); y_i, s_i; \Sigma) - \{s_i \log \Phi(g(x_i)) + (1 - s_i) \log \Phi(-g(x_i))\}] \quad (6)$$

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 $s_i = 1$. Using cross-validation to optimize hyperparameters, we adopted a final Heckman-style multimodal transformer as our "lender model."

4.2 GenAI Rewrites and Loan Fundability

As GenAI helps produce high-quality writing, it may help loan applicants enhance fundability. To see if the probability of obtaining funding would increase after rewriting, we compute the predicted funding probability using the rewritten vs. the original text as the input. Results are shown in Figure 6.

In Figure 6(a), each point represents a loan. The x-axis represents the predicted probability of funding using the original text, and the y-axis represents that using the text rewrites. We found that



(a) (b)

Figure 6: GenAI Rewrites Increase Fundability for Most Loans.

with rewritten text as the input, the lender model generally produces a higher predicted probability of being funded, as most data points are above the 45-degree line. To see which loans improved more, we further visualize the same set of loans in Figure 6(b), with the y-axis representing the improvement of fundability (defined as the predicted fundability using rewrites minus that using the original text). We can see that the largest improvements occurred to lower fundability loans, with an original funding probability of 0.5 or less. Overall, this result indicates a widespread improvement of fundability. From the borrower's perspective, it seems to suggest a trivial solution in terms of equilibrium adoption.

4.2.1 Attribution Test

Our next question then is which feature(s) led to greater improvements of fundability when loan description text is rewritten by GenAI? 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 rewrite text and that using the original text. Then, we fit an predictive model using the increased fundability as the target variable, text features of the original text, change in text features between the original and the rewrite text, WQI, and the

tabular features as predictors.[8] Based on this model, we evaluate the feature importance, and a list of the most important features are visualized in Figure 7. We can see that WQI is among the most important variables in predicting the improvement of fundability.

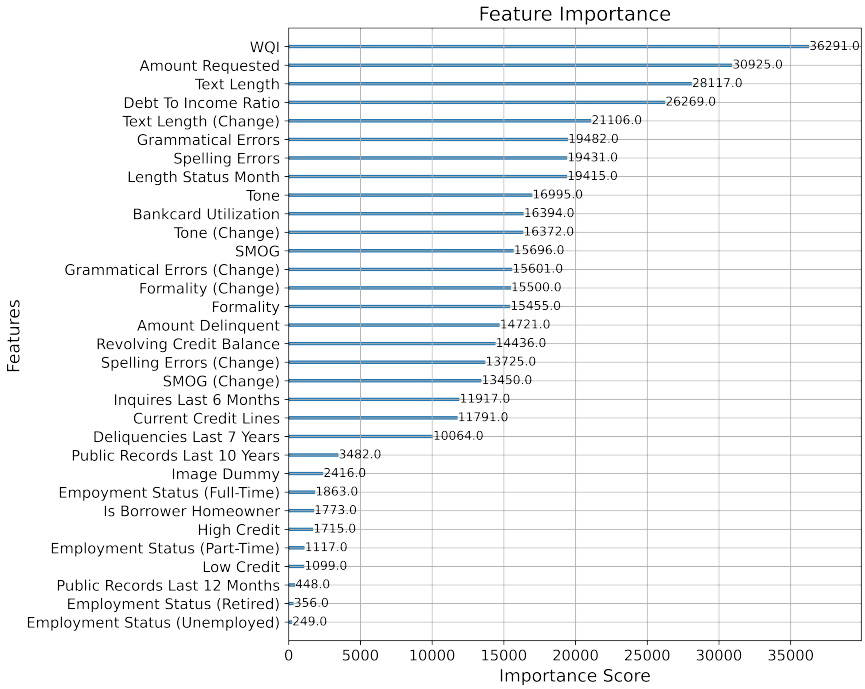


Figure 7: Feature Importance When Predicting Increased Fundability.

To picture the relationship between WQI and the improvement in fundability, we employ SHAP-based explanation (Lundberg and Lee, 2017) based on the joint predictive model. Specifically, we plot the partial dependence of the expected improvement of fundability by WQI in Figure 8. We can see that higher WQI values correspond to smaller fundability improvements. Since larger WQI values indicate that the original and the rewrite are more similar, the implication is there is less room for improvement in fundability by improving the text. Additionally, for lower WQI values (approximately 0.78 or less), the corresponding expected fundability improvement stays somewhat the same (at around 9.35 percentage points).

[8] We used the XGBoost model as it is widely known as one of the best performing models. Additionally, since several of the predictors have a strong correlation, a regression model may be subject to multicollinearity.

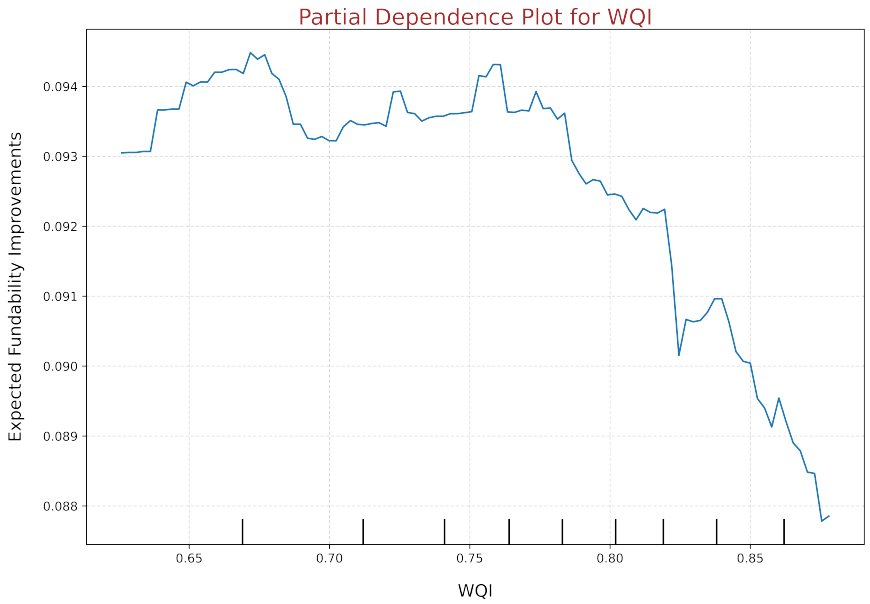


Figure 8: Partial Dependence Plot for WQI.

4.3 Soft Information: An Utility-Based Measurement

We measure soft information as the additional predictability of textual data on loan quality/outcomes on top of other non-textual information. In Eqn. (7), the utility function U would be a performance metric such as AUC or ROI assessed on the test set using a model trained on the training set.

$$\text{soft-information} = U(\text{text}, \text{tabular}) - U(\text{tabular}) \tag{7}$$

We randomly partition the loans into a training and a test set at the 4:1 ratio, develop the lender model based on the training set, and assess its performance on the test set. Results are reported in the first row of Table 4. First, when using AUC as the measure of soft information, as detailed in Table A11, the best performing model when using tabular features only would be the ensemble model, achieving an AUC of 72.5%. When both text and tabular data are used, the AUC reaches 74.7% by our lender model (i.e., the multimodal transformer with BERT for text embedding), showing an increase of 2.2 percentage points, which may be considered as soft information quantified by AUC.

Additionally, considering the risk-return trade-off between defaulted and repaid loans, we also

examine the soft information as measured by ROI.[9] Similarly, we found that the best performing model when using tabular features only would be the ensemble model, achieving an ROI of 16.02%. When both text and tabular data are used, the ROI reaches 18.63% by our lender model, showing an increase of 2.61 percentage points. This increase of ROI when using text in the predictive model may be considered as soft information as quantified by ROI.

Table 4: Measuring Soft Information Using AUC and ROI

	AUC			ROI		
Scenario	Model 1A (Tabu)[a]	Model 1B (Tabu + Text)	Diff.[b] (Soft Info.)	Model 2A (Tabu)[a]	Model 2B (Tabu + Text)	Diff.[b] (Soft Info.)
No GenAI		74.7%	2.2%***		18.63%	2.61%***
Small-Scale Adoption	72.5%	73.8%	1.3%***	16.02%	15.68%	-0.34
Large-Scale Adoption		74.0%	1.5%***		16.78%	0.76

[a] The best performing model using tabular features only is the ensemble model. The performance metric does not change by GenAI adoption since the model does not use text.
[b] Sig. levels: * $p < .05$, ** $p < .01$, *** $p < .001$. All AUCs and ROIs are averaged across 10 replications of the ten-fold mean.

4.4 Soft Information Lost in GenAI Adoption

The above findings are based on a pre-GenAI scenario, in which no one used GenAI to rewrite the loan descriptions. We call it the "*No GenAI*" scenario. This scenario is straightforward and corresponds to the traditional P2P lending investment environment as studied in most existing literature (Netzer et al., 2019; Gao et al., 2023). With GenAI, borrowers may start to use tools such as ChatGPT to produce higher quality writing for the loan description. However, whether a

[9] We adopt the method described in Netzer et al. (2019) to calculate ROI, which incorporates interest rates in 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, which is based on data published by crowdfunding consulting agencies. We then rank the loans by their expected profits, and select the top 1,000 loans with the highest expected profit to create an investment portfolio. Assuming \$100 is invested in these 1,000 loans, we calculate the actual profit (or loss) of each loan using their observed default status. In summary, 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.

borrower used GenAI's help in writing the loan description is not disclosed to lenders. Therefore, lenders may not have responded to the market by adjusting their strategies. We call this scenario "*Small-Scale Adoption*". To simulate this scenario using our data, we use the same lender model as trained in the no-GenAI scenario, while updating the text description of the test data using our ChatGPT prompts. This scenario represents the online credit market where GenAI is adopted by borrowers, and lenders' decision-making process has not yet accounted for this change.

The second row of Table 4 shows results under small-scale adoption. Specifically, our lender model reports a 73.8% AUC and 15.68% ROI. Compared with the utility without leveraging text, the quantified soft information would be reduced to a 1.3 percentage point by AUC and nearly none by ROI (a 0.34 percentage point decline without statistical significance). Comparing small-scale adoption with the no-GenAI scenario, we observe a decline in utility, which suggests a reduction in soft information.

The reduction of soft information may be partially attributed to the convergence of writing. To demonstrate this point, we compute a group homogeneity index as

$$GroupHomogeneity = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} CosineSimilarity(emb(\mathbf{x}_i), emb(\mathbf{x}_j)), \quad (8)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ represent the text description of loan i and loan j, respectively and n represents the total number of samples in this group. We compute this group homogeneity index by credit level and compare them before and after rewrites.[10] Results are in Figure 9, from which we can see several interesting patterns. Firstly, borrowers with better credit profiles show much less homogeneity than those with worse credit profiles before the rewrites. This indicates that high-score borrowers write in a more diverse manner than low-score borrowers. Secondly, after rewrites, the homogeneity of each group increased substantially, with the high-score groups increased relatively more than the low-score groups, narrowing the gap in homogeneity across credit levels. This finding suggests that the application of GenAI increases the homogeneity of writing. The convergence in writing may explain the reduction in soft information when lenders attempt to distinguish high quality loans from low quality ones.

[10] Due to the substantial computational expense of calculating similarity for all sample pairs, we employed a bootstrap sampling approach (Hernández et al., 2017) to estimate the group homogeneity. Specifically, we randomly sampled 20% of the data within each group for computation with 10 repetitions.

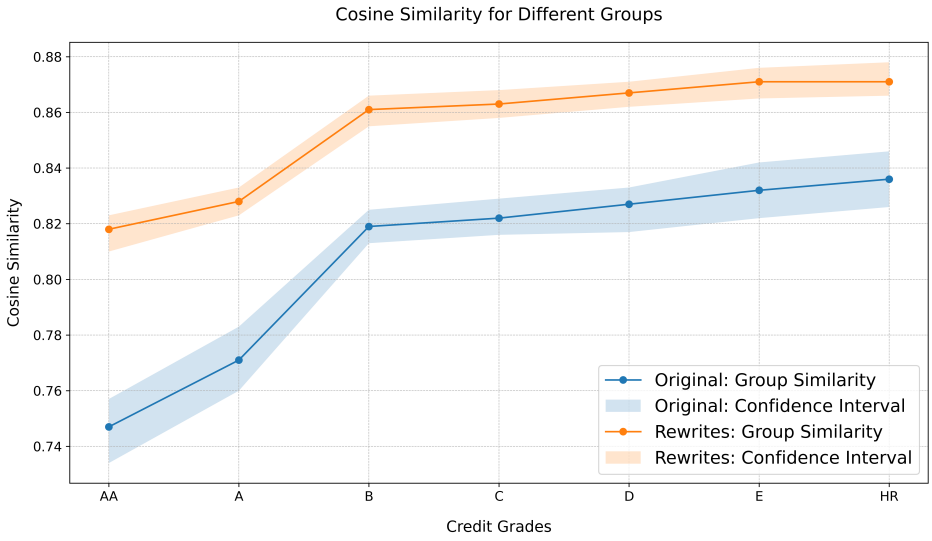


Figure 9: Text Similarity by Credit Grades Before and After Rewriting.

5 Lender Response and General Equilibrium

A caveat to the analysis above is that while we assume that 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 model, especially when GenAI adoption is widespread. If they retrain their model based on loans whose text descriptions are improved with the help of GenAI, they can pick up the signals, or lack of, that remain in the text. We call this scenario "*Large-Scale Adoption*." With large-scale adoption of GenAI, the decision-making processes of lenders are likely to be adjusted. In other words, the predictive model for simulating lender decisions can now be primarily trained on GenAI-generated data. To simulate this scenario, we first rewrite the original text using GenAI, including both the training and the test data. Then, a lender model is trained using the updated training data, whose text has been rewritten by GenAI. The test data also has their text descriptions rewritten by GenAI. Figure 10 illustrates the three cases of adoption.

In this section, we first evaluate the extent of misallocation in each of the GenAI adoption scenarios and then derive the market equilibrium between borrowers and lenders.

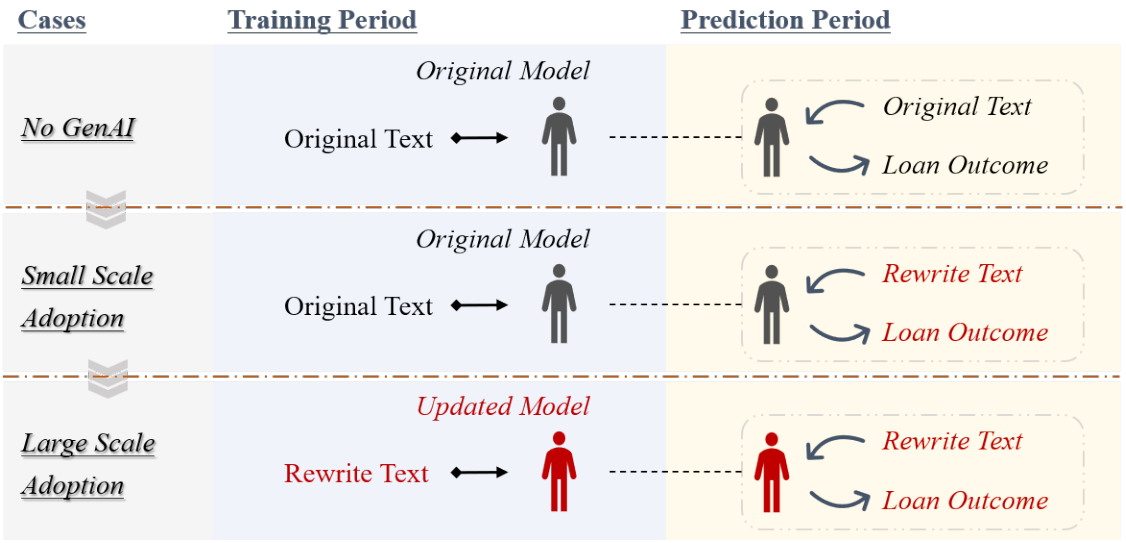


Figure 10: GenAI Adoption Scenarios.

5.1 Misallocation

To analyze the impact of GenAI adoption on lenders' misallocation, we first train a lender model using the training set (with the original text and Heckman correction) and use its prediction as the counterfactual ground truth. Then, we evaluate misallocation using FPR and FNR as the metrics.[11] Results are summarized in Table 5.

We can see that without GenAI, the baseline FPR and FNR would be at 17.50% and 20.86%, respectively. In the context of small-scale adoption, FPR and FNR significantly increase to 23.12% and 23.28%, reflecting an increase of 5.62 and 2.42 percentage points (or a 32.1% and 11.6% relative increase), respectively. This indicates a substantially increase in misallocation if borrowers leverage GenAI to improve writing while lenders have not updated their decision model. When lenders do update their decision model, as in the large-scale adoption scenario, the FPR would become 20.52%, which is still much higher than the no-GenAI scenario but reflects a substantial (2.6 percentage points, or a 11.2% relative) reduction compared with the small-scale adoption scenario. Additionally, the FNR would become 20.28%, which is equivalent to the no-GenAI

[11] Considering repayment as the positive outcome, a case is considered positive (P) if the model classifies it as repaid and negative (N) if the model classifies it as defaulted. A false positive (FP) case would be one that the model classifies it as repaid and it actually defaulted, and a false negative (FN) case would be one that the model classifies it as defaulted and it was actually repaid. The false positive rate is calculated as 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).

scenario (i.e., no statistical difference found) and reflects a substantial (3.0 percentage points, or a 12.9% relative) reduction compared with the small-scale adoption scenario. These trends indicates that with borrowers' adoption of GenAI, if the lender updates the decision model, the FPR and FNR can partially recover.

Table 5: Misallocation in Different GenAI Adoption Scenarios

Scenario	FPR	DIFF.[a]	FNR	DIFF.[a]
No GenAI	17.50%		20.86%	
Small-Scale Adoption	23.12%	**+5.62%*****	23.28%	**+2.42%*****
Large-Scale Adoption	20.52%	**+3.02%*****	20.28%	- 0.58%

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

How do the misallocation errors reflect in the ROI? The third row of Table 4 shows the ROI results under large-scale adoption. Specifically, our lender model reports a 74.0% AUC and 16.78% ROI. Compared with the utility without leveraging text, the quantified soft information would be reduced to a 1.5 percentage point by AUC and nearly none by ROI (a 0.76 percentage point increase without statistical significance). Comparing large-scale adoption with the small-scale adoption scenario, we observe an improvement in the lender's utility, which suggests that soft information is somewhat recaptured and financial loss mitigated.

5.2 A Generative General Equilibrium

Based on the above analysis, it may be reasonable to assume that lenders shall always update their decision model (and use both tabular and text features). Borrowers need to decide whether to rewrite. No one knows the true equilibrium adoption rate among all borrowers, once GenAI is widely adopted. We follow Cong (2024) to leverage the power of GenAI itself to analyze a counterfactual long-run equilibrium.

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 decision and the loan applications generated, we then train the proprietary lender's model. Finally, we verify that the equilibrium adoption is indeed the conjectured value and the borrowers' adoption

decisions are rational. We repeat the process until all the equilibria are identified.

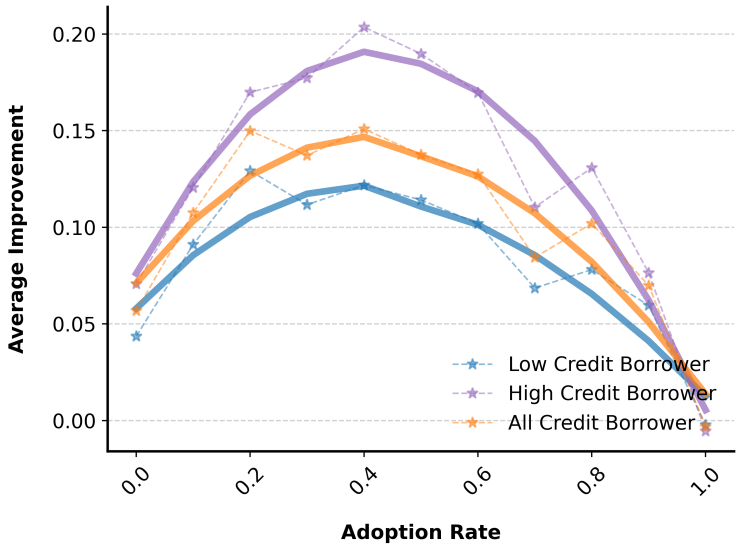


Figure 11: Market Equilibrium. (Dashed lines are results computed at increments of 10% market adoption rate, and solid lines are smoothed curves.)

Specifically, we define borrowers' utility as the improvement in fundability based on the rewrite when compared with the fundability using the original text. We also conjecture that high-score borrowers have the highest incentives to adopt GenAI. Results are shown in Figure 11. We can see that as the adoption rate increases from 0 to about 0.4, there is a rapid increase in the fundability improvements for all borrowers, but the largest increase is for high-score borrowers. As the adoption rate continues to increase from about 0.4 to 1, this improved funding rate begins to taper off, but at the same rate of adoption, high-score borrowers still benefit the most; as the probability of adoption nears 1, the increased gains for all borrowers reach a minimum and are essentially similar. Despite some heterogeneity (e.g., higher-credit borrowers receive a higher utility, mid-level market adoption leads to higher personal utility) almost all borrowers, irrespective of their credit levels or assumed market adaoption rates, receive a positive utility from GenAI adoption. Therefore, if borrowers are rational and the cost of GenAI use is minimal, all borrowers are motivated to adopt GenAI, resulting a unique full-adoption long-run equilibrium.

In this equilibrium, every borrower adopts GenAI for online loan applications, and lenders' decision model is data-driven and trained on the synthetic rewrites. As we have shown earlier, the loss of soft information is somewhat mitigated relative to the partial equilibrium outcomes, but the

lender still relies more on hard information compared to the case without GenAI.

6 Conclusion

We utilize a predominant large language model to investigate how various dimensions of writing quality of online loan applications affect credit allocation. We also study the potential impacts of GenAI on online credit markets, in which borrowers can use unstructured text to communicate their financial situation. To imitate lender decisions, we train a multimodal transformer, which extends Heckman correction to deep learning models. Specifically, we find that GenAI can help significantly improve the writing quality and fundability. Lower credit borrowers exhibit higher writing quality, as measured by a novel Writing Quality Index (WQI), than higher credit borrowers. Despite the enhanced predictive power of newer deep learning models, GenAI rewrites have reduced predictive power of the text description. By making high-quality writing accessible to all borrowers, GenAI poses a significant challenge to the transmission of soft information due to the convergence in writing. Moreover, borrowers' strategic use of GenAI could lead to a systematic gap between the perceived and actual loan risk, increasing the likelihood of credit misallocation. However, once lenders update their decision model to account for this new reality, the loss in soft information and the misallocation are mitigated. We characterize the unique general equilibrium with full GenAI adoption.

Our findings have the following implications in the context of GenAI and online credit markets. First, lenders should anticipate an increase in writing quality and a decrease in information value from the text-based loan description. The weakened linkage between writing and financial behavior due to GenAI has profound implications for the future of financial decision-making. If borrower-generated text can no longer reliably signal behavioral traits, lenders will need to reevaluate the weight they assign to textual information in their credit assessment models. This shift may necessitate a greater reliance on alternative forms of soft information, such as digital footprints or social network data, or a renewed focus on hard data like credit scores. Second, while low-score borrowers have leveraged the free text as their chance to explain and persuade investor for funding, high-score borrowers did not make as much an effort in producing high quality writing. With GenAI tools, improving writing quality gets much easier and thus high-score borrowers, especially

early adopters, may substantially benefit from GenAI-aided writing. Finally, the platform might take an active role in facilitating platform participants. For example, they may integrate an advanced AI detector to warn investors that an narrative might used help from GenAI. Although it is still extremely hard to reliably detect whether a description is improved by GenAI as of today, providing a probabilistic guesstimate with full transparency may still prove helpful. As another example, the platform might also use custom GenAI prompts to standardize text written by different individuals to remove unwanted or misleading information. In that case, a metric like WQI, which measures the similarity between the original text and the rewrite (of a custom prompt) may also provide helpful predictive value. In summary, the rise of AI-generated text could spur the development of new analytical tools designed to detect and adjust for AI-driven enhancements, thereby preserving the integrity of text-based assessments for loan quality.

References

Turan G Bali, Bryan T Kelly, Mathis Mörke, and Jamil Rahman. Machine forecast disagreement. Technical report, National Bureau of Economic Research, 2023.

Fabio Caldieraro, Jonathan Z Zhang, Marcus Cunha Jr, and Jeffrey D Shulman. Strategic information transmission in peer-to-peer lending markets. *Journal of Marketing*, 82(2):42–63, 2018.

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.

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 Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423.

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 & Finance*, 64: 169–187, 2016.

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.

Anindya Ghose and Panagiotis G Ipeirotis. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498–1512, 2010.

Ken Gu and Akshay Budhkar. A package for learning on tabular and text data with transformers. In *Proceedings of the Third Workshop on Multimodal Artificial Intelligence*, pages 69–73, Mexico City, Mexico, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.maiworkshop-1.10. URL https://www.aclweb.org/anthology/2021.maiworkshop-1.10.

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.

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.

Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 4768–4777, Red Hook, NY, USA, 2017.

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.

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.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China, 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1410. URL https://aclanthology.org/D19-1410.

Joe Ryan, Katya Reuk, and Charles Wang. To fund or not to fund: Determinants of loan fundability in the prosper.com marketplace. Technical report, The Stanford Graduate School of Business, January 2007. URL https://www.prosper.com/downloads/research/prosper_regression_projectfundability_study.pdf. Confidential.

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, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Lifang Zhang, Jianzhou Wang, and Zhenkun Liu. What should lenders be more concerned about? developing a profit-driven loan default prediction model. *Expert Systems with Applications*, 213: 118938, 2023.

Christoph Zott and Quy Nguyen Huy. How entrepreneurs use symbolic management to acquire

resources. *Administrative Science Quarterly*, 52(1):70–105, 2007.

Appendix A Details of Features

Based on existing research, we have compiled the commonly used tabular features in the online lending market, primarily classified into two categories: Verified Hard Credit Information and Non-Credit Information (Gao et al., 2023; Netzer et al., 2019; Caldieraro et al., 2018). For each loan application in the online credit market, we focus on two aspects: (1) whether the loan is funded, and (2) whether the loan is repaid (for funded loans only). For more details, see Table A1.

Table A1: Definitions of Tabular Features

ID	Feature	Description[a]
Panel A: Verified Hard Credit Information		
1	CreditGrade	Letter assigned to application based on the borrower's credit score. The seven credit grades are AA (least risky), A, B, C, D, E, HR (high risk; most risky).
2	BackcardUtilization	The sum of the balances owned by the borrower on open bankcards divided by the sum of the cards' credit limits.
3	AmountDelinquent	Natural logarithm of (1 + dollar amount delinquent).
4	DelinquenciesLast7Years	Natural logarithm of (1+number of delinquencies in prior 7 years).
5	InquiriesLast6Months	The number of credit inquiries in previous 6 months.
6	PublicRecordsLast10Years	The number of public records (bankruptcy, civil judgments, and tax liens) in the prior 10 years.
7	PublicRecordsLast12Months	The number of public records (bankruptcy, civil judgment, and tax liens) in the previous year.
8	CurrentCreditLines	Natural logarithm of (1 + borrower's number of current credit lines.)
9	OpenCreditLines	Natural logarithm of (1 + number of borrower's open credit lines.)
10	RevolvingCeditBalance	Natural logarithm of (1 + borrower's dollar amount of revolving credit balance.)
11	AmountRequested	Natural logarithm of the loan amount requested by the borrower.
Panel B: Non-Credit Information		
12	DebtToIncomeRatio	Borrowers' debt-to-income ratio.
13	IsBorrowerHomeowner	Indicator variable for home ownership.
14	Income	Indicators for income level.
15	EmploymentStatus	Employment status indicators.
16	LengthStatusMonths	The length in months of the employment status at the time the listing was created.
17	ImageDum	An indicator for listings with photograph of at least one adult.
Panel C: Loan Info.[b]		
18	BorrowerMaximumRate	Maximum rate a borrower is willing to pay.
19	BorrowerRate	The is the rate the borrower pays if the loan were to close at this point in time.
20	LenderRate	The rate that lenders would receive on the listing if the loan were to close at this point in time.
21	BidMaximumRate	The Maximum Rate in which a bidder will be able to lend money in order to create a winning bid.
22	BidCount	The total number of bids a listing receives.
Panel D: Labels		
23	Fund_label	An indicator that equals to one if a listing is fully funded and becomes a loan.
34	Repay_label	An indicator that equals to one if the loan fully repaid.

[a.] Variable description is adapted from Gao et al. (2023) when they used them; and from the Prosper website (https://www.prosper.com/Downloads/Services/Documentation/ProsperDataExport_Details.html) otherwise.
[b.] Panel C features are observed during the bidding stage. To avoid information leakage, they are used to predict 'Repay', but not used to predict 'Fund'.

These tabular features include categorical variables such as *Credit Grade* and *Income*, as well as numerical variables such as *Debt To Income Ratio* and *Amount Requested*. Table A2 presents the descriptive statistics of these tabular features.

To examine the relationship between loan status (funded - unfunded; repay - default) and tabular features, we also compared the differences in loan characteristics within different groups.

Table A2: Descriptive Statistics for Tabular Features

		Unfunded (n=19,484)		Funded (n=19,484)		Default (n=6,886)		Repaid (n=12,598)	
ID	Numeric Variable	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
1	AmountRequested	8.999	0.814	8.434	0.853	8.393	0.845	8.509	0.862
2	BidCount	24.74	76.65	154.5	148.8	156.2	145.9	151.4	154.0
3	BidMaximumRate	0.178	0.085	0.181	0.077	0.164	0.073	0.210	0.074
4	BorrowerMaximumRate	0.181	0.085	0.218	0.080	0.202	0.080	0.246	0.072
5	BorrowerRate	0.179	0.085	0.182	0.077	0.166	0.073	0.213	0.074
6	LenderRate	0.178	0.085	0.181	0.077	0.165	0.073	0.211	0.074
7	DebtToIncomeRatio	0.574	1.408	0.325	0.892	0.297	0.793	0.378	1.046
8	BankcardUtilization	0.554	0.371	0.552	0.373	0.524	0.361	0.604	0.388
9	AmountDelinquent	2.827	3.710	1.715	3.135	1.385	2.882	2.317	3.473
10	DelinquenciesLast7Years	1.034	1.300	0.820	1.204	0.725	1.146	0.994	1.287
11	InquiriesLast6Months	3.313	4.258	2.580	3.551	2.031	2.863	3.584	4.372
12	PublicRecordsLast10Years	0.440	0.956	0.368	0.885	0.305	0.746	0.483	1.084
13	PublicRecordsLast12Months	0.044	0.254	0.038	0.235	0.032	0.207	0.049	0.279
14	CurrentCreditLines	2.270	0.613	2.195	0.617	2.205	0.591	2.176	0.662
15	OpenCreditLines	2.138	0.609	2.056	0.617	2.066	0.591	2.038	0.661
16	RevolvingCreditBalance	8.051	2.960	7.880	2.925	7.986	2.773	7.686	3.176
17	LengthStatusMonths	23.37	56.23	23.70	54.80	21.80	53.04	27.19	57.71
	Categorical Variable	n	(%)	n	(%)	n	(%)	n	(%)
18	ImageDum	10,542	54.1%	13,114	67.3%	4,594	66.7%	8,520	67.6%
19	IsBorrowerHomeowner	9,919	50.9%	10,279	52.8%	3,457	50.2%	6,822	54.2%
20	Income (Not displayed): 0	202	1.0%	72	0.4%	23	0.3%	49	0.4%
	Income ($0 or unable to verify): 1	0	0.0%	564	2.9%	221	3.2%	343	2.7%
	Income ($1-24999): 2	2,731	14.0%	2376	12.2%	866	12.6%	1,510	12.0%
	Income ($25,000-49,999): 3	7,256	37.2%	7272	37.3%	2,646	38.4%	4,626	36.7%
	Income ($50,000-74,999): 4	4,893	25.1%	4940	25.4%	1,708	24.8%	3,232	25.7%
	Income ($75,000-99,999): 5	2,117	10.9%	2214	11.4%	758	11.0%	1,456	11.6%
	Income ($100,000+): 6	2,121	10.9%	1915	9.8%	631	9.2%	1,284	10.2%
	Income (Not Employed): 7	164	0.8%	131	0.7%	33	0.5%	98	0.8%
21	CreditGrade AA	2,571	13.2%	2,571	13.2%	448	6.5%	2,123	16.9%
	CreditGrade A	2,529	13.0%	2,529	13.0%	654	9.5%	1,875	14.9%
	CreditGrade B	3,373	17.3%	3,373	17.3%	1,132	16.4%	2,241	17.8%
	CreditGrade C	4,166	21.4%	4,166	21.4%	1,523	22.1%	2,643	21.0%
	CreditGrade D	3,568	18.3%	3,568	18.3%	1,436	20.9%	2,132	16.9%
	CreditGrade E	1,657	8.5%	1,657	8.5%	791	11.5%	866	6.9%
	CreditGrade HR	1,620	8.3%	1,620	8.3%	902	13.1%	718	5.7%
22	EmploymentStatusFT	16,414	84.2%	16,722	85.8%	5,843	84.9%	10,879	86.4%
	EmploymentStatusNE	203	1.0%	157	0.8%	42	0.6%	115	0.9%
	EmploymentStatusPT	755	3.9%	754	3.9%	209	3.0%	545	4.3%
	EmploymentStatusR	650	3.3%	377	1.9%	148	2.1%	229	1.8%
	EmploymentStatusSE	1,462	7.5%	1,474	7.6%	644	9.4%	830	6.6%

Note: All features were determined during the pre-funding stage.

For comparing categorical variables across groups, we encoded them as follows: Credit Grades were represented by numbers 1-7, with 1 for the lowest grade (HR) and 7 for the highest (AA); Employment Status was coded as 1 for full-time and 0 otherwise; Is Borrower Homeowner and Image Dummy, both binary, were coded 1 for "yes" and 0 for "no". Income was encoded based on Prosper coding rules. Details are shown in Table A3.

Table A3: Difference between Subsamples

	Stage 1 Outcome: Funded vs. Not				Stage 2 Outcome: Default vs. Repay			
	Funded	Unfunded	Diff.(F-U)[a]	T	Repay	Default	Diff.(R-D)[a]	T
Panel A: Tabular variables for loan applications.								
Amount Requested	8.434	8.999	**-0.565*****	-66.83	8.393	8.509	**-0.116*****	-9.074
Debt To Income Ratio	0.325	0.574	**-0.249*****	-20.80	0.297	0.378	**-0.081*****	-6.074
Bankcard Utilization	0.552	0.554	-0.002	-0.463	0.524	0.604	**-0.080*****	-14.27
Amount Delinquent	1.715	2.827	**-1.112*****	-31.95	1.385	2.317	**-0.932*****	-20.04
Delinquencies Last 7 Years	0.820	1.034	**-0.214*****	-16.80	0.725	0.994	**-0.269*****	-15.00
Inquiries Last 6 Months	2.580	3.313	**-0.733*****	-18.46	2.031	3.584	**-1.533*****	-29.84
Public Records Last 10 Years	0.368	0.440	**-0.072*****	-7.734	0.305	0.483	**-0.178*****	-13.51
Public Records Last 12 Months	0.038	0.044	**-0.006***	-2.398	0.032	0.049	**-0.017*****	-4.805
Current Credit Lines	2.195	2.270	**-0.075*****	-12.15	2.205	2.176	**0.029****	3.170
Open Credit Lines	2.056	2.138	**-0.082*****	-13.13	2.066	2.038	**0.028****	3.115
Revolving Credit Balance	7.880	8.051	**-0.171*****	-5.750	7.986	7.686	**0.300*****	6.842
Length Status Months	23.70	23.37	+0.596	0.596	21.80	27.19	**-5.287*****	-6.567
Bid Count	154.5	24.74	**+129.8*****	108.2	156.2	151.4	**4.876***	2.187
Bid Maximum Rate	0.181	0.178	**+0.003****	2.711	0.164	0.210	**-0.046*****	-41.89
Borrower Maximum Rate	0.218	0.181	**+0.037*****	43.96	0.202	0.246	**-0.044*****	-37.74
Borrower Rate	0.182	0.179	**+0.003*****	3.665	0.166	0.213	**-0.047*****	-42.52
Lender Rate	0.181	0.178	**+0.003****	3.286	0.165	0.211	**-0.046*****	-41.89
Credit Grades	4.226	4,226	-	-	4.503	3.781	**0.785*****	30.45
Employment Status	0.858	0.842	**0.016*****	4.375	0.864	0.849	**0.015****	2.873
Is Borrower Homeowner	0.472	0.509	**-0.037*****	-7.240	0.458	0.498	**-0.040*****	5.280
Income	3.612	3.657	**-0.045*****	-3.573	3.639	3.562	**0.077*****	4.030
Image Dummy	0.673	0.541	**0.132*****	26.92	0.676	0.667	0.009	1.301
Panel B: Textual variables from loan description.								
Text Length	222.2	182.8	**39.38*****	29.73	215.1	235.1	**-19.98*****	-9.590
Spelling Errors	0.033	0.036	**-0.003*****	-12.44	0.034	0.033	**0.001*****	3.782
Grammatical Errors	0.059	0.068	**-0.009*****	-22.17	0.057	0.063	**-0.006*****	-10.51
Formality	0.236	0.243	**-0.007*****	-12.60	0.235	0.237	-0.002	-1.828
Readability	10.66	10.24	**0.415*****	10.85	10.71	10.56	**0.157****	3.132
Tone	0.752	0.739	**0.013*****	7.643	0.755	0.744	**0.011*****	4.713

[a] Significance were determined using two-sample T tests. Sig. levels: * p <.1, ** p <.05, *** p <.01. The Diff. column represents the difference in means of the funded (or repaid) group minus the unfunded (or default) group.
[c] The unit for Amount Requested is USD 1,000. The unit for StatusTime is year (12 months).

Appendix B Text Features and Fundability

B.1 Text Features and WQI

we analyze the relationship between text features and WQI, as shown in Table A4. From Table 1, we observe a significant correlation between WQI and all text features, with the text features explaining 11.0% of WQI. This suggests that WQI captures both text feature-related information and additional information beyond the text features.

Table A4: Relationship Between Text Features and WQI

	DV: WQI						
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
DWN	0.135***						0.114***
	(0.003)						(0.003)
SEP		-0.804***					-0.393***
		(0.024)					(0.028)
GEP			-0.195***				0.082***
			(0.012)				(0.014)
LWP				-0.203***			-0.050***
				(0.009)			(0.010)
SMOG					0.090***		0.060***
					(0.003)		(0.003)
Tone						-0.010***	-0.025***
						(0.003)	(0.003)
Constant	0.737***	0.800***	0.785***	0.821***	0.738***	0.780***	0.758***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)
Observations	38,968	38,968	38,968	38,968	38,968	38,968	38,968
R-squared	0.084	0.043	0.010	0.021	0.024	0.000	0.110

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

B.2 Text Features on Fundability

To check the relationship between text features and getting funded, we employ a logistic regression model and select credit features as control variables to conduct the significance test. The model

may be specified as follows:

$$\begin{aligned}\text{logit}(Outcome_i) =& \beta_0 + \beta_1 \times IsHighCredit_i + \beta_2 \times IsLowCredit_i + \beta_3 \times TF_i \\ &+ \beta_4 \times TF_i \times IsHighCredit_i \\ &+ \beta_5 \times TF_i \times IsLowCredit_i + \boldsymbol{\gamma} \cdot \boldsymbol{c}_i + \varepsilon_i, \end{aligned} \tag{9}$$

where $Outcome_i$ represents whether loan i is funded, TF represents each text 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 text feature would be $\hat{\beta}_3$ 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. Table A5 summarizes the results where each column represents a model with the text feature indicated in the model name interacting with the credit levels. Estimated effects by credit level are visualized in Figure 3.

Also note that with the same level of model complexity (i.e., allowing one text feature to interact with credit level while controlling for all the other text features), WQI reaches the highest R^2 (18.0% for WQI as seen in Model 4 of Table 3, as opposed to 17.4–17.5% for all models in Table A5), which supports its comprehensiveness in measuring text characteristic.

B.3 Text Features on Fundability Improvements

We examine the relationship between features and fundability improvements using SHAP-based explanations (Lundberg and Lee, 2017) based on the XGBoost model as described in Section 4.2.1. Focusing on baseline text features (those of the original text) and their change values (i.e., the difference in text features between rewrite and the original text), we present the partial dependence plot in Figure 12.

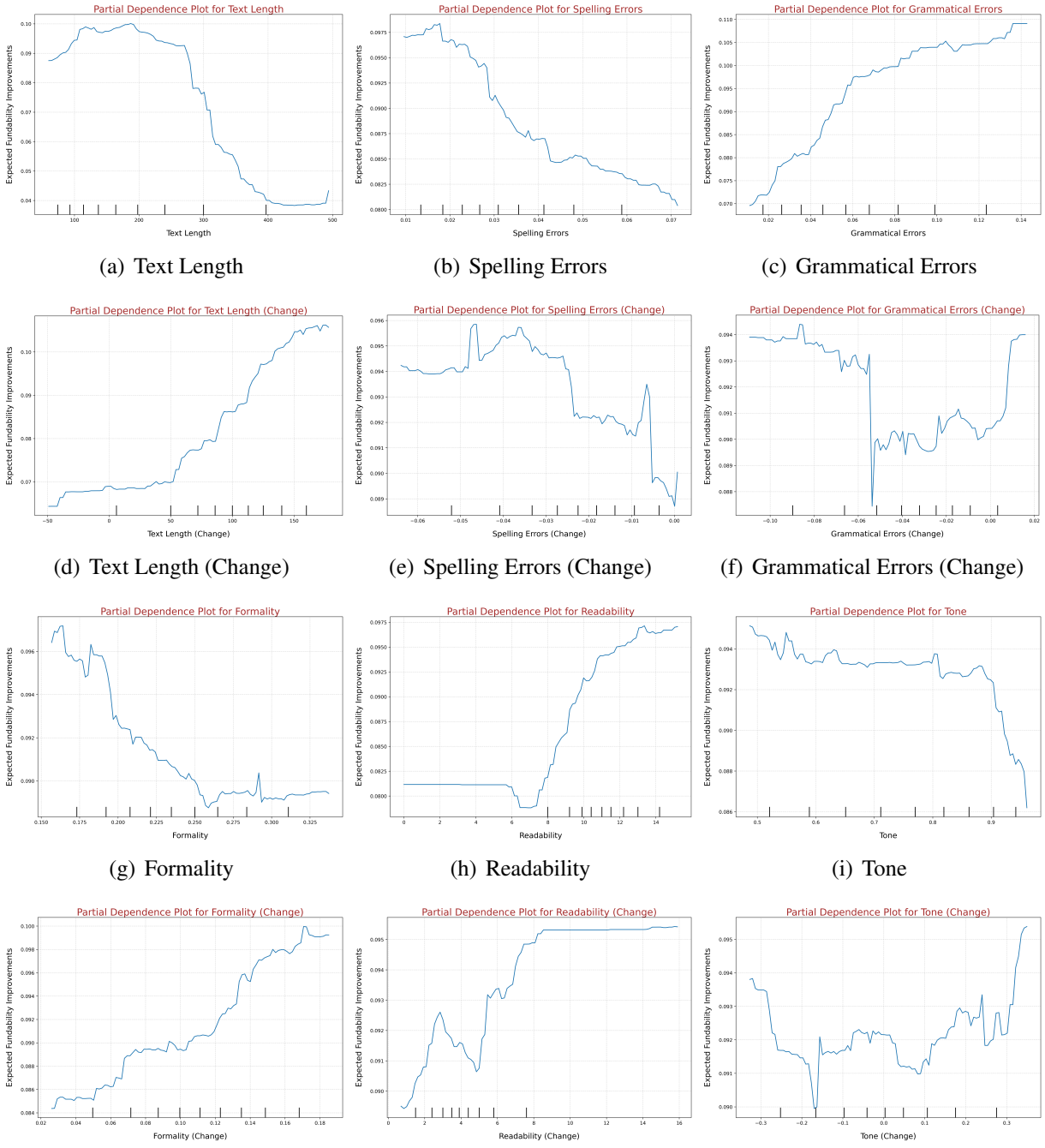
From Figure 12(a), we can see that shorter baseline text length corresponds to greater improvement in fundability. Proportionately, larger changes in text length made by GenAI is linked to greater improvement in fundability, as shown in Figure 12(d). This suggests that when the baseline text length is low, there is more room to make changes; and since text length is positively linked to fundability, larger increase in text length is linked to a larger increase in fundability. As we can see in Figure 12(g) and Figure 12(j), a similar pattern can be found in formality.

Table A5: Relationship Between Text Features and Funding Outcome

	DV: Funding (Funded: 1, Unfunded: 0)					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
TF ▷	**Text Length**	**Spelling Error**	**Grammatical Error**	**Formality**	**Readability**	**Tone**
TF × High Credit	**0.134*****	**-0.132*****	-0.011	**-0.098*****	**0.073***	0.015
	(0.032)	(0.026)	(0.032)	(0.027)	(0.029)	(0.028)
TF × Low Credit	**0.314*****	-0.060	**-0.260*****	**-0.169*****	-0.049	**0.094****
	(0.033)	(0.037)	(0.034)	(0.033)	(0.032)	(0.030)
Text Length	**0.270*****	**0.363*****	**0.355*****	**0.36*****	**0.360*****	**0.364*****
	(0.019)	(0.015)	(0.016)	(0.015)	(0.015)	(0.015)
SEP	0.003	**0.068*****	-0.001	0.001	0.008	-0.001
	(0.015)	(0.020)	(0.014)	(0.014)	(0.015)	(0.014)
GEP	**-0.118*****	**-0.124*****	**-0.071*****	**-0.112*****	**-0.117*****	**-0.112*****
	(0.015)	(0.015)	(0.017)	(0.014)	(0.014)	(0.014)
Formality	**0.050*****	**0.050*****	**0.052*****	**0.108*****	**0.049*****	**0.054*****
	(0.014)	(0.013)	(0.013)	(0.017)	(0.013)	(0.013)
Readability	**0.057*****	**0.052*****	**0.058*****	**0.058*****	**0.045****	**0.057*****
	(0.012)	(0.012)	(0.012)	(0.012)	(0.015)	(0.012)
Tone	0.022	0.021	0.022	0.023	0.021	0.002
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.015)
WQI	-0.245	-0.289	-0.257	-0.277	-0.303*	-0.285
	(0.155)	(0.154)	(0.154)	(0.154)	(0.154)	(0.154)
Tabular Features Controlled	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,968	38,968	38,968	38,968	38,968	38,968
Pseudo R^2	17.5%	17.4%	17.5%	17.4%	17.4%	17.4%

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

Similarly, higher percentages of baseline spelling and grammatical errors provide greater opportunities for correction, leading to significant increases in fundability. In terms of readability, Figure 12(h) shows that improvements are consistently linked to positive impacts on fundability. For tone, however, a U-shaped relationship with fundability is observed. This phenomenon carries two implications. First, for loan descriptions with low tone, increasing the positivity leads to fundability improvements. Second, for loan descriptions with excessively high tone, reducing the level of positivity similarly enhances fundability. This observation suggests that the tone of loan descriptions should be balanced, avoiding extremes of excessive positivity.



(a) Text Length (b) Spelling Errors (c) Grammatical Errors

(d) Text Length (Change) (e) Spelling Errors (Change) (f) Grammatical Errors (Change)

(g) Formality (h) Readability (i) Tone

(j) Formality (Change) (k) Readability (Change) (l) Tone (Change)

Figure 12: Text Features, Changes and Fundability.

Appendix C Example Rewrites

In this section, we present some examples of GenAI rewrites. As shown in Table A6, GenAI's rewrites significantly improve the readability and formality of the loan descriptions.

Table A6: 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 D Human Assessments

Our findings rest on the assumption that machine learning algorithms replicate human decision-making processes. This assumption has gained traction in recent research, where machine learning models are increasingly used to simulate the judgments of human experts (Bali et al., 2023). However, it remains possible that what an algorithm identifies as fundable may not fully align with human investor perceptions. To address this potential discrepancy and verify the robustness of our results, we conducted an online experiment with human testers.

To generate the stimuli, we drew a stratified sample of our dataset based on funding status (unfunded, funded), credit level (high, medium, low), and WQI (divided into four quantiles). From the 24 possible combinations of these 3 factors, we randomly selected 5 loan applications per combination, and obtained their rewrite version. In other words, we have 120 loan application samples, with their text description before and after rewrite.

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, while the other evaluated their willingness to fund based on the text. Each participant evaluated 48 texts, which include 24 original descriptions and their corresponding rewrites, with the sentence order randomly shuffled. Participants were not informed whether the texts were original or rewritten versions, and the correspondence between the original and rewritten texts was not disclosed as all 48 pieces of text were randomly shuffled.

D.1 Does the GenAI-rewritten version exhibit a higher writing quality?

One group of human raters were shown 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 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.*"

As shown in Table A7, the writing quality of loan applications is enhanced when rewritten by GenAI. This result is consistent with our expectations and provides additional robustness to the findings presented earlier.

D.2 Does the GenAI-rewritten version truly enhance fundability?

The other group of human raters were shown 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 below, please respond with* ***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.*"

As shown in Table A7, loan descriptions rewritten by GenAI received higher funding intentions, indicating that the application of GenAI could increase loan funding rates, which is consistent with our previous findings.

Table A7: Human Rater Results

	DV: Writing Quality	DV: Willingness to Lend
Rewrite	**2.342*****	**1.450*****
	(0.223)	(0.249)
Constant	5.283***	4.500***
	(0.158)	(0.176)
Listing Fixed Effect	Yes	Yes
Rater Fixed Effect	Yes	Yes
Observations	240	240
Pseudo R^2	73%	72%

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

Appendix E Robustness to Prompt Design

Rule settings are relatively flexible and exhibit significant variation among users, due to differences in their abilities to use GenAI tools and their own cognitive levels. In this paper, we consider several common scenarios: (1) We examine the special case without a role setting, where borrowers directly use GenAI tools to rewrite their loan applications; (2) We assess the scenario where GenAI is adopted to enhance the readability and emphasizes content that could assist lenders in decision-making, termed "View 1"; (3) We assess the scenario where GenAI is adopted to make the reason compelling and improve the trustworthy of the loan description, termed "View 2"; (4) Building on existing research, we designed settings from the perspective of text features that have been proven to increase fundability, termed "View 3". Notably, each setting aims to improve writing quality to enhance fundability. Details are shown in Figure 13.

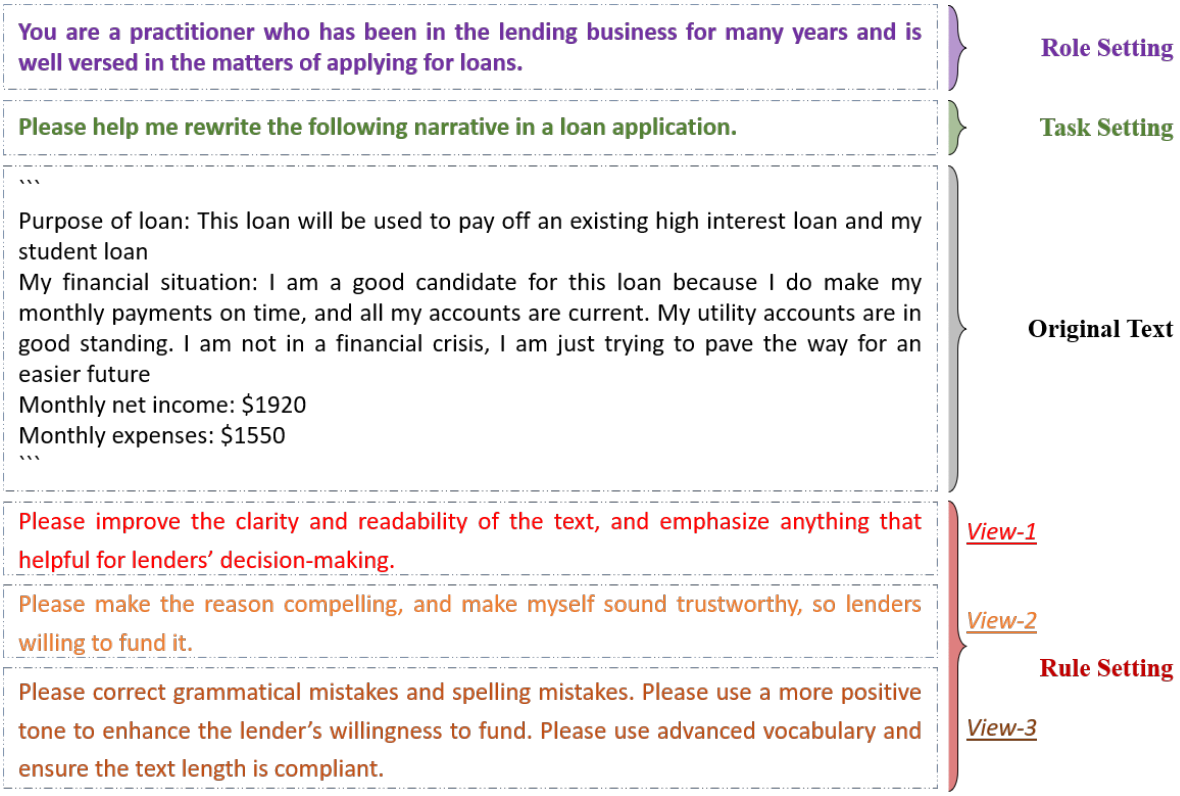


Figure 13: An Example ChatGPT Prompt Showing Different Levels of Specification.

To ensure the robustness of our results, we constructed several prompt templates, as illustrated in Table A8.[12]

[12] We opt to progressively broaden the rule settings stems from the fact that, with a deeper understanding of GenAI,

Table A8: Prompt Templates Based on Different Settings

Prompt Template	Role Setting	Task Setting	Rule Setting
1	✓	✓	None
2	✓	✓	"View 1"
3	✓	✓	"View 1 + View 2"
4	✓	✓	"View 1 + View 2 + View 3"

In the following, we use P1R through P4R to represent results corresponding to the rewrites produced by applying prompt templates 1 through 4.

E.1 Prompt Design and Text Features

Based on the previous setup, we tested the changes in text features across different prompt templates, as presented in Table A9.

Table A9: Textual Features Before and After Rewrite

	Original	P1R		P2R		P3R		P4R	
	Mean (SD)	Mean (SD)	DIF* (%)	Mean (SD)	DIF* (%)	Mean (SD)	DIF* (%)	Mean (SD)	DIF* (%)
Text Length	202.5 (132.2)	240.2 (109.7)	37.7*** (+18.6%)	248.1 (99.35)	45.6*** (+22.5%)	292.5 (90.54)	90.0*** (+44.4%)	281.1 (99.44)	78.6*** (+38.8%)
Spelling Error	0.035 (0.021)	0.009 (0.008)	-0.026*** (-74.3%)	0.009 (0.009)	-0.026*** (-74.3%)	0.008 (0.007)	-0.027*** (-77.1%)	0.009 (0.007)	-0.026*** (-74.3%)
Grammatical Error	0.064 (0.041)	0.035 (0.030)	-0.029*** (-45.3%)	0.038 (0.030)	-0.026*** (-40.6%)	0.026 (0.021)	-0.038*** (-59.4%)	0.028 (0.023)	-0.036*** (-56.3%)
Formality	0.239 (0.058)	0.342 (0.042)	0.103*** (+43.1%)	0.345 (0.039)	0.106*** (+44.4%)	0.349 (0.033)	0.110*** (+46.0%)	0.367 (0.035)	0.128*** (+53.6%)
Readability	10.45 (3.775)	14.71 (2.664)	4.26*** (+40.8%)	14.78 (2.115)	4.33*** (+41.4%)	15.26 (1.502)	4.81*** (+46.0%)	15.75 (1.684)	5.30*** (+50.7%)
Tone	0.745 (0.157)	0.773 (0.146)	0.028*** (+3.8%)	0.775 (0.145)	0.030*** (+4.0%)	0.753*** (0.144)	0.008*** (+1.1%)	0.754 (0.144)	0.009*** (+1.2%)

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

Table A9 indicates that while GenAI's impact on text features varies across different prompt templates, the overall trend remains consistent. Specifically, GenAI adoption enhances length compliance, formality, and readability, while reducing grammatical and spelling errors. Furthermore, the tone of the rewritten text becomes slightly more positive. These results are consistent with the findings in Section 3.1.

users' techniques and abilities will consistently advance.

E.2 Prompt Design and WQI

Since the WQI is calculated based on the cosine similarity of loan descriptions before and after GenAI rewriting, we analyzed the WQI under different prompt templates, with the results shown in Table A10.

Table A10: WQI by Credit Level and Funding Outcome

	P1R			P2R			P3R			P4R		
	U	D	R	U	D	R	U	D	R	U	D	R
High	0.736	0.771	0.769	0.745	0.774	0.772	0.734	0.768	0.763	0.726	0.764	0.759
	(0.009)	(0.006)	(0.007)	(0.008)	(0.006)	(0.006)	(0.008)	(0.006)	(0.006)	(0.009)	(0.006)	(0.006)
Medium	0.790	0.786	0.783	0.790	0.788	0.787	0.779	0.776	0.773	0.779	0.773	0.770
	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.005)	(0.007)	(0.007)	(0.005)	(0.006)	(0.007)
Low	0.796	0.795	0.788	0.799	0.798	0.792	0.803	0.790	0.781	0.790	0.784	0.776
	(0.004)	(0.008)	(0.009)	(0.004)	(0.007)	(0.008)	(0.004)	(0.008)	(0.009)	(0.004)	(0.008)	(0.009)

[a] U: Unfunded Loans; D: Default Loans; R: Repaid Loans.
[b] Values in parentheses represent standard deviations.

From Table A10, we observe that although there are differences in prompt templates, the variation in WQI across different different groups remains consistent. Specifically, we find that low-credit borrowers have higher WQI, while high-credit borrowers have lower WQI, with this difference being particularly noticeable for unfunded loans. Additionally, we observe that funded high-credit borrowers have significantly higher WQI compared to their unfunded counterparts, which implies that some high-credit borrowers may undervalue the importance of writing loan descriptions. Finally, we find that within funded loans, the WQI difference between repay and default loan statuses is relatively small, which potentially suggests that WQI is more influenced by funding status.

Appendix F Technical Information

F.1 Architectures for Text Embeddings

F.1.1 BERT

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

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

The pre-train process of BERT relies on two key techniques: Masked Language Model (MLM) and Next Sentence Predict (NSP). The MLM sets the objective to maximize the likelihood of the masked tokens. For a masked token, w_m, the objective can be formulated as:

$$L_{MLM} = -\sum_{i \in M} \log P(w_m | context) \tag{10}$$

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

$$L_{NSP} = -(y \log(P)) + (1 - y) \log(1 - P) \tag{11}$$

where, L_{NSP} is 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).

F.1.2 RoBERTa

Robustly Optimized BERT Pre-training Approach(RoBERTa), is a machine learning model for natural language processing (NLP) tasks. It was developed by Facebook AI researchers and is based on Google's BERT (Bidirectional Encoder Representations from Transformers) model. RoBERTa modifies key hyperparameters in BERT, including removing the next-sentence pre-training objective and training with much larger mini-batches and learning rates.

Key features and improvements of RoBERTa include the following. Training Data and Duration: RoBERTa is trained on a larger dataset and for a longer time compared to BERT. This leads to improved performance on various NLP benchmarks. Dynamic Masking: Unlike BERT, which generates a static mask before training, RoBERTa applies dynamic masking where the masked tokens change during the training process. Hyperparameter Tuning: It uses different batch sizes and learning rates, which contribute to its robustness and efficiency. No Next-Sentence Prediction (NSP): RoBERTa removes the NSP task used in BERT, focusing solely on the masked language modeling task. byte-Pair Encoding (BPE): RoBERTa uses BPE as its tokenization mechanism, allowing for more efficient handling of rare words.

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

$$L = \sum_{i=1}^{N} \max(0, \alpha + d(a_i, p_i) - d(a_i, n_i)), \tag{12}$$

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 i-th triplet, respectively.

F.2 Multimodal Transformer

To make predictions using both tabular features and text features, we leverage a multimodal transformer model, illustrated in Figure 14.[13] 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.

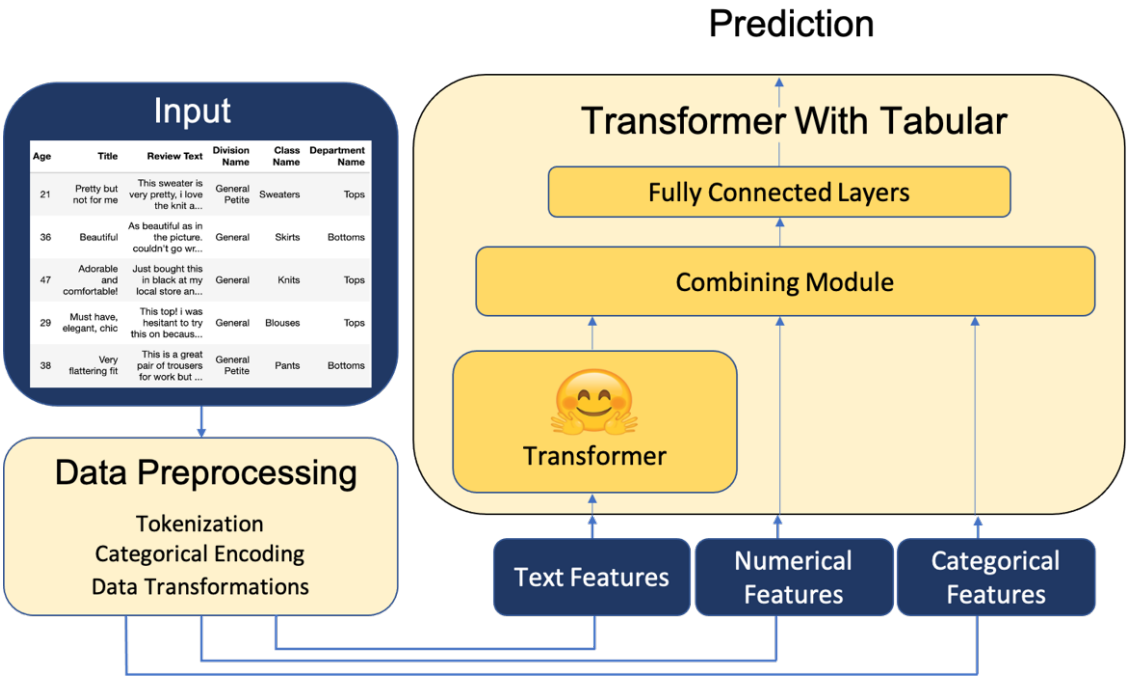


Figure 14: Architecture of the Multimodal Transformer (Gu and Budhkar, 2021)

F.3 Performance Benchmarking with Traditional Classifiers

Following Netzer et al. (2019), we consider regularized logistic regressions (GLM), random forest (RF), and extremely randomized trees (ERT), as well as an emsemble model consisting of these base models. More detailed description of these methods can be found in Netzer et al. (2019). We also include light gradient-boosting machine (LGBM), an algorithm renowned for its effectiveness in loan evaluation (Zhang et al., 2023), as one of the base models. For machine learning classifiers. We use tabular features, explicit text features, and bigram features in a setting similar to Netzer et al. (2019). The Ensemble Model was constructed by stacking the above models, each with an using the aforementioned models.

[13] https://github.com/georgian-io/Multimodal-Toolkit.

Table A11: Performance Benchmarking by Predicting Default Among Funded Loans

	AUC[c]			
	Model 1 (Tabu Only)	Model 2 (Text Only)	Model 3 (Tabu+Text)	Soft Information
Panel A: machine learning models[a]				
Ensemble	**72.5%**	64.6%	73.7%	1.2%
GLM Only	71.4%	63.8%	71.4%	0.0%
RF Only	72.3%	63.0%	73.0%	0.7%
LGBM Only	72.3%	63.6%	73.3%	1.0%
ERT Only	71.7%	63.0%	72.3%	0.6%
Panel B: deep learning models[b]				
BERT	-	69.6%	74.7%	2.2%
RoBERTa	-	**71.0%**	**75.8%**	**3.3%**

[a] Ensemble = an Ensemble model of all; GLM = Generalized linear model; RF = Random forest; LGBM = Light gradient boosting machine; ERT = Extremely randomized trees (extra trees).
[b] BERT and RoBERTa are designed for text data, therefore, they do not have a result for Model 1. While BERT and RoBERTa are large language models designed for text, when we combine them with tabular data, we employed the multimodal transformer architecture (Gu and Budhkar, 2021).
[c] Note: All AUCs are averaged across 10 replications of the ten-fold mean.

In online credit markets, Netzer et al. (2019) applied several advanced machine learning models and AUC to evaluate the informational value of text data. Building on their research, we used both machine learning models, such as generalized linear models (GLM) and random forests (RF), and deep learning models, including BERT and RoBERTa, to evaluate the value of soft information in default prediction within online credit markets. The results are shown in Table A11. For these machine learning models, soft information is captured through bi-grams, whereas in deep learning models (Netzer et al., 2019), it is obtained through text embeddings (Vaswani et al., 2017).

In Table A11, Model 1 is based solely on tabular features, Model 2 uses only text, while Model 3 combines both tabular features and text as predictors. For BERT and RoBERTa, since they do not have a Model 1 result, we use the best performance of machine learning models (i.e., ensemble) to impute it. The value of soft information is quantified by the performance variation from Model 1 to Model 3. In general, soft information improved the performance of loan default prediction models, although there were variations across different credit levels. This pattern confirms that text has predictive value, which is even larger when using deep learning models.

F.4 Hyperparameters

In this section, we document the hyperparameters and optimization methods of the machine learning models involved in our study. This final estimator is trained through cross-validation. We employ Bayesian methods for the optimization of hyperparameters. All these models were implemented using sklearn, a highly renowned Python package extensively used in the field of machine learning.[14]

Table A12: Hyperparameter Settings for Classifiers

Model	Parameter	Description	Range
GLM	C	Inverse of regularization strength; must be a positive float. Smaller values specify stronger regularization.	(0.01, 10)
	penalty	Specify the norm of the penalty.	['l1', 'l2']
	solver	Algorithm to use in the optimization problem.	[0, 1]
Tree-Based[a]	n_estimators	The number of trees in the forest.	(10, 1000)
	max_depth	The maximum depth of the tree.	(5, 30)
	max_features	The number of features to consider when looking for the best split.	(0.1, 0.999)
	min_samples_leaf	The 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	n_estimators	The number of trees.	(10, 1000)
	max_depth	The maximum depth of the tree.	(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	The number of leaves in a tree.	(24, 45)
DL[a]	learning_rate_text	The learning rate of the pre-trained language models.	[3e-6]
	learning_rate_tabu	The learning rate of the MLP model, which employed to learn the features of tabular 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	Train epochs in each round.	[6, 7, 8]
	round	Train round for each group of parameters.	[10]
	batch_size	Batch size.	[15, 30]
	dropout_rate	The dropout rate of models.	[0.45]
	activate_func	Activate functions.	['relu','leaky_relu']

[a] Tree-based models include RF and ERT; deep learning (DL) models include BERT and RoBERTa.

When considering implicit text features such as BERT and RoBERTa, we directly add a multi-layer perceptron (MLP) layer with cross entropy loss to map numeric vector representations of text into the outcome class label. Due to random initialization, we ran the same model setting 16

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

times and report results based on the average. We implemented them with Pytorch[15] and Huggingface.[16] We adopted grid search to find the best parameters for BERT and RoBERTa. Details of the parameters are presented in Table A12.

[15] https://pytorch.org/
[16] https://huggingface.co/