Information Processing and Management 61 (2024) 103655

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Risk disclosure and entrepreneurial resource acquisition in crowdfunding digital platforms: Evidence from digital technology ventures

Hong Huo [a], Chen Wang [a], Chunjia Han [b,*], Mu Yang [b], Wen-Long Shang [c,d,e,**]

[a] School of Management, Harbin University of Commerce, Harbin 158020, China
[b] Department of Management, Birkbeck, University of London, London WC1E 7HX, United Kingdom
[c] Beijing Key Laboratory of Traffic Engineering, College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China
[d] School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100008, China
[e] Centre for Transport Studies, Imperial College London, London SW7 2AZ, United Kingdom

ARTICLE INFO

Keywords:
Crowdfunding
Digital entrepreneurship
Risk disclosure
Text mining
Digital technology ventures

ABSTRACT

The widespread development of digital technology facilitates the emergence of new entrepreneurial modes, of which crowdfunding digital platforms are one. In the digital environment of crowdfunding platforms, digital entrepreneurs can obtain the essential resources necessary for their startups' rapid and cost-effcient development. However, the information asymmetry derived from the digital nature of crowdfunding platforms leads to a lower chance of success for entrepreneurial ventures in this market, especially those in digital technology, limiting the important role that crowdfunding platforms can play in digital entrepreneurship. To this end, we focus on the risk disclosure section introduced by crowdfunding platforms to alleviate information asymmetry and explore the influence mechanism of the content of risk disclosure on entrepreneurial resource acquisition in crowdfunding digital platforms. By employing a novel text mining technique structural topic modelling, we analyse the risk disclosure texts of 4,284 digital technology crowdfunding projects and successfully identify various factors that constrain the development of digital technology ventures in crowdfunding platforms. Furthermore, we find that the risk topics digital entrepreneurs disclose negatively affect entrepreneurial resource acquisition. However, this relationship is moderated by the reward structure setting in the context of reward-based crowdfunding. The findings of this study not only enrich the literature on crowdfunding and digital entrepreneurship but also provide valuable practical implications on how crowdfunding digital platforms can be used to promote the development of digital entrepreneurship.

1. Introduction

Crowdfunding, as a medium for digital entrepreneurship, plays a pivotal role in the digital entrepreneurship system (Zheng et al., 2022). It serves as an alternative source of financing based on digital platforms and connects digital entrepreneurs with potential

* Corresponding author.
** Corresponding author at: Beijing Key Laboratory of Traffic Engineering, College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China.
E-mail addresses: chunjia.han@bbk.ac.uk (C. Han), shangwl_imperial@bjut.edu.cn (W.-L. Shang).

https://doi.org/10.1016/j.ipm.2024.103655
Received 2 March 2023; Received in revised form 9 October 2023; Accepted 8 January 2024
Available online 26 January 2024

public funders (also known as backers) from which digital entrepreneurs obtain the funding they need to help them fill the gap in venture financing (Nakagawa & Kosaka, 2022). In particular, this study concentrates on reward-based crowdfunding, which is based on a pre-sale agreement in which individual backers fund entrepreneurial projects in return for some product or service known as rewards (Xiao et al., 2021). In contrast to conventional entrepreneurial models, digital entrepreneurs encounter distinct circumstances when solicting contributions from potential backers on crowdfunding platforms (Nzembayie et al., 2019). This has given rise to novel entrepreneurial risks and challenges for crowdfunding ventures, especially those focused on cutting-edge digital technologies, as they strive to secure the necessary funding effectively. The proliferation of digital technology has led to a surge in the number of digital technology ventures utilising crowdfunding platforms, but this type of ventures often faces a relatively low successful chance. However, we have no idea about the risks and challenges that digital technology ventures may encounter on the crowdfunding platform. With the aim of motivating entrepreneurs to contemplate, deliberate, and ultimately address any potential issues that may emerge in their projects, Kickstarter implemented a new policy in September 2012. This policy mandates the inclusion of a "Risks and challenges" section on each project page, which discloses the risks that the venture may encounter and provides appropriate courses of action should they materialise. Consequently, this policy directly opens up an avenue for risk research in the realm of crowdfunding for digital technology ventures.

As a digital platform, crowdfunding platforms do not allow digital entrepreneurs and funders to interact with each other directly, which leads to information asymmetry (Costello & Lee, 2022). Information asymmetry makes potential backers face a high degree of ambiguity about the content and quality of the project (Jiang et al., 2020), therefore, discourages potential backers from contributing. In traditional financial markets, listed companies are often mandated by law and fiduciary regulations to disclose risk information to inform investors of potential risks, thereby helping to reduce information asymmetry (Xu & Zhang, 2013). In reward-based crowdfunding, while platforms are not subject to the appropriate regulations, some crowdfunding platforms have attempted to improve risk disclosure through the implementation of various strategies in order to keep the crowdfunding industry sustainable, with the introduction of the "Risks and challenges" section is one of the attempts by Kickstarter. However, it is not yet known how the information in the "Risks and challenges" will affect the pledging decisions of potential backers.

The information disclosed by digital entrepreneurs often covers different topics, which may have a different impact on fundraising outcomes (Jiang et al., 2020; Xiao et al., 2021). It is therefore reasonable to assume that the different topics in risk disclosure may also have a different impact on the crowdfunding performance. However, while previous research has clearly noted the necessity of collecting and analysing data in the "Risks and challenges" section used for project risk disclosure (Anglin et al., 2018; Yang et al., 2020), we do not know how the topics in risk disclosure affect crowdfunding performance. Moreover, in the reward-based crowdfunding model, setting the reward structure is crucial and will be an important factor in influencing backers' decisions (Yang et al., 2020). Therefore, to make the study more rigorous and in-depth, attention should also be paid to the impact of the number of optional reward tiers. Based on these research requirements, this study seeks to answer the research questions below:

RQ1: What are the main risk topics mentioned in the risk disclosure for digital technology ventures? What insights can we gain from them?
RQ2: How will topics in risk disclosure for digital technology ventures affect crowdfunding performance?
RQ3: How does the number of reward tiers affect the relationship between topics in risk disclosure and crowdfunding performance in a reward-based context?

Crowdfunding performance is commonly defined as the eventual outcome achieved by a crowdfunding project at the end of fundraising (Tian & Zhang, 2023; Wang et al., 2022). Ahlers et al. (2015) suggest that crowdfunding performance is a multifaceted concept that can be measured by various indicators, including fundraising success, the fundraising ratio, the amount raised, and the number of backers etc., which have also been widely used in existing studies (Allison et al., 2017; Xu, 2018). In this study, we define crowdfunding performance as the total amount of funds raised by the project and the total number of backers who funded the project. Because this can capture two aspects of the outcome of a crowdfunding project at the same time, the total amount of funds raised can reflect the final financial outcome of a project, while the number of backers can usually reflect the popular outcome of a project.

Furthermore, a big challenge in answering the above research question is deriving meaningful information from the vast scale of unstructured text data. The rapid development of text mining techniques, such as machine learning, addresses this challenge, allowing us to gain insights and important findings from this unstructured data with good speed and accuracy (Hashem et al., 2016). More specifically, we analyse the risk disclosure texts of 4,284 digital technology ventures launched on the Kickstarter platform between August 2018 and August 2022 by employing an unsupervised machine learning method, Structural Topic Modelling (STM). After extracting the main topics, we further explored the main effects of these risk information related topics on crowdfunding performance and the moderating effects of reward tiers through regression analysis.

Several interesting findings were uncovered in this study. Firstly, our STM model successfully extracts a number of topics about the risks and challenges that may confront initiating a digital technology venture on a crowdfunding platform. We prove the viability and efficacy of employing the STM model to derive meaningful insights from large volumes of unstructured text data from crowdfunding platforms. Secondly, our findings indicate that risk topics disclosed by digital entrepreneurs can have a negative impact on the amount of funds raised and the number of backers, including those related to how to respond to risks and challenges. On the positive side, however, we provide new evidence that the risk disclosure content contributes to alleviating information asymmetry in the crowdfunding market. Thirdly, as we suspected, the number of reward tiers significantly moderates the relationship between risk topics and crowdfunding performance in a reward-based context. The adverse impact of risk topics is mitigated if the creator sets more optional reward tiers. Fourthly, by observing changes in the proportions of these topics over time, we find that the topics mentioned by entrepreneurs in risk disclosure are constantly changing, for example, the probability of campaign or product delays increasing significantly after the COVID-19 explosion. Beyond the academic contribution, this research also provides valuable practical implications. By examining the impact of risk information on crowdfunding performance and the moderating role of reward structure, we provide useful guidance for digital technology ventures and platform design, and offer seemingly sensible suggestions for solving the financing problem in digital entrepreneurship.

2. Literature review and hypotheses development

2.1. Information asymmetry and signal theory

Due to the digital nature of crowdfunding platforms, there can easily be an information asymmetry between entrepreneurs and potential backers (Sannajust et al., 2014). On the one hand, entrepreneurs can manage the flow and disclosure of information (Bai et al., 2022; Thies et al., 2016), and to prevent the possibility of imitation and replication by competitors, some entrepreneurs may withhold detailed information about their projects. On the other hand, most potential backers in crowdfunding platforms do not have extensive investment experience, and they cannot adequately investigate detailed information about the project or assess its investment value (Jiang et al., 2020). As a result, information asymmetry is considered among the most essential aspects influencing the outcome of crowdfunding (Chakraborty & Swinney, 2021).

Signal theory is built around the concept of information asymmetry. The primary principle of the theory is to lessen the information asymmetry between the parties (Akerlof, 1978; Spence, 1978). The signal theory substantially explains transactional behaviour in crowdfunding (Courtney et al., 2017) since signals about projects can effectively reduce information asymmetry between entrepreneurs and backers (Kirmani & Rao, 2000). For example, previous research discovered that crowdfunding project creators who convey sustainable value signals (Siebeneicher & Bock, 2022), project quality signals (Tajvarpour & Pujari, 2022), and rhetorical signals (Anglin et al., 2018) in their project texts can help improve crowdfunding performance. Crowdfunding platforms such as Kickstarter offer various sections to convey project signals to help overcome information asymmetry between entrepreneurs and backers (Zheng et al., 2016). The "Risks and challenges" section, for example, contains information on the risks and challenges associated with the project as well as the plans and commitments of the entrepreneurs to address them, and these risk signals are also expected to impact crowdfunding performance.

2.2. Text information in crowdfunding and disruptive text mining technology

Text information is an important medium for entrepreneurs to convey signals to potential backers in crowdfunding platforms (Jiang et al., 2020). Entrepreneurs can use text descriptions to disclose information, present project details, and demonstrate their projects' investment value to help crowdfunding succeed. Several studies have focused on the impact of textual information on crowdfunding performance, including project abstract (Westerlund et al., 2021), project description (Chan et al., 2021), reward description (Jiang et al., 2021), and project update information (Xiao et al., 2021). Furthermore, text information is viewed as an important source of insight (Berger et al., 2020; Yang et al., 2022). For example, based on the online self-introduction of female entrepreneurs, Wang et al. (2022) successfully identified the four main digital identities presented by female entrepreneurs on crowdfunding platforms. Lee and Sohn (2019) found that the use of smart services technology is rapidly emerging globally by mining project text data in the software category.

In order to process and analyse large-scale unstructured text data, a range of disruptive text mining techniques has been developed and adopted to benefit many research areas. For example, Cai et al. (2023) explored the impact of social robots on the mechanisms of information dissemination in social networks based on machine learning methodology. Shah et al. (2021) mine online doctor reviews based on STM to analyse the driving factors of patient satisfaction and dissatisfaction. Disruptive text mining techniques are also crucial for research related to crowdfunding texts, as they can provide fast, effective, and low-cost analysis. To examine the textual data of crowdfunding projects, for instance, one of the most widely used topic modelling techniques, Latent Dirichlet Allocation (LDA), has been implemented (Jiang et al., 2021; Wang et al., 2022; Xiao et al., 2021).

2.3. Disclosure of risk information

Risk disclosure is a standard mechanism firms adopt to decrease information asymmetry and enhance the efficiency of the capital markets (Liu et al., 2022; Madsen & McMullin, 2020). In traditional financial markets, laws and fiduciary regulations enforce listed firms to implement risk disclosure, ensuring that individual investors have access to appropriate information about potential risks and avoiding excessive risky investment behaviour (Xu & Zhang, 2013). For example, Kim et al. (2022) found that increased project risk awareness is often associated with poor financing outcomes for crowdfunding projects. In reward-based crowdfunding, however, entrepreneurs can subjectively and voluntarily disclose the potential risks and challenges of their projects, and the platform will not assess or check the information disclosed. Potential backers may develop multiple understandings of such subjective and unverifiable disclosures. Some backers may choose to be risk-averse, while others may develop a greater sense of trust because the information reduces the uncertainty of the project (Madsen & McMullin, 2020). For example, Michels (2012) has demonstrated that unverifiable disclosures lead to greater participation in the peer-to-peer lending market. These varied results challenge forecasting how potential backers will react to the content of the risk information disclosed about the crowdfunding project.

While several recent studies have explored the impact of project risk salience on backer funding decisions (Kim et al., 2022; Madsen

& McMullin, 2020), these studies have not explored specific topics in risk disclosure and the impact of these topics on crowdfunding outcomes in the context of digital entrepreneurship. Previous crowdfunding studies have examined the key role of topics in the project text (Jiang et al., 2020; Westerlund et al., 2021; Xiao et al., 2021), and the findings reveal that different topic content in a crowdfunding text may influence backers' contributions differently. Based on the above research evidence, we hypothesise that different risk disclosure topics may impact crowdfunding performance differently. Specifically, topics that relate to the risks faced by the project may increase the risk perceptions of potential investors, resulting in lower crowdfunding performance, while topics that relate to the plans and commitments made by the entrepreneur to address the risks may increase the credibility and persuasiveness of the project, leading to a positive impact on crowdfunding performance. We therefore hypothesise:

H1a. In the risk disclosure, topics that relate to the risks faced by the project will negatively impact crowdfunding performance.
H1b. In the risk disclosure, topics that relate to the entrepreneur's response to risk will positively impact crowdfunding performance.

2.4. The moderating effect of reward tiers

Reward-based crowdfunding has a unique feature in that entrepreneurs can offer several rewards to backers, typically representing products or services that can be offered if the venture successfully funds its fundraising efforts (Kunz et al., 2017). In particular, the rewards given by entrepreneurs are generally structured in a tiered form, with the entrepreneur offering a fully functional product or service in return for a larger pledge and another product or service with limited functionality in exchange for a smaller pledge (Xiao et al., 2021). Several studies have found that the likelihood of project success increases with the number of optional reward tiers (Jiang et al., 2021; Kunz et al., 2017). Studies have also demonstrated that the number of optional reward tiers affects the decision-making process of backers in a wide range of ways. For example, Frydrych et al. (2014) argue that the setting of the reward structure serves as a signal that may indicate the preparation performed by the project creator, entrepreneurs who offer more reward options are likely to be perceived as more serious and reliable by backers. In addition, potential backers have a degree of heterogeneity, with different views and levels of perception of the risk information disclosed about the project (Xiao et al., 2021). An increase in reward tiers enhances the freedom of backers, who have a wider chance of finding the option that best suits their needs, thus increasing their intrinsic motivation (Kunz et al., 2017). Based on the above evidence, we examine the moderating effect of the number of optional reward tiers. We propose that the increase in the number of reward tiers will help reduce the perception of uncertainty in the project by backers and understand the risk information of the project in a more positive way. We therefore hypothesise:

H2: The number of reward tiers positively moderates the relationship between topics in risk disclosure and crowdfunding performance, i.e., a greater number of reward tiers weakens the negative impact of topics on crowdfunding performance and enhances the positive impact of topics on crowdfunding performance.

3. Data and methodology

Using unsupervised machine learning approach, this study examines ventures related to digital technology on the rewards-based crowdfunding platform Kickstarter, extracting topics from the risk disclosure text in order to gain insights hidden in unstructured big data and to explore the impact of the topics in risk disclosure on crowdfunding performance. Specifically, we crawl data from the Kickstarter website from August 2018 to August 2022 and then clean it up into an experimental dataset. Next, an unsupervised machine learning technique, STM, is employed to extract topics from the text. Finally, ordinary least squares (OLS) regression is used to investigate the relationship between each of these topics and crowdfunding performance and the moderating effects of reward tiers.

3.1. Data crawling and processing

Similar to other recent studies (Chan et al., 2021; Siebeneicher & Bock, 2022),we utilise data from the rewards-based crowdfunding platform Kickstarter, a popular data source for research using text analytics (Sandouka, 2019). Since 2009, over 21 million funders have supported over $7 billion in projects, successfully funding over 230,000 projects (Kickstarter, 2022). The Kickstarter platform comprises an enormous number of similarly structured projects in a digital setting that are all subject to the same rules, allowing us to examine the interactions between individuals in a regulated setting (Chan et al., 2021). The relevant information from these interactions finally constitutes the dataset that enables us to directly uncover potential information in the projects and measure how the topics in risk disclosure and the reward tiers affect potential backer engagement in crowdfunding.

To collect a suitable sample of data from the Kickstarter platform, we first collect URLs from webrobots website (https://webrobots.io/kickstarter-datasets/) for all projects launched on Kickstarter between August 2018 and August 2022, removing duplicate projects from the list. We select sub-categories related to digital technology under the technology category, which includ 3D Printing, Apps, Robots, Software, and Web, and keep only the successful or failed projects. A total of 4,284 projects are collected, 3,560 failed and 724 successful. The success rate of 16.9% is much lower than the average success rate of 40.39% for all categories on Kickstarter (Kickstarter, 2022), which proves that digital technology ventures often have a lower likelihood of success on crowdfunding platforms. The data columns include project title, subcategory, creation date, risks disclosure text, URL, duration, funding goal, amount raised, number of backers, number of optional reward tiers, number of images, etc. Finally, we perform text pre-processing work using the tidytext package (Silge & Robinson, 2016) in the R programming. This process includes character deletion (removing non-English text, URLs, numbers, special symbols and stop words), normalisation (converting all letters to lower case) and stemming (removing word affixes), which is consistent with the processing of previous similar topic extraction research (Han et al., 2021; Wang et al., 2022). The final dataset includes 4,143 projects, 3,465 failed and 678 successful.

3.2. Topic extraction based on STM

Topic modelling, a natural language processing (NLP) approach integrated with unsupervised machine learning, detects potential topics in a text corpus by examining co-occurrence relationships between words (Xu, 2018). A range of topic modelling techniques, such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003), have been used to analyse textual information about crowdfunding projects and have proven to perform effectively in topic analysis (Nakagawa et al., 2022; Xiao et al., 2021). Based on a comparison of techniques widely used in previous text analysis studies, we selected STM (Roberts et al., 2014) to calculate the probability of topic composition in risk disclosure text and to analyse trends of topics embedded in unstructured text data. STM is a technique similar to LDA but allows for the metadata incorporated into the data sample to explain the prevalence of topics (Roberts et al., 2016). This technique has emerged in the analysis of online reviews (Ding et al., 2020), social media data (Han et al., 2021) and speech reports (Savin & Teplyakov, 2022).

Specifically, STM is selected for this study for three main reasons. Firstly, STM is a generative probabilistic model that reduces subjective bias by automatically extracting topics from the corpus. The output of the STM provides not only top words related to topics but also the percentage distribution of different topics in each risk disclosure text, which allows us to explore how these risk topics impact crowdfunding performance. Using STM can greatly improve the interpretability of the crowdfunding risk disclosure text in our study. Secondly, an improvement of STM over other topic modelling techniques, such as LDA, is its ability to directly estimate the impact of document metadata on topic prevalence (Büschken & Allenby, 2016). By employing STM, we can easily determine how the proportion of different risk topics disclosed by entrepreneurs varies over time, allowing us to monitor the "dynamics" of risk disclosure information about digital technology ventures more effectively, which aligns with our research objective. Thirdly, due to the rich variety of datasets used in our study, taking metadata as covariates in the topic estimation process may enhance the fit of the resulting model to the input data (Schmiedel et al., 2019), enabling us to identify risk topics disclosed by entrepreneurs in greater detail and with greater accuracy from a large amount of textual content.

In STM, a document is defined as a mixture of topics, meaning that the document is composed of multiple topics, and each topic represents a probability distribution of many words (Roberts et al., 2014). In this study, we consider the textual description of the risk disclosure as a document, and the stm package in the R programming language (Roberts et al., 2019) was used to model our analysis, where the project risk disclosure text content was input as a document, the prevalence function is set as follows:

$$prevalence \sim state + s(project\,created\,date) \tag{1}$$

where s is the smooth function of time, and state indicates the funding state of the project (i.e., successful or failed), state equals 1 if the project is funded successfully, and 0 otherwise.

Although STM is an unsupervised text analysis technique, it is necessary to determine the number of topics K, which is one of STM's most crucial user-specified parameters and contributes to achieving a substantive interpretation of the modelling results (Li et al., 2013). We use the searchK function from the stm package in R programming to test models trained on a sparse matrix in parallel with a range of different K values, and the optimal number of topics for our dataset is found in the range of 3 to 20 topics. We statistically assess the model output based on the training models using three criteria: semantic coherence, held-out likelihood, and residuals, Fig. 1 reports the detailed diagnostic values of the three criteria for different topics. Higher values of semantic coherence and held-out likelihood and lower values of residuals will result in better model performance (Roberts et al., 2019, 2014). As shown in Fig. 1, no model outperforms the others absolutely, as the number of topics increases, semantic coherence and residuals typically decrease, while held-out likelihood increases. Hence, a model with 10 topics that can produce relatively decent results in all three aspects is selected for this study. With $K = 10$, semantic coherence is locally maximal at -89.79, held-out likelihood reaches a high value of -6.40, and residuals at a relatively low value of 2.24.

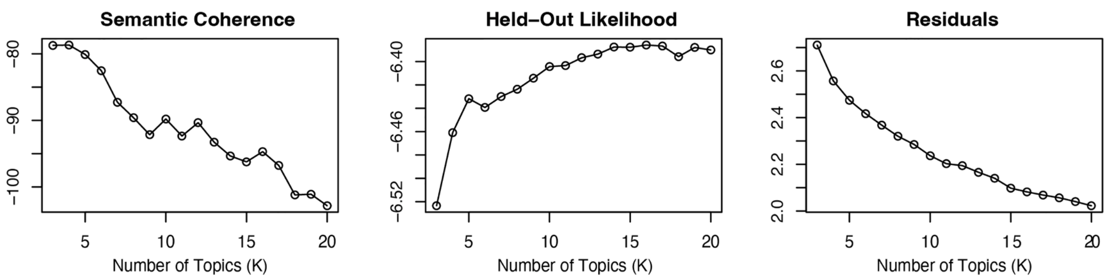


Fig. 1. Diagnostic values by number of topics.

3.3. Measures

3.3.1. Dependent variable: crowdfunding performance

Considering that 3,560 of the 4,284 projects we collected failed in crowdfunding, focusing solely on whether the project met its funding target may have led us to miss some important information (Ba et al., 2021). In addition, since the target funds was controlled in this study, the fundraising ratio (funds raised/target funds) was not selected as the dependent variable. Eventually, after considering and referring to previous crowdfunding studies, we measured crowdfunding performance using the total amount of funds raised (*Raised*) and the number of backers (*Backers*).

3.3.2. Independent and moderator variables

The probability of each topic in the risk disclosure for each project, as determined using the STM model, was used as the independent variable. Regarding our moderator variable, "Reward tiers", we exploit the fact that the number of optional reward tiers is crucial in reward-based crowdfunding and that they influence the decision-making process of crowdfunding backers in multiple potential ways (Yang et al., 2020). Based on this evidence, we examine the number of optional reward tiers for each project to explore whether it could help reduce backers' perceptions of project uncertainty and understand the risk information provided by entrepreneurs in a more positive way.

3.3.3. Control variables

In line with earlier studies on crowdfunding performance (Jiang et al., 2021; Mollick, 2014; Siebeneicher & Bock, 2022), our regression analysis model incorporates various project-level factors that have been found to have an impact on crowdfunding success, to exclude potential confounding effects, and we log-transformed some of these control variables as we found them to be significantly skewed (Calic & Mosakowski, 2016). Our control variables include the funding goal (*Goal*), the duration of the fundraising (*Duration*), the number of images (*Images*, log-transformed), whether the project has a video exhibition (*Video exhibition*), the number of comments (*Comments*, log-transformed), the number of updates (*Updates*), whether is marked as "Loved Project" by Kickstarter (*Loved project*), and whether is launched in the United States (*USA*). Furthermore, we also control the total number of words in the "Risks and challenges" description (*Words in R&C*, log-transformed), as existing research has demonstrated that word count contributes to the perception of project quality (Parhankangas & Renko, 2017). Finally, we also create dummy variables for the specific year in which the project was created and the subcategory to which the project belongs to control the potential impact of the time and the type of project. Table 1 summarises the description and measures of all the variables used in this study.

3.3.4. Data analysis model

Regarding the data analysis model, we conducted an OLS regression analysis, which has been widely used in existing studies on crowdfunding performance (Chan et al., 2021; Kim et al., 2022). Our dependent variables, which included the amount raised (log--transformed) and the number of backers (log-transformed), were both non-negative continuous variables. Independent and dependent variables revealed a linear relationship upon testing, whereas the residuals showed a normal distribution. In this context, OLS regression served as our best option for estimating the quantitative impact of topics in risk disclosure on crowdfunding performance and the moderating effect of reward tiers. Based on the research hypothesis proposed, we constructed the model as follows:

$$\log(Performance)_i = \beta_0 + \beta_1(Topic1_i : Topic10_i) + \beta_2(Rewardtiers_i) + \beta_3[(Rewardtiers_i) \times (Topic1_i : Topic10_i)] + \beta_4(Controls_i) + \varepsilon_i \quad (2)$$

Table 1
Summary of variables description and measures.

Variable	Description	Value
Dependent variables		
Raised	Total amount of funds ultimately raised by the project (in USD).	Continuous variable
Backers	Total number of backers pledging to the project.	Continuous variable
Independent variables		
Topic	Topic probabilities for each topic in the risk disclosure text.	Continuous variable
Moderator variables		
Reward tiers	Number of different reward tiers offered.	Continuous variable
Control variables		
Goal	Funding goal (in USD).	Continuous variable
Duration	The duration of the fundraising for the project (in days).	Continuous variable
Images	The number of images used in the project introduction.	Continuous variable
Video exhibition	Whether the project page has a video exhibition.	Yes = 1, No = 0
Comments	The number of comments posted by backers during the funding period.	Continuous variable
Updates	The number of project updates posted by creators during the funding period.	Continuous variable
Words in R&C	Total number of words in the "Risks and challenges" description.	Continuous variable
Loved Project	Whether the project has been selected as a "Loved Project" by Kickstarter.	Yes = 1, No = 0
USA	Whether the project is initiated in the USA.	Yes = 1, No = 0
Year	The year in which the project was initiated on Kickstarter.	Dummy variable
Subcategory	Subcategory to which the project belongs.	Dummy variable

where β_0 is the intercept term, β_j ($j = 1, 2, 3, 4$) is the regression coefficient, i represents the project, and ε_i is the residual term.

4. Results and analysis

4.1. Topics identified in the risk disclosure

Our STM model approach successfully identifies ten distinct and internally consistent topics from the risk disclosure description of the project, the obtained topics show the risks and challenges digital technology ventures may face on the crowdfunding platform. The results are presented in Table 2. It is important to note that the second and third columns of the table are the output of the STM model, where the top words are those with the highest probability of appearing in the topic but the lowest probability of appearing in other topics. The first column is the topic labels manually assigned by the two researchers. We followed the approach of Schmiedel et al. (2019), where two researchers coded the labels for each topic independently by scrutinising and analysing the ranking and meaning of the top words for each topic. In the case of different label assignments between the two researchers, a thorough discussion and examination are carried out until a consensus is reached on each label.

We then categorised the ten topics. Firstly, we determined four categories, namely external risks and challenges, internal risks and challenges, response to risks and challenges, and possible consequences, based on the enterprise risk management framework proposed by COSO (COSO, 2004), which identifies "risk identification, assessment, response, and evaluation of management effectiveness" as the main components of enterprise risk management. The four categories were also determined by being referenced to the categories of risk topics summarised by Costello and Lee (2022). Next, our two researchers manually coded the ten topics into four different categories. Specifically, the two coders were requested to categorise the topics by the semantics of the topic labels and representative examples without interfering with each other, and they were required to be very confident of the match before putting the topic into a category (Short et al., 2010) After coding, we applied Holsti's formula ($PA_O = 2A \ / \ (n_A + n_B)$, with the agreement rate PA_O, A congruent categorizations, n_A and n_B total categorisations by both coders) to test reliability between coders (Holsti, 1969), and the results showed an inter-coder reliability score of 0.89, which is greater than 0.70, indicating a high reliability of the topic classification results (Short et al., 2010). Finally, the two coders reached a consensus on the classification of all ten topics after sufficient discussion and validation. Table 2 shows the results of the classification of topics.

It is observable that most of the risks and challenges faced by digital technology ventures are derived from external factors, as demonstrated by the top two topics, namely "Funding issues" (15.78%) and "Ideas not understood" (14.35%) accounting for almost one-third of all topics. The topic "Funding issues" revolves around the fact that digital technology ventures need a lot of funding to start and entrepreneurs face challenges in raising enough funds. On the other hand, the topic "Ideas not understood" mainly expresses the concern of digital entrepreneurs that their entrepreneurial ideas are not understood and accepted by people. Examples of the two topics are as follows.

[Funding issues] Right now I have a team of developers waiting to build the app. The only problem is funding. To make the app it will cost around $75,000…

[Ideas not understood] The only real danger for this project is that no one will like the idea and everyone but me thinks it's stupid…

The remaining two topics in the external risks and challenges category are "Market competition" (11.48%) and "Political and legal environment" (4.10%), which focus on the competition that digital technology ventures may face in the market and the constraints of different policies and regulations in different regions and countries.

The second largest topic category is internal risks and challenges, with three topics classified in this category: "Application development" (10.57%), "Data collection/complex algorithms" (6.78%), and "Design /production" (6.70%), which focus on the risks and challenges that entrepreneurs may face in the design, development, testing and production of their products. For example, the largest topic under this topic category, "Application development", is mainly related to the problems that may be encountered during the development of software or applications, such as the creation of new features, version upgrades, and whether the Appstore can

Table 2
Topic summary in risks and challenges information disclosure.

Topic label	Topic Proportions	Top words
Topic Category: External risks and challenges		
Topic 5: Funding issues	15.78%	fund, rais, need, develop, market, help, enough
Topic 1: Ideas not understood	14.35%	peopl, want, know, like, mani, challeng, way
Topic 3: Market competition	11.48%	user, busi, platform, social, market, servic, competit
Topic 10: Political and legal environment	4.10%	ensur, secur, event, legal, concern, reduc, area
Topic Category: Internal risks and challenges		
Topic 6: Application development	10.57%	app, develop, launch, version, releas, date, featur,
Topic 9: Data collection/complex algorithms	6.78%	data, compani, technolog, industri, user, solut, code
Topic 8: Design/production	6.70%	test, design, part, time, work, print, produc
Topic Category: Response to risks and challenges		
Topic 2: Confidence to overcome challenges	12.83%	risk, project, challeng, develop, face, biggest, overcom
Topic 7: Problem solving for customers	5.56%	custom, problem, provid, inform, issu, servic, solv
Topic Category: Possible consequences		
Topic 4: Campaign or product delays	11.85%	product, delay, time, deliv, confid, kickstart, campaign

successfully support them. An example of this topic is as follows.

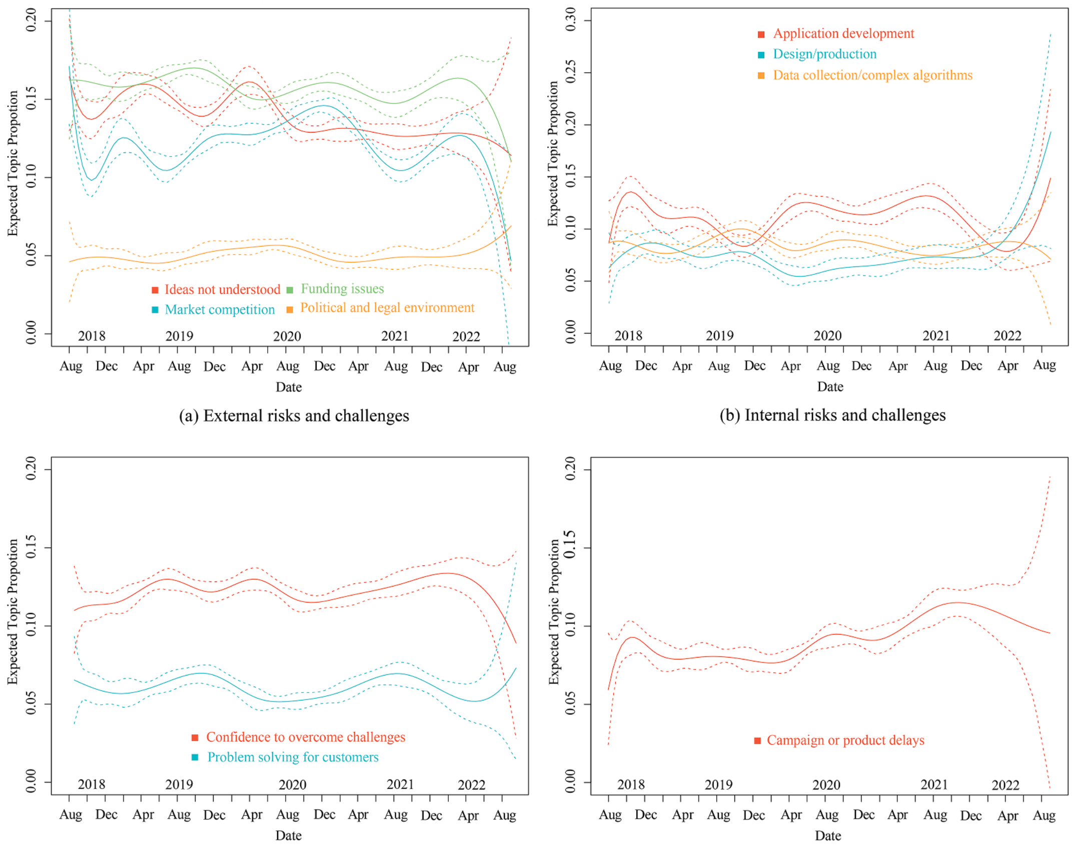
[Application development] One common risk for app developers is pending approval from the Apple and Google Play Store. Our app could be rejected if testing doesn't completely cover any flaws in the application on multiple devices…

The topic of "Data collection/complex algorithms" is mainly related to the fact that digital technology ventures may face obstacles in data collection and protection as well as complex algorithms. The topic "Design/production" is mainly related to problems such as technical immaturity or equipment failure that a product may encounter at the time of design or production. In addition to disclosing their ventures' potential risks and challenges, digital entrepreneurs often discuss how they will respond to these risks and challenges in risk disclosure. By reading the examples in the topic "Confidence to overcome challenges" (12.83%), we discover that under this topic, digital entrepreneurs usually state that their projects have almost no risk or are confident that they can overcome risks and challenges if they occur.The topic "Problem solving for customers" (5.56%) focuses on the promise of what entrepreneurs will do if a problem arises. The final topic "Campaign or product delays", accounting for 11.85% of all topics, focused on the fact that projects may not be completed within the planned schedule after problems are encountered, resulting in project or reward product delays.

4.2. *Topic trend analysis*

With the unique advantages of the STM topic model, we analyse the shifting trends of topics hidden in unstructured text data. Fig. 2 shows the trend in the prevalence of each topic over time, and it can be observed that most of the topics show an increasing and decreasing trend in proportion during the examined period, suggesting that the risks and challenges that digital entrepreneurs face when launching digital technology ventures in crowdfunding platforms are dynamically changing.

Fig. 2(a) shows the changing trend in the prevalence of the four topics under the external risks and challenges topic category. It is observable that the proportion of Topic 1 "Ideas not understood" has been trending downwards since April 2020 and has gradually stabilised, indicating that crowdfunding platforms continue to improve the disclosure of project information through various website policies and designs (Chen et al., 2020), allowing digital entrepreneurs to present their entrepreneurial ideas in more detail, thus reducing the concern of ideas not being understood by others. The proportion of Topic 3 "Market competition" also shows a substantial change during the examined period, with digital entrepreneurs' discussions on this topic increasing between June 2019 and December 2020. Furthermore, Topic 5 "Funding issues" and Topic 10 "Political and legal environment" have remained relatively flat, but the



(a) External risks and challenges (b) Internal risks and challenges

(c) Response to risks and challenges (d) Possible consequences

Fig. 2. Topic prevalence over time estimated by the STM model (The dashed curves indicate the 95% confidence intervals.).

proportion of the two topics has changed more significantly since March 2022, which may be due to the fact that the outbreak of the Russian-Ukrainian conflict has changed the main risks faced by digital entrepreneurs, with the instability of the external political environment becoming an increasing concern for entrepreneurs compared to funding issues.

As shown in Fig. 2(b), the three topics under the category of internal risks and challenges have changed significantly at both the December 2019 and April 2022 timeframes, providing further evidence that global contingencies such as COVID-19 and the Russian-Ukrainian conflict can impact the entrepreneurial activities of digital entrepreneurs, making them shift their concern and discussion about the possible risks to their projects. The proportion of the two topics grouped in the response to risks and challenges category (Fig. 2(c)) also exhibits a generally stable trend over our examined period. Interestingly, however, the proportion of the two topics shows a clear reverse trend since March 2022. The outbreak of the Russia-Ukraine conflict caused a serious disruption to the global supply chain (Mbah & Wasum, 2022), and some entrepreneurs are worried about the impact this may have on the manufacture and delivery of their products. Such unpredictable factor makes entrepreneurs less confident in overcoming risks and challenges, instead trying to provide backers with solutions if the unexpected occurs, which may explain the significant decrease in the proportion of Topic 2 "Confidence to overcome challenges" while the increasing trend in the proportion of Topic 7 "Problem solving for customers" since March 2022.

Finally, it is worth noting that we observed an overall upward trend in the proportion of the topic "Campaign or product delays" (Fig. 2(d)) since April 2020, and by analysing representative examples, we find that many entrepreneurs mention the influence of the COVID-19 outbreak on their campaigns, suggesting that the pandemic may have exposed their entrepreneurial projects to more uncertainty, leading to a much higher likelihood of delays in the campaign or reward delivery.

4.3. Empirical results

Table 3 provides descriptive statistics for all our variables, and the results of the correlation matrix analysis of the variables show that the correlation coefficients between all the independent variables are less than 0.5, indicating a low correlation between the independent variables. Note that introducing of all ten topic probabilities would have led to serial correlations between all topics in the regression model. Therefore, we used only nine of the ten topics and excluded Topic 4 "Campaign or product delays" from the regression model. By removing it, each models' variance inflation factor (VIF) values are distributed between 1 and 2.539, below the acceptable level of 10, indicating that our results are not significantly affected by the multicollinearity problem.

The regression results for our eight models are shown in Table 4, where Models 1–4 represent regressions utilising log (Raised) as the dependent variable, whereas Models 5–8 reprent regressions employing log (Backers). Among them, Model 1 and Model 5 contain only control variables and serve as the baseline model, and the independent variables are added to Model 2 and Model 6 to test the main effect of risk topics on crowdfunding performance. Reward tiers as a moderator variable is added in Model 3 and Model 7, and the interaction terms between reward tiers and risk topics are included in Model 4 and Model 8 to test the moderating effect of the number of reward tiers on the main effect. The values of Adjusted R^2 for all eight models are > 0.5, indicating that our regression models fit well.

Table 3
The descriptive statistics of variables.

Variables	Mean	Std.	Min	Max	N
Dependent variables					
Raised	47,485.744	1059,436.040	0.000	54,970,803.460	4143
Backers	71.980	512.818	0.000	13,052.000	4143
Independent variables					
Topic 1	0.143	0.127	0.005	0.805	4143
Topic 2	0.128	0.088	0.006	0.714	4143
Topic 3	0.115	0.128	0.001	0.868	4143
Topic 4	0.119	0.149	0.002	0.893	4143
Topic 5	0.158	0.129	0.004	0.806	4143
Topic 6	0.106	0.123	0.003	0.864	4143
Topic 7	0.056	0.081	0.007	0.935	4143
Topic 8	0.067	0.108	0.003	0.869	4143
Topic 9	0.068	0.083	0.003	0.799	4143
Topic 10	0.041	0.057	0.004	0.833	4143
Moderator variables					
Reward tiers	5.080	3.783	0.000	44.000	4143
Control variables					
Goal	60,910.477	292,668.574	0.745	10,804,000.000	4143
Duration	39.530	14.170	1.000	74.000	4143
Images	7.900.000	13.461	0.000	145.000	4143
Video exhibition	0.260	0.441	0.000	1.000	4143
Comments	35.840	430.001	0.000	22,073.000	4143
Updates	2.310	6.102	0.000	82.000	4143
Words in R&C	88.920	86.333	3.000	1224.000	4143
Loved project	0.003	0.054	0.000	1.000	4143
USA	0.530	0.499	0.000	1.000	4143

Table 4
Estimation results for OLS regression.

	Dependent variable: log (Raised)				Dependent variable: log (Backers)			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Topic 1		−0.920*** (0.164)	−0.812*** (0.161)	−0.965*** (0.166)		−0.316*** (0.063)	−0.273*** (0.061)	−0.307*** (0.064)
Topic 2		−1.342*** (0.219)	−1.266*** (0.214)	−1.435*** (0.219)		−0.420*** (0.084)	−0.390*** (0.082)	−0.439*** (0.084)
Topic 3		−1.134*** (0.160)	−1.038*** (0.157)	−1.221*** (0.162)		−0.399*** (0.061)	−0.361*** (0.060)	−0.410*** (0.062)
Topic 5		−0.858*** (0.158)	−0.787*** (0.155)	−0.956*** (0.161)		−0.266*** (0.061)	−0.238*** (0.059)	−0.275*** (0.062)
Topic 6		−0.799*** (0.176)	−0.743*** (0.172)	−0.957*** (0.179)		−0.253*** (0.067)	−0.231*** (0.066)	−0.287*** (0.068)
Topic 7		−1.105*** (0.222)	−0.973*** (0.217)	−1.103*** (0.233)		−0.492*** (0.085)	−0.439*** (0.083)	−0.440*** (0.089)
Topic 8		−0.431** (0.193)	−0.309 (0.189)	−0.433** (0.199)		−0.332*** (0.074)	−0.283*** (0.072)	−0.299*** (0.076)
Topic 9		−0.663*** (0.213)	−0.608*** (0.209)	−0.787*** (0.212)		−0.233*** (0.082)	−0.211*** (0.080)	−0.260*** (0.081)
Topic 10		−1.809*** (0.287)	−1.714*** (0.282)	−1.932*** (0.289)		−0.573*** (0.110)	−0.536*** (0.108)	−0.608*** (0.110)
Reward tiers			0.062*** (0.005)	0.068*** (0.005)			0.025*** (0.002)	0.027*** (0.002)
Topic 1 * Reward tiers				0.171*** (0.038)				0.061*** (0.014)
Topic 2 * Reward tiers				0.202*** (0.052)				0.042** (0.020)
Topic 3 * Reward tiers				0.029 (0.037)				0.004 (0.014)
Topic 5 * Reward tiers				0.091** (0.037)				0.034** (0.014)
Topic 6 * Reward tiers				0.110*** (0.042)				0.035** (0.016)
Topic 7 * Reward tiers				0.066 (0.044)				−0.001 (0.017)
Topic 8 * Reward tiers				0.002 (0.041)				−0.017 (0.016)
Topic 9 * Reward tiers				0.227*** (0.048)				0.070*** (0.019)
Topic 10 * Reward tiers				0.134** (0.054)				0.049** (0.021)
Log (Goal)	−0.013 (0.021)	0.017 (0.021)	−0.001 (0.021)	0.005 (0.021)	−0.039*** (0.008)	−0.032*** (0.008)	−0.039*** (0.008)	−0.037*** (0.008)
Duration	−0.004*** (0.001)	−0.003*** (0.001)	−0.003** (0.001)	−0.002** (0.001)	−0.002*** (0.0004)	−0.002*** (0.0004)	−0.001*** (0.0004)	−0.001*** (0.0004)
Log (Images)	0.679*** (0.036)	0.605*** (0.036)	0.459*** (0.037)	0.434*** (0.037)	0.248*** (0.014)	0.225*** (0.014)	0.167*** (0.014)	0.159*** (0.014)
Video exhibition	0.160*** (0.039)	0.151*** (0.039)	0.128*** (0.038)	0.124*** (0.038)	0.037** (0.015)	0.035** (0.015)	0.026* (0.015)	0.025* (0.015)
Log (Comments)	0.881*** (0.038)	0.782*** (0.040)	0.775*** (0.039)	0.798*** (0.039)	0.570*** (0.014)	0.545*** (0.015)	0.542*** (0.015)	0.553*** (0.015)
Updates	0.027*** (0.004)	0.025*** (0.003)	0.018*** (0.003)	0.021*** (0.003)	0.018*** (0.001)	0.017*** (0.001)	0.015*** (0.001)	0.016*** (0.001)
Loved project	0.757*** (0.286)	0.680** (0.283)	0.596** (0.277)	0.586** (0.276)	0.363*** (0.109)	0.350*** (0.108)	0.316*** (0.106)	0.307*** (0.105)
USA	0.198*** (0.031)	0.192*** (0.031)	0.179*** (0.03)	0.177*** (0.03)	0.075*** (0.012)	0.071*** (0.012)	0.066*** (0.012)	0.065*** (0.011)
Log (Words in R&C)	0.323*** (0.057)	0.328*** (0.058)	0.254*** (0.057)	0.244*** (0.057)	0.092*** (0.022)	0.104*** (0.022)	0.074*** (0.022)	0.071*** (0.022)
Constant	0.825*** (0.136)	1.623*** (0.178)	1.522*** (0.175)	1.654*** (0.179)	0.599*** (0.052)	0.873*** (0.068)	0.833*** (0.067)	0.861*** (0.069)
R^2	0.502	0.514	0.534	0.541	0.667	0.673	0.688	0.692
Adjusted R^2	0.500	0.511	0.531	0.536	0.666	0.671	0.686	0.689
MAE	0.811	0.802	0.783	0.782	1.312	1.309	1.303	1.302
MSE	0.962	0.938	0.899	0.897	0.140	0.138	0.132	0.131
VIF	1.010–2.185	1.013–2.466	1.014–2.467	1.040–2.539	1.010–2.185	1.013–2.466	1.014–2.467	1.040–2.539

Notes: $N = 4143$.
[*] $p < 0.1$.
[**] $p < 0.05$.
[***] $p < 0.01$. Robust standard errors are provided in parentheses. In each model, we control for year and subcategory but not reported in the table above.

Hypothesis 1a proposes that topics related to the risks faced by the project in the risk disclosure will have a negative impact on crowdfunding performance. According to Model 2 and Model 6, the seven topics related to the risks faced by the project, namely Topic 1 "Ideas not understood" (Model 2: $\beta = -0.920$, $p < 0.01$; Model 6: $\beta = -0.316$, $p < 0.01$), Topic 3 "Market competition" (Model 2: $\beta = -1.134$, $p < 0.01$; Model 6: $\beta = -0.399$, $p < 0.01$), Topic 5 "Funding issues" (Model 2: $\beta = -0.858$, $p < 0.01$; Model 6: $\beta = -0.266$, $p < 0.01$), Topic 6 "Application development" (Model 2: $\beta = -0.799$, $p < 0.01$; Model 6: $\beta = -0.253$, $p < 0.01$). Topic 8 "Design/production" (Model 2: $\beta = -0.431$, $p < 0.05$; Model 6: $\beta = -0.332$, $p < 0.01$), Topic 9 "Data collection/complex algorithms" (Model 2: $\beta = -0.663$, $p < 0.01$; Model 6: $\beta = -0.233$, $p < 0.01$) and Topic 10 "Political and legal environment" (Model 2: $\beta = -1.809$, $p < 0.01$; Model 6: $\beta = -0.573$, $p < 0.01$) all show statistically significant negative coefficients on the amount raised and the number of backers, indicating that these topics significantly have a negative impact on the crowdfunding performance, supporting hypothesis 1a.

Hypothesis 1b proposes that topics related to entrepreneurs' response to risk will positively impact crowdfunding performance. However, the regression results for Model 2 indicate that the topics related to entrepreneurs' response to risk, that is, Topic 2 "Confidence to overcome challenges" (Model 2: $\beta = -1.342$, $p < 0.01$; Model 6: $\beta = -0.420$, $p < 0.01$) and Topic 7 "Problem solving for customers" (Model 2: $\beta = -1.105$, $p < 0.01$; Model 6: $\beta = -0.492$, $p < 0.01$), also show a significant negative effect on the amount raised and the number of backers, and therefore hypothesis 1b is not supported.

Hence, the results of our main effects examination suggest that topics in risk disclosure significantly negatively impact both the amount raised and the number of backers, whether they relate to the risks faced by the project or to the entrepreneur's response to the risks. After disclosing the specific content of risks, creators encountered more challenges in improving the crowdfunding performance of their projects.

After examining the main effects, we tested the moderating effect of reward tiers on the relationship between topics and the amount raised through Model 4 and the relationship between topics and the number of backers through Model 8. Specifically, for Topic 1 "Ideas not understood" (Model 4: $\beta = 0.171$, $p < 0.01$; Model 8: $\beta = 0.061$, $p < 0.01$), Topic 2 "Confidence to overcome challenges" (Model 4: $\beta = 0.202$, $p < 0.01$; Model 8: $\beta = 0.042$, $p < 0.05$), Topic 5 "Funding issues" (Model 4: $\beta = 0.091$, $p < 0.05$; Model 8: $\beta = 0.034$, $p < 0.05$), Topic 6 "Application development" (Model 4: $\beta = 0.110$, $p < 0.01$; Model 8: $\beta = 0.035$, $p < 0.05$), Topic 9 "Data collection/complex algorithms" (Model 4: $\beta = 0.227$, $p < 0.01$; Model 8: $\beta = 0.070$, $p < 0.01$), Topic 10 "Political and legal environment" (Model 4: $\beta = 0.134$, $p < 0.05$; Model 8: $\beta = 0.049$, $p < 0.05$) six topics, the regression coefficients of their interaction terms with the reward tiers are all significant and positive, indicating that for most of the topics, their negative impact on the amount raised and the number of backers can be moderated by the number of reward tiers. In other words, digital entrepreneurs can help attenuate the negative impact of risk information disclosure on crowdfunding performance by setting more optional reward tiers. Therefore, Hypothesis 2 is supported.

Furthermore, as shown in Model 1 and Model 5, our control variables, such as the number of images, video exhibition, the number of comments, the number of updates, selected as the "Loved project", and launched in the USA all have a significant positive impact on crowdfunding performance, while the funding goal and the duration of the fundraising negatively affect crowdfunding performance, which is consistent with existing research (Nakagawa & Kosaka, 2022; Wessel et al., 2022). In particular, we also reveal an interesting result that the word count of the risk disclosure has a significant positive impact on the amount raised ($\beta = 0.323$, $p < 0.01$) and the

Table 5
Estimation results for 2SLS regression of main effects.

	Dependent variable: log (Raised)	Dependent variable: log (Backers)
Topic 1	-2.147*** (0.340)	-0.858*** (0.131)
Topic 2	-2.903*** (0.427)	-1.182*** (0.164)
Topic 3	-1.683*** (0.255)	-0.705*** (0.098)
Topic 5	-1.860*** (0.274)	-0.707*** (0.105)
Topic 6	-2.355*** (0.369)	-0.956*** (0.142)
Topic 7	-2.779*** (0.480)	-1.182*** (0.185)
Topic 8	-1.344*** (0.396)	-0.769*** (0.153)
Topic 9	-1.990*** (0.339)	-0.852*** (0.130)
Topic 10	-3.435*** (0.502)	-1.464*** (0.193)
Reward tiers	0.060*** (0.005)	0.024*** (0.002)
Log (Goal)	0.014 (0.021)	-0.030*** (0.008)
Duration	-0.003** (0.001)	-0.001*** (0.000)
Log (Images)	0.396*** (0.039)	0.136*** (0.015)
Video exhibition	0.119*** (0.039)	0.023 (0.015)
Log (Comments)	0.678*** (0.042)	0.496*** (0.016)
Updates	0.017*** (0.003)	0.014*** (0.001)
Loved project	0.544* (0.281)	0.286*** (0.108)
USA	0.163*** (0.031)	0.059*** (0.012)
Log (Words in R&C)	0.259*** (0.059)	0.077*** (0.023)
Constant	2.704*** (0.297)	1.372*** (0.114)
R^2	0.518	0.672
Adjusted R^2	0.514	0.670

Notes: $N = 4143$.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$. Robust standard errors are provided in parentheses. Year and subcategory are controlled but not reported in the table above.

number of backers ($\beta = 0.092$, $p < 0.01$) to the project.

4.4. Robustness check

This paper measures the crowdfunding performance of a project with two different continuous dependent variables, namely the amount raised and the number of backers. Also, this study separately conducts OLS regression analyses to examine the effect of topics in risk disclosure on crowdfunding results and the moderating role of reward tiers. The empirical results demonstrate that the two dependent variables show similar results for both our main effects and moderating effects examinations, proving the relative robustness of our OLS regression results.

Furthermore, our regression model and empirical results may be confounded by endogeneity, with the possibility that some factors simultaneously influence topics in risk disclosure and crowdfunding performance of a project. In response to this, in addition to minimising omitted variables or reverse causality problems by including multiple control variables in the baseline regression, we also applied the instrument variables (IVs) method to address possible endogeneity problems (Greene, 2003; Heckman, 1997; Yu et al., 2022), whereby the ranking of the proportion of each topic in every project's risk disclosure is used as instrument variables in two-stage least square (2SLS) regression. Specifically, the ranking of the proportion of each topic in every project's risk disclosure text, which measures the extent of its risk disclosure without affecting the crowdfunding outcomes, satisfies the relevance assumption and the exogeneity assumption of the IVs (Greene, 2003). In addition, the coefficients of all our proposed IVs are significant in the first-stage regression of the 2SLS procedure, with the F-statistics are all greater than 10, indicating the validity of the IVs we have selected. In the second stage regression, the coefficients of the nine explanatory variables remain significant, and Table 5 summarises the estimation results of the 2SLS regression of the main effects. It can be observed that the 2SLS results generally support the OLS results presented above. Therefore, the findings of this paper still hold after considering endogeneity.

5. Discussion and conclusions

This study focuses on digital technology ventures in reward-based crowdfunding platforms and explores the topics in risk disclosure of digital technology ventures and their impact on crowdfunding performance. Specifically, by using structural topic modelling, we first extracted a wide range of topics discussed by digital entrepreneurs in the risk disclosure texts of their projects and gained a series of insights from these topics. Afterwards, we investigated how these risk topics affected crowdfunding performance through regression analysis and analysed the moderating effect of reward tiers on this effect. In addition, we empirically examined our proposed hypothesis using a large amount of real-world data collected from Kickstarter, and the results showed several exciting findings.

To begin with, our STM model successfully extracted ten topics from the risk disclosure text of digital technology ventures, representing the main contents discussed by digital entrepreneurs in risk disclosure. The results of our topic modelling illustrate that the inability of digital entrepreneurs to raise enough funds for their projects is the biggest concern. This finding confirms the argument of Wessel et al. (2022) that technology ventures frequently suffer from issues related to their viability and, therefore, have difficulty obtaining the required funds through crowdfunding channel. In addition, with the unique advantage of the STM model, we also analysed the trends in topics hidden in big data. We found that the risks and challenges confronted by digital entrepreneurs on crowdfunding platforms are dynamically changing. We also established that changes in the external environment can also impact ventures. For example, after the COVID-19 outbreak, the likelihood of delays in entrepreneurial campaigns and reward delivery increased significantly.

Secondly, our regression results suggest that in contrast to the role of extensive information in increasing the likelihood of a project achieving its funding goal, the increased project risk topic content actually decreases fundraising success, which is measured by the amount raised and the number of backers. This result also supports the finding of previous studies that digital entrepreneurs experience more difficulty raising funds following increased project risk salience (Kim et al., 2022; Madsen & McMullin, 2020). However, to our expectation, the topic of responding to risks and challenges mentioned by digital entrepreneurs in their risk disclosures also negatively affects crowdfunding performance. It has been demonstrated that when positive emotions are expressed in serious documents, it is usually because the communicator seeks to hide the relevant negative message (Humpherys et al., 2011). This view seems to provide an explanation for our finding that if positive topics are talked about in risk disclosure, potential backers may have negative language expectation violations (Costello & Lee, 2022) because the credibility of risk disclosure is reduced, which affects potential backers' willingness to invest. Furthermore, our regression results reveal another interesting finding that the text length of risk disclosure positively affects crowdfunding outcomes, suggesting that the word count of risk disclosure contributes to the perception of project quality (Parhankangas & Renko, 2017) and that more detailed information provided by entrepreneurs can help reduce backers' uncertainty and increase the credibility of risk information. The key takeaway is that our study provides evidence that risk disclosure can help reduce information asymmetry.

Finally, this study shows that the impact of risk disclosure topics on crowdfunding performance is significantly moderated by reward tiers, and that having more optional reward tiers reduces the negative impact of risk topics on the amount of funds raised and the number of backers. This conclusion is consistent with our study hypothesis, and the findings of previous studies provide support for this result. In a reward-based context, the setting of the reward structure plays a crucial role in the decision-making process of the backers (Yang et al., 2020). Increasing the number of optional reward tiers enhances the freedom of backers, who have a greater chance of finding the option that best suits them (Kunz et al., 2017). In addition, having more optional reward tiers leads backers to perceive entrepreneurs as more serious and reliable (Frydrych et al., 2014), thus increasing backers' tolerance of project risk and alleviating their concerns about project uncertainty, thus positively moderating the negative impact of the topic in risk disclosure on crowdfunding performance.

5.1. Theoretical implications

The theoretical implications of this study are mainly in the following three aspects. Firstly, this study provides evidence of the viability and efficacy of using an unsupervised machine learning approach, structural topic modelling, to derive insights from large volumes of unstructured text data from crowdfunding platforms, filling a research gap on how to use unstructured risk disclosure texts to investigate the potential risks faced by digital technology ventures on crowdfunding platforms. Furthermore, investigation based on structural topic modelling provides new opportunities for future crowdfunding research. Most previous crowdfunding research has been limited to experimental or interview-based survey methods (Scheaf et al., 2018; Wessel et al., 2022). With the rapid development of digital technologies, our study provides further insights into the application of big data to crowdfunding research in the context of digital entrepreneurship, such as through the use of web crawling and machine learning to expand research samples and analyse unstructured data. With more entrepreneurial activity possibly turning to digital platforms, this approach will not replace but enhance previous research methods.

Secondly, this study demonstrates the usefulness of risk signals in project texts by demonstrating that risk signals in project risk disclosure can significantly influence crowdfunding participants' intention to contribute. We provide evidence that risk signals in texts can contribute to reducing information asymmetry and that funders can use these signals to identify project risks and make wiser choices quickly. In this sense, our study enriches the literature on the application of signal theory to crowdfunding and digital entrepreneurship.

Finally, this study provides evidence of the importance of topics in risk disclosure texts on crowdfunding performance. Unlike previous studies that primarily focused on project risk salience (Kim et al., 2022; Madsen & McMullin, 2020), our study is the first to use a structural topic modelling to extract topics from risk disclosure texts in reward-based crowdfunding, confirming the role of risk disclosure topics in influencing crowdfunding outcomes. We employed structural topic modelling to quantify risk disclosure content as topic probabilities and found the amount of funds raised and the number of backers varies with these probabilities. We complemented previous research on how project risk disclosure affects crowdfunding performance by focusing on these issues.

5.2. Practical implications

Our findings also have important practical implications. Firstly, our research sheds light on the risks and challenges that may arise when launching a digital technology venture on crowdfunding platforms, providing insights for digital entrepreneurs. With the onset of a wave of digital entrepreneurship, an increasing number of digital technology ventures are moving to crowdfunding platforms. Our discoveries can guide digital entrepreneurs, helping them assess the risks and challenges they may encounter in advance and increasing the likelihood of them succeeding in their entrepreneurial campaigns.

Secondly, our research shows that the topics in risk disclosure can significantly impact crowdfunding outcomes. Therefore, entrepreneurs initiating digital technology ventures on crowdfunding platforms should not only focus on common project metrics such as fundraising time and video presentations but also on the content of risk disclosure to improve information transparency. Although our results identify a negative relationship between risk topics and crowdfunding outcomes, this does not mean entrepreneurs should reduce the discussion of risk topics in their risk disclosure. To facilitate smooth transactions and avoid unnecessary trouble, entrepreneurs should explain project risks in-depth, instill confidence in potential backers, and ensure they completely understand the project risks. We also confirm that the risk disclosure text's length significantly impacts on crowdfunding performance. Furthermore, entrepreneurs can also mitigate the negative impact of risk topics by appropriately increasing the number of optional reward tiers.

Finally, for crowdfunding platforms, our findings demonstrate that mandatory risk disclosure can help reduce information asymmetry in the digital environment and contribute to the sustainability of the crowdfunding industry. The long-term benefits of risk disclosure policies may outweigh the short-term detriments (Kim et al., 2022). Therefore, platforms should compel entrepreneurs to disclose risk information in detail and review risk disclosure content to help increase the credibility of information and reduce uncertainty for potential backers, thereby mitigating the negative effects of risk disclosure.

5.3. Limitations and future research

As with most studies, there are limitations to our work that should be further improved in future studies. Firstly, we only selected reward-based crowdfunding as the context of our study, which is just one type of crowdfunding model. For various types of crowdfunding models, crowdfunding participants may have different motivations and require diverse fundraising strategies. Future research can try to compare the different impacts of risk disclosure topics in other types of crowdfunding models, such as donation-based or equity-based crowdfunding, which is a potentially interesting question. Secondly, the absolute discretion of entrepreneurs in disclosing risk information about their projects may lead some entrepreneurs to conceal part of the risk information to reduce negative impacts. Therefore, the insights we gain from "Risks and challenges" on digital entrepreneurship risks may have some limitations. Future research may alternatively gain more authentic insights about the potential risks faced by digital entrepreneurship projects from the backer reviews, which often contain backers' objective evaluations of the projects and the projects' true performance after fundraising. Finally, this study only explored the impact of risk disclosure topic content on crowdfunding performance. However, in the digital context, some other textual cues are equally considered important, such as linguistic ambiguity (Costello & Lee, 2022), emotional tendency (Jiang et al., 2020), and credibility (Kim et al., 2016). Future research can utilise more advanced text mining techniques to determine the impact of other features in risk disclosure texts on crowdfunding performance, contributing to how entrepreneurs disclose risks more strategically.

Funding

This research was funded by the Chinese National Funding of Social Sciences, grant no. 23BJY191 and Natural Science Foundation of Heilongjiang Province of China, grant no. LH2022G014.

CRediT authorship contribution statement

Hong Huo: Investigation, Conceptualization, Project administration, Funding acquisition, Supervision. **Chen Wang:** Investigation, Formal analysis, Software, Data curation, Writing – original draft, Writing – review & editing. **Chunjia Han:** Investigation, Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Validation. **Mu Yang:** Formal analysis, Data curation, Software, Visualization, Writing – review & editing. **Wen-Long Shang:** Conceptualization, Methodology, Software, Visualization, Validation.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

Data will be made available on request.

Acknowledgment

The authors would like to thank the Crowdfunding Net (www.Kickstarter.com) for supplying the data.

References

Ahlers, G. K., Cumming, D., Günther, C., & Schweizer, D. (2015). Signaling in equity crowdfunding. *Entrepreneurship theory and practice, 39*(4), 955–980.
Akerlof, G. A. (1978). The market for "lemons": quality uncertainty and the market mechanism. *Uncertainty in economics* (pp. 235–251). Elsevier.
Allison, T. H., Davis, B. C., Webb, J. W., & Short, J. C. (2017). Persuasion in crowdfunding: An elaboration likelihood model of crowdfunding performance. *Journal of Business Venturing, 32*(6), 707–725.
Anglin, A. H., Short, J. C., Drover, W., Stevenson, R. M., McKenny, A. F., & Allison, T. H. (2018). The power of positivity? The influence of positive psychological capital language on crowdfunding performance. *Journal of Business Venturing, 33*(4), 470–492.
Ba, Z., Zhao, Y. C., Song, S., & Zhu, Q. (2021). Understanding the determinants of online medical crowdfunding project success in China. *Information Processing & Management, 58*(2), Article 102465.
Bai, S., Bi, X., Han, C., Zhou, Q., Shang, W.-L., Yang, M., … He, H. (2022). Evaluating R&D efficiency of China's listed lithium battery enterprises. *Frontiers of Engineering Management, 9*(3), 473–485.
Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing, 84*(1), 1–25.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022. Jan.
Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science, 35*(6), 953–975.
Cai, M., Luo, H., Meng, X., Cui, Y., & Wang, W. (2023). Network distribution and sentiment interaction: Information diffusion mechanisms between social bots and human users on social media. *Information Processing & Management, 60*(2), Article 103197.
Calic, G., & Mosakowski, E. (2016). Kicking off social entrepreneurship: How a sustainability orientation influences crowdfunding success. *Journal of Management Studies, 53*(5), 738–767.
Chakraborty, S., & Swinney, R. (2021). Signaling to the crowd: Private quality information and rewards-based crowdfunding. *Manufacturing & Service Operations Management, 23*(1), 155–169.
Chan, H. F., Moy, N., Schaffner, M., & Torgler, B. (2021). The effects of money saliency and sustainability orientation on reward based crowdfunding success. *Journal of Business Research, 125*, 443–455.
Chen, Y. J., Dai, T., Korpeoglu, C. G., Körpeoğlu, E., Sahin, O., Tang, C. S., et al. (2020). OM forum—innovative online platforms: Research opportunities. *Manufacturing & Service Operations Management, 22*(3), 430–445.
COSO, I. (2004). Enterprise risk management-integrated framework. Committee of Sponsoring Organizations of the Treadway Commission, 2.
Costello, F. J., & Lee, K. C. (2022). Exploring investors' expectancies and its impact on project funding success likelihood in crowdfunding by using text analytics and Bayesian networks. *Decision Support Systems, 154*, Article 113695.
Courtney, C., Dutta, S., & Li, Y. (2017). Resolving information asymmetry: Signaling, endorsement, and crowdfunding success. *Entrepreneurship Theory and Practice, 41*(2), 265–290.
Ding, K., Choo, W. C., Ng, K. Y., & Ng, S. I. (2020). Employing structural topic modelling to explore perceived service quality attributes in Airbnb accommodation. *International Journal of Hospitality Management, 91*, Article 102676.
Frydrych, D., Bock, A. J., Kinder, T., & Koeck, B. (2014). Exploring entrepreneurial legitimacy in reward-based crowdfunding. *Venture Capital, 16*(3), 247–269.
Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
Han, C., Yang, M., & Piterou, A. (2021). Do news media and citizens have the same agenda on COVID-19? An empirical comparison of twitter posts. *Technological Forecasting and Social Change, 169*, Article 120849.
Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., et al. (2016). The role of big data in smart city. *International Journal of Information Management, 36*(5), 748–758.
Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources*, 441–462.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading. MA: Addison-Wesley (content analysis).
Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems, 50*(3), 585–594.
Jiang, C., Han, R., Xu, Q., & Liu, Y. (2020). The impact of soft information extracted from descriptive text on crowdfunding performance. *Electronic Commerce Research and Applications, 43*, Article 101002.
Jiang, H., Wang, Z., Yang, L., Shen, J., & Hahn, J. (2021). How rewarding are your rewards? A value-based view of crowdfunding rewards and crowdfunding performance. *Entrepreneurship Theory and Practice, 45*(3), 562–599.
Kickstarter (2022). Kickstarter stats. Retrieved December 31, 2022 from https://www.kickstarter.com/help/stats.
Kim, J. G., Kong, H. K., Karahalios, K., Fu, W. T., & Hong, H. (2016). The power of collective endorsements: Credibility factors in medical crowdfunding campaigns. In *Proceedings of the CHI conference on human factors in computing systems* (pp. 4538–4549).
Kim, K., Park, J., Pan, Y., Zhang, K. P., & Zhang, X. Q. (2022). Risk disclosure in crowdfunding. *Information Systems Research, 33*(3), 1023–1041.
Kirmani, A., & Rao, A. R. (2000). No pain, no gain: A critical review of the literature on signaling unobservable product quality. *Journal of Marketing, 64*(2), 66–79.
Kunz, M. M., Bretschneider, U., Erler, M., & Leimeister, J. M. (2017). An empirical investigation of signaling in reward-based crowdfunding. *Electronic Commerce Research, 17*, 425–461.
Lee, W. S., & Sohn, S. Y. (2019). Discovering emerging business ideas based on crowdfunded software projects. *Decision Support Systems, 116*, 102–113.
Li, K., Zhang, Q., Kwong, S., Li, M., & Wang, R. (2013). Stable matching-based selection in evolutionary multiobjective optimization. *Ieee Transactions on Evolutionary Computation, 18*(6), 909–923.
Liu, Z., Qian, Q., Hu, B., Shang, W.-L., Li, L., Zhao, Y., … Han, C. (2022). Government regulation to promote coordinated emission reduction among enterprises in the green supply chain based on evolutionary game analysis. *Resources, Conservation and Recycling, 182*, 106290.
Madsen, J. M., & McMullin, J. L. (2020). Economic consequences of risk disclosures: Evidence from crowdfunding. *Accounting Review, 95*(4), 331–363.
Mbah, R. E., & Wasum, D. F. (2022). Russian-Ukraine 2022 War: A review of the economic impact of Russian-Ukraine crisis on the USA, UK, Canada, and Europe. *Advances in Social Sciences Research Journal, 9*(3), 144–153.
Michels, J. (2012). Do unverifiable disclosures matter? Evidence from peer-to-peer lending. *The Accounting Review, 87*(4), 1385–1413.
Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing, 29*(1), 1–16.
Nakagawa, K., & Kosaka, G. (2022). What social issues do people invest in? An examination based on the empathy–altruism hypothesis of prosocial crowdfunding platforms. *Technovation, 114*, Article 102508.
Nzembayie, K. F., Buckley, A. P., & Cooney, T. (2019). Researching pure digital entrepreneurship–a multimethod insider action research approach. *Journal of Business Venturing Insights, 11*, e00103.
Parhankangas, A., & Renko, M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Venturing, 32*(2), 215–236.
Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., et al. (2014). Structural topic models for open-ended survey responses. *American journal of political science, 58*(4), 1064–1082.
Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association, 111*(515), 988–1003.
Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software, 91*, 1–40.
Sandouka, K. (2019). Text analysis of crowdfuding: A literature review.
Sannajust, A., Roux, F., & Chaibi, A. (2014). Crowdfunding in France: A new revolution? *Journal of Applied Business Research (JABR), 30*(6), 1919–1928.
Savin, I., & Teplyakov, N. (2022). Topics of the nationwide phone-ins with Vladimir Putin and their role for public support and Russian economy. *Information Processing & Management, 59*(5), 20. Article 103043.
Scheaf, D. J., Davis, B. C., Webb, J. W., Coombs, J. E., Borns, J., & Holloway, G. (2018). Signals' flexibility and interaction with visual cues: Insights from crowdfunding. *Journal of Business Venturing, 33*(6), 720–741.
Schmiedel, T., Müller, O., & Vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods, 22*(4), 941–968.
Shah, A. M., Yan, X., Tariq, S., & Ali, M. (2021). What patients like or dislike in physicians: Analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach. *Information Processing & Management, 58*(3), Article 102516.
Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational Research Methods, 13*(2), 320–347.
Siebeneicher, S., & Bock, C. (2022). Sustainable aim and personal gain? How sustainable value affects the relation between personal value and crowdfunding success. *Technological Forecasting and Social Change, 183*, Article 121938.
Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software, 1*(3), 37.
Spence, M. (1978). Job market signaling. *Uncertainty in economics* (pp. 281–306). Elsevier.
Tajvarpour, M. H., & Pujari, D. (2022). The influence of narrative description on the success of crowdfunding campaigns: The moderating role of quality signals. *Journal of Business Research, 149*, 123–138.
Thies, F., Wessel, M., Rudolph, J., & Benlian, A. (2016). Personality matters: How signaling personality traits can influence the adoption and diffusion of crowdfunding campaigns. In ECIS (p. ResearchPaper36).
Tian, Y., & Zhang, Y. (2023). The impact of the reward scheme design on crowdfunding performance. *Technological Forecasting and Social Change, 194*, Article 122730.
Wang, Y., Li, Y., & Wu, J. (2022). Digital identities of female founders and crowdfunding performance: An exploration based on the LDA topic model. *Gender in Management: An International Journal, 37*(5), 659–678.
Wessel, M., Thies, F., & Benlian, A. (2022). The role of prototype fidelity in technology crowdfunding. *Journal of Business Venturing, 37*(4), 24. Article 106220.
Westerlund, M., Singh, I., Rajahonka, M., & Leminen, S. (2021). Technology project summaries as a predictor of crowdfunding success. *Technology Innovation Management Review, 11*(11–12), 33–44.
Xiao, S., Ho, Y. C., & Che, H. (2021). Building the momentum: Information disclosure and herding in online crowdfunding. *Production and Operations Management, 30*(9), 3213–3230.
Xu, S. X., & Zhang, X. Q. (2013). Impact of wikipedia on market information environment: evidence on management disclosure and investor reaction. *MIS Q., 37*(4), 1043–1068.
Xu, L. Z. (2018). Will a digital camera cure your sick puppy? Modality and category effects in donation-based crowdfunding. *Telematics and Informatics, 35*(7), 1914–1924.
Yang, L. S., Wang, Z. Y., & Hahn, J. (2020). Scarcity Strategy in crowdfunding: An empirical exploration of reward limits. *Information Systems Research, 31*(4), 1107–1131.
Yang, Z., Shang, W.-L., Zhang, H., Garg, H., & Han, C. (2022). Assessing the green distribution transformer manufacturing process using a cloud-based q-rung orthopair fuzzy multi-criteria framework. *Applied Energy, 311*, 118687.
Yu, H., Jiang, Y., Zhang, Z., Shang, W., Han, C., & Zhao, Y. (2022). The impact of carbon emission trading policy on firms' green innovation in China. *Financial Innovation, 8*(1), 55.
Zheng, H. C., Hung, J. L., Qi, Z. H., & Xu, B. (2016). The role of trust management in reward-based crowdfunding. *Online Information Review, 40*(1), 97–118.
Zheng, Y., Wang, Y., & Mian, S. A. (2022). Strategic positioning of projects in crowdfunding platforms: Do advanced technology terms referencing, signaling and articulation matter? *International Journal of Entrepreneurial Behavior & Research(ahead-of-print)*.