Technological Forecasting & Social Change 198 (2024) 122971

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Investigating reviewers' intentions to post fake vs. authentic reviews based on behavioral linguistic features

Jong Min Kim [a], Keeyeon Ki-cheon Park [b], Marcello Mariani [c,d,*], Samuel Fosso Wamba [e]

[a] College of Social Science of Konkuk University, South Korea
[b] College of Education, Kongju National University
[c] Henley Business School, University of Reading, Greenlands, Henley on Thames, Oxfordshire RG9 3AU, United Kingdom
[d] University of Bologna, Italy
[e] TBS Business School, 1 Place Alphonse Jourdain, TOULOUSE, 31068, France

A R T I C L E I N F O

Keywords:
Fake reviews
reviewers' intention
Online review policy
Textual content
Sentiment analysis
Topic modeling

A B S T R A C T

Growing interest in peer-generated online reviews for product promotion has incentivized online review manipulation. The latter is challenging to be detected. In this study, to discern reviews that are likely authentic vs. fake, we leverage interpersonal deception theory (IDT) and then investigate verbal and nonverbal features that reflect reviewers' intentions to post fake vs. authentic reviews by using topic modeling techniques. Our findings show topic differences between fake vs. authentic reviews. Based on the results, review manipulators tend to write reviews recommending particular movies, while authentic reviewers are likely to provide movie content information in their reviews. Also, we reveal that review manipulation happens at the early stage of product diffusion and contributes to increasing review ratings. Lastly, we discover that manipulated/fake reviews are more informative and positive. Our findings contribute to extend research on online fake reviews literature by innovatively examining review-writing intentions with topic differences, sentiment, and informativeness. To the best of our knowledge, this is the first attempt to introduce topic factors in the fake review detection literature.

1. Introduction

Digital transformation has drastically changed the product purchase process (Gartner, 2019; Zhang and Ghorbani, 2020) in both positive and negative ways. An outcome of digital transformation is electronic word-of-mouth (eWOM) which has become an essential source of information for potential customers' purchase decisions (Shukla et al., 2021). However, the increased importance of online reviews is accompanied by an increased prevalence of fake and misleading information, such as fake online reviews and advertisements (Kim et al., 2023; Zhang and Ghorbani, 2020). In several countries (both in the West and in the East), marketing companies and freelancers intentionally manipulate online reviews by posting promotional or fake reviews (Heinzman, 2019) that can mislead consumers. Furthermore, those marketing companies are increasingly using advanced technologies such as artificial intelligence (AI) and machine learning (ML) to develop automated review generation software that enables the creation of numerous fake reviews on websites.[1] The pervasiveness of fake reviews has become a critical concern in industry and academia (Salminen et al., 2022), because fake reviews compromise consumer decision-making.

To understand the phenomenon at hand, marketing and information management scholars have analyzed textual features with AI and ML techniques to increase detection accuracy (Ozbay and Alatas, 2020; Shu et al., 2017; Zhang et al., 2016). For instance, several researchers have proposed different approaches with linguistic features for fake review detection (Jindal et al., 2010; Kumar et al., 2019; Ludwig et al., 2016; Zhang and Ghorbani, 2020). However, fake reviews may include similar linguistic characteristics to authentic reviews (Chen and Chen, 2015), by imitating authentic reviews (Lappas, 2012). Analyzing only superficial features in textual parts of online reviews such as linguistic, as well as ratings, is not sufficient to distinguish fake reviews from authentic reviews. In the relatively recent research stream of online review

* Corresponding author at: Greenlands, Henley on Thames, Oxfordshire RG9 3AU, United Kingdom.
E-mail addresses: jxk101020@konkuk.ac.kr (J.M. Kim), parkkc88@kongju.ac.kr (K.K.-c. Park), m.mariani@henley.ac.uk (M. Mariani), s.fosso-wamba@tbs-education.fr (S.F. Wamba).

[1] http://internettrend.co.kr/trendForward.tsp.

https://doi.org/10.1016/j.techfore.2023.122971
Received 19 July 2022; Received in revised form 30 October 2023; Accepted 1 November 2023
Available online 18 November 2023

Table 1
Fake review literature review.

Field	Research	Algorithms	Constructs
Information Management & Information System	Banerjee and Chua (2017)	Text mining technique (LIWC) Logistic Regression	Comprehensibility, specificity, exaggeration, negligence
	Cano-Marin et al. (2023)	Text and Network analysis	Text-based
	Shan et al. (2021)	Text mining technique (LIWC, SentiWordNet 3.0) Machine learning classification (SVM, NB, CART, RF, MLPNN)	Sentiment, language styles
	Zhang et al. (2016)	Text minig technique (NLTK 3.0) Machine learning classification (SVM, NB, RF, DT)	Text-based and reviewer-related
	Kumar et al. (2023)	Confirmation factor analysis	Joy of missing out, government regulation, information quality, perceived believability
Marketing	Salminen et al. (2022)	Machine learning-based algorithm (fakeRoBERTa)	Text-based
	Moon et al. (2021)	Text mining technique (LIWC, Text Miner)	Text-based (Emotions, pronouns, cognitive heuristics, time orientations)
	Hajek et al. (2023)	Machine learning-based algorithm (ABAE)	POS tagging, readability, complexity, rating, word embedding

manipulation, there is still an ongoing call for research on fake online review detection that could be based on a thorough analysis of online review features. Therefore, this study extends the nascent research line of online review manipulation to capture the in-depth nature of fake reviews that reflects online review writers' intentions.

According to interpersonal deception theory (IDT) (Buller and Burgoon, 1996), deceptive behavior such as review manipulation is a form of interpersonal communication happening between the sender and the receiver of a message. Since intentional deception entails a complex cognitive process, it is critical to examine the information embedded in communication to understand the communicators' intentions and ultimately detect deception. Such intentions can be reflected in textual characteristics through nonverbal (non-textual) and verbal (textual) cues of the communicators. So far, research investigating the characteristics of fake reviews has focused either on non-textual content, such as online review ratings, or very superficial proxies of textual content, such as the number of words in a document, without considering the intentions of review manipulators that are reflected in textual content (Lau et al., 2012; Luca and Zervas, 2016). Drawing on IDT, this study innovatively investigates the intentions of review manipulators considering both nonverbal (non-textual) and verbal (textual) cues of communication to distinguish likely fake/manipulated reviews from authentic reviews.

One common challenge in fake review detection research is to identify an appropriate dataset that hosts labeled reviews (Wu et al., 2020; Zhang et al., 2016). To overcome such a challenge, in the footsteps of Mayzlin et al. (2014), we use a unique dataset (stemming from the website, Naver.com) that includes both verified and unverified online reviews related to movies in the platform. Fake reviews tend to be short-lived, with decreasing motivational levels along with time to evade detection systems (Hu et al., 2012; Allcott and Gentzkow, 2017), such as the beginning of the product life cycle. Consequently, our paper focuses on short life cycle products such as movies released in cinemas and uses the metadata of posting time to identify fake reviews. Furthermore, this study uses text-mining techniques such as topic modeling and sentiment analysis to extract both verbal and nonverbal features that help revealing reviewers' intentions. Subsequently, those features are examined empirically. In so doing, we advance scholarly understanding of online review manipulation in relation to reviewers' intention and contribute to the lively debate on fake review detection. In the methodological perspective, this paper contributes to fake review literature by freshly introducing a significant factor, topic differences, based on text analysis.

The rest of the paper is organized as follows. Next section discusses related studies in the field of review manipulation and Interpersonal Deception Theory. Section 3 describes our hypotheses. Then, Section 4 presents our method including text-mining techniques for review manipulation. Section 5 presents the empirical results to describe critical factors to differentiate fake reviews with authentic reviews. Section 6 summarizes both theoretical and practical contributions of our work and identifies the limitations and directions for future work.

2. Literature review

Given the significant impact of online reviews on consumers' decision-making process and product evaluation (Shukla et al., 2021), numerous fields including information systems, information management, and marketing have paid increasing attention to manipulation in online contexts. Table 1 presents studies about fake review detection in various fields. Furthermore, some studies investigated the impact of fake reviews, such as consumers' decision-making process (Ananthakrishnan et al., 2020; Zhang and Gupta, 2018). In this section, we discuss related literature in online review manipulation in both methodological and theoretical perspectives. Then, we find research gap that our study can contribute to.

2.1. Online review manipulation

Manipulated online reviews serve the purpose of promoting the products of a focal firm at the expense of competitors' products (Ren and Ji, 2017). In other words, they are produced to mislead consumers in their decision process (Luca and Zervas, 2016). For these reasons, understanding and detecting manipulation in online reviews has become critical and urgent, also because fake review can undermine the credibility and reliance of online reviews.

Advancement in digital technologies has contributed to the over-generation of manipulated reviews (Zhang and Ghorbani, 2020). In response, detection techniques using machine learning methods have been developed to classify reviews into fake and authentic categories (Barbado et al., 2019; Kumar et al., 2018). Several researchers have analyzed textual features in combination with text-mining techniques (Banerjee and Chua, 2017) to effectively detect review manipulation with higher accuracy. For instance, Shan et al. (2021) use machine learning methods to detect fake review detection with textual features that characterize review inconsistency. Banerjee and Chua (2017) identify linguistic characteristics such as linguistic diversity, and the use of particular types of words, which differentiate fake online reviews from authentic reviews.

However, detecting review manipulation only using textual features is insufficient because manipulators generate well-written reviews that are similar to authentic reviews (Chen and Chen, 2015; Zaman et al., 2023). When manipulators have strong deception skills, they become more strategic in hiding information as much as they want and tend to imitate authentic reviews to avoid detection (Lappas, 2012). Moreover, some reviews often use language similar to that used in authentic reviews, to reinforce credibility (Chen and Chen, 2015). Accordingly, to identify more effectively manipulation, it is important to understand online reviewers' intentions. To understand the hidden intentions within online reviews, it is important to uncover patterns in their comments, such as topic features. Therefore, we analyze inherent topics in online reviews to recognize the reviewers' underlying purpose.

2.2. Interpersonal deception theory

Manipulation of reviews is fundamentally a practice of deception (Grazioli and Jarvenpaa, 2003). To discern how online reviews are manipulated as a form of deception, we draw upon the interpersonal deception theory (IDT) (Buller and Burgoon, 1996). Deception can be broadly defined as a deliberate action of a sender to mislead the receiver (Burgoon et al., 1994). IDT considers deception as an interpersonal communication process between sender and receiver (Buller and Burgoon, 1996). Thus, deceptive action can be explained in terms of the key components of communication, such as the receiver, the context, the message, the feedback, and the channel (Infante et al., 2010; Kim et al., 2021) that, conjointly, can reveal the deceivers' intentions.

IDT explains that the communicators' intentions are likely to be reflected in two types of behaviors: 1) strategic behavior (verbal cues) and 2) nonstrategic behaviors (leakage of nonverbal cues) during deception (Buller and Burgoon, 1996). Unlike truth-tellers, deceivers often adopt unusual verbal and nonverbal behaviors to mislead people effectively. Thus, we need to examine both types of behaviors to detect deception more accurately. However, most of the existing studies examining online review manipulation detection heavily rely only on strategic behaviors such as verbal cues (Hajek et al., 2023; Moon et al., 2021; Wang et al., 2012) because it is challenging to detect nonstrategic behaviors in online reviews. Zhang et al. (2016) categorize both verbal and nonverbal behavior features in the online review environment based on IDT. They also demonstrate that incorporating nonverbal features such as rating, reviewer's characteristics, positive ratio, etc., can significantly improve detection performance for online review manipulation. These results of the study conducted by Zhang et al. (2016) also emphasize the importance of examining both strategic and nonstrategic behaviors of deceivers for better performance. Therefore, in line with the research gap identified, this study analyses strategic and nonstrategic behaviors in online settings with reference to the intentions of online review manipulators.

3. Hypotheses development

Since review manipulators attempt to mimic authentic reviews, detecting manipulated reviews is a difficult task. So, in the footsteps of Mayzlin et al. (2014), we adopt an indirect approach to infer vulnerability to review manipulation. We empirically exploit a key difference between two types of reviews: verified vs. unverified reviews. If a customer purchases a product by using the website and creates a review about the product, this customer is classified as verified, which means his/her review is authentic. Otherwise, the customer is classified as unverified, and the review is known to be vulnerable to manipulation (Mayzlin et al., 2014).

3.1. Reviewers' nonverbal behavior

Nonverbal behaviors in interpersonal deception theory (IDT) entail facial expressions, gestures, and body movements (Zhou et al., 2013). However, those are not available in online reviews. Dennis et al. (2008) show that the media capabilities in online communication environments such as rehearsability and reprocessability enable to capture nonverbal behavior in alternative forms. As a result, nonverbal behaviors can be inferred from the reviewers' behaviors in the context of online reviews (Banerjee and Chua, 2014), such as posting time.

Review posting time refers to the time when review manipulators choose to mislead readers. For instance, when businesses release new products or services, they might promote their products or services through manipulated reviews at the initial stage of the product lifecycle to boost their sales because online reviews influence consumers' purchasing behavior (Casaló et al., 2015). In particular, products that have "short" product life cycles such as movies released in cinemas may require strong marketing activities at the initial stage, because sales decline steeply and after a limited period of maturity (Goldman, 1982). In other words, the early period after the movie release can be strategically considered a promotional period for movie distributors and investors. Then, they concentrate their limited promotion budget on this promotional period to increase the sales. Based on this inference, we expect that review manipulation for product promotion will be concentrated over the promotional period in the movie industry, as movie products released in cinemas have relatively short life cycles.

Therefore, review posting time is a behavioral feature of reviewers that can be used to distinguish fake reviews from authentic reviews. Considering explicitly the posting time in fake review detection allows us to extract the fake reviews generated for promotion purposes, as promotion is concentrated on the early period after the cinema release date in the context of movie industry. Whereas the broad definition of fake reviews, this study focuses on fake reviews for promotional purposes that can ultimately translate into revenue maximization. Based on the discussion above, we hypothesize that in the context of products with short life cycles:

H1. Movie promoters are likely to generate a larger volume of fake online reviews during the promotional period.

Based on signaling theory, signals are critical factors in communicating product quality and reducing information asymmetry (Spence, 1978). In particular, reliance on signals becomes higher in experience goods because judgments on quality are difficult without having an experience or consumption of the good (Huang et al., 2011). In the context of manipulation, using signals such as review volume and numeric ratings may be effective in reducing uncertainty about product quality and increasing reputation. Therefore, when unethical businesses manipulate online reviews, it is likely that they will post numerous reviews with high numeric ratings. Accordingly, we hypothesize:

H2. Movie promoters who intend to manipulate online reviews are likely to post higher review ratings during the promotional period.

3.2. Textual content: review informativeness, topics, and sentiments

Since manipulative reviews are often crafted like authentic reviews and manipulation behavior is hidden in the guise of text, investigating textual features in online reviews does not provide sufficient clues to detect review manipulation. To address this potential issue, we thoroughly examine reviewers' intentions in online reviews more thoroughly, such as informativeness (Banerjee and Chua, 2017), sentiment (Hu et al., 2012) and, innovatively also review topics.

Based on the information manipulation theory, the difference between authentic and manipulated texts depends on the quality (McCornack, 1992). Informativeness is a feature that represents the richness in details of the written text (Banerjee and Chua, 2017). Previous literature suggested that authentic writing is richer in nouns (Rayson et al., 2002; Banerjee and Chua, 2017). However, when consumers browse reviews, consumers carefully examine review content and the competence and perceived knowledge of reviewers in their language style (Connors et al., 2011). Reviews that are not informative might be the byproduct of the incompetence of reviewers, and consumers are likely to be reluctant to follow and use these reviews to support their decisions. In addition, manipulators can use AI and ML-

Fig. 1. Viewer-type and netizen-type reviews.

empowered software, including Generative AI (Dwivedi et al., 2023), for automated online review generation to make their reviews more informative to deceive readers more effectively. Therefore, it is likely that review manipulators would use informative writing to make their reviews more credible in the eyes of consumers.

H3. Review manipulators are more likely to use more informative language than authentic reviewers.

Manipulating sentiment is one of the persuasion strategies that reviewers use to persuade a reader to think in a particular way (Hu et al., 2012). Such a manipulated sentiment is commonly adopted in many areas where the writer intends to influence the third party's opinion, such as political news (Kahn and Kenney, 2002). For example, when the media reports news about a local firm with fewer negative words to persuade readers to consider the firm favorably, the firm's market value increases (Gurun and Butler, 2012). Building upon this, review manipulators would tend to use a positive slant in the form of sentiment to promote their products or services and persuade consumers' purchase decisions.

H4. Review manipulators post more reviews with positive sentiment than authentic reviewers.

Our research also attempts to find the pattern or semantic structure within the reviews, such as topic features, to understand the potentially different states of mind that review manipulators may have while creating their reviews. In the exiting literature, the use of topic features in review manipulation detection is very limited or unclear (Wang et al., 2018). However, to understand in-depth information in online reviews such as intentions, it is important to uncover hidden patterns in text comments.

The previous deception research suggests that the deceivers tend to offer fewer details on the fact (or product) being reviewed because they fabricate stories and have less familiarity with the topic (Burgoon et al., 2003; DePaulo et al., 2003). We assume that the review manipulators show similar linguistic profiles to deceptive people because they mostly did not consume the product (e.g., they did not watch a movie) before creating their reviews. Since these manipulators have less information and familiarity with the content of a product (e.g., a movie), they are more likely to include fewer details about the product (e.g., a movie) itself. Instead, the purpose of review manipulation is mostly promotion, thus, review manipulators are more likely to post stories that persuade consumers to purchase the product. Therefore, we propose that review manipulators post more "recommendation-related" topics and less "content–related" topics than authentic reviewers.

H5. Review manipulators post more "recommendation-related" reviews than authentic reviewers.

H6. Review manipulators post fewer "content-related" reviews than authentic reviewers.

4. Research approach

To extract the intention-related variables from text comments in reviews, we applied several machine learning algorithms. First, we used Deep Neural Network to conduct sentiment analysis that is suitable for our movie dataset. Second, we performed topic modeling that uses unsupervised machine learning to discover the abstract topics in a collection of text comments. Lastly, we use text-mining technique to find the number of nouns used in each text comment.

4.1. Data sources

Before conducting our method, we collected online review data for movies from Naver.com in February 2017. To scrape online reviews about movies from Naver.com, we utilized Visual Basic.NET. Visual Basic.NET (VB.NET) is a programming language developed by Microsoft, based on the .NET framework, and is an object-oriented programming language. This program has proven advantageous for identifying and extracting HTML elements from web pages, making it a popular choice for scraping online reviews in previous research (Yousaf and Kim, 2023). The most financially successful 58 movies from 2014 to 2016 were selected (20 in 2014, 20 in 2015, and 18 in 2016).[2] Naver.com is the leading portal website, having over 60 % search engine market share in South Korea.[3] Unlike other movie review platforms, it has an online review policy allowing customers to write two types of customer reviews within the platform: the so-called viewer-type and netizen-type reviews (Fig. 1). After purchasing a movie ticket via the website, if a customer posts a review about the movie, the review is classified as a viewer-type review (verified-type review). Without doing it via the website, if a customer posts a review about a movie, it is classified as a netizen-type of review (unverified-type review). The left side of Fig. 1 provides the summary statistics of the viewer- and netizen-type reviews for a movie and the right-side shows examples of the viewer- and netizen-type reviews. The average rating of netizen-type reviews for the movie is 9.30, while that of viewer-type ones is 9.00. There usually exist some differences in the average rating between the two.

4.2. Data description

The data set consists of 1,106,165 online reviews for 58 movies. We

[2] There are two reasons why we chose the most successful 58 movies from 2014 to 2016. The first reason is that they dominated the movie industry for the years. This means that they are relatively less heterogeneous. This is helpful to empirically test our proposed hypotheses. Secondly, the most successful movies are also those covered by the highest amount of online reviews to make LDA more accurate. If we chose less successful movies, the accuracy of LDA might decrease.

[3] https://www.interad.com/en/category/insights/korean-search-engine-market-share.html.

Table 2
Summary statistics of customer reviews for 60 days after movie release.

Variable	Viewer-type reviews	Netizen-type reviews
Number of observations	465,815	640,350
Mean	8.71	8.41
Standard deviation	1.63	2.81
Minimum	1	1
Maximum	10	10

collected the reviews posted during the 2 months after the movie release. Among the movies, 29 are domestic movies. In South Korea, there are 4 major distributors[4] (CJ ENM, Lotte Entertainment, Showbox, and Next Entertainment). Among the 58 movies, 31 movies were distributed by these major distributors.

Table 2 shows the composition of online reviews related to the movies. Among 1,106,165 reviews, 465,815 were posted by viewers, while the remaining ones were posted by netizens (i.e., online reviewers who did not make a verified purchase). The average rating of viewer-type reviews is 8.71, which is higher than that of netizen-type reviews. Netizen-type reviews have a higher value of the standard deviation.

For the empirical analysis, we aggregate individual reviews into daily data. For example, review ratings are aggregated into a daily value by calculating the average. The empirical analysis consists of two parts. The first part focuses on the impact of review manipulation on nonstrategic behaviors such as leakage of nonverbal cues based on the entire dataset (individual review-level data) aggregated into daily data (H1 and H2). The second part focuses on the strategic behaviors of review manipulators such as verbal cues.

4.3. Sentiment analysis

The sentiment classification attempts to classify the reviews into 'positive' or 'negative' classes. For Korean texts, there are a relatively small number of pre-trained and well-built sentiment models. We developed our own model with the Naver sentiment movie corpus (https://github.com/e9t/nsmc), which is based on the method of (Maas et al., 2011). This dataset consists of a total of 200,000 reviews (train: 150,000, test: 50,000) by collecting 100 reviews per movie among Naver movie reviews. The scores from 1 to 4 were assigned as negative, while scores 9 and 10 were assigned as positive. The neutral ratings (5 to 8) were excluded from this paper because we consider the objective (neutral) texts are less informative (Pang and Lee, 2004). Therefore, we focus only on the polar subjective statements (positive or negative) to improve the sentiment classification.

We used a deep neural network that provides relatively high accuracy on predicting sentiments of movie reviews. Fig. 2 shows the structure of our sentiment model. We preprocessed the input data, online reviews by tokenizing the input data, correcting misspelled sentences and spacing, and removing stop words. Then, the tokenized texts are vectorized by the word embedding, then they are fed into DNN layers. After obtaining the final representation, it was finally fed into the Sigmoid classifier layer. Then, we can classify the sentiments of our dataset as the output of the DNN. Our model performs with high accuracy (83.93 %) based on the Python Keras accuracy indicator (Binary accuracy). Compared to other studies on sentiment analysis, our average accuracy is relatively high. For instance, Duncan and Zhang (2015) had an average accuracy of 74.15 %, and Indriani and Nugrahadi (2016) achieved 72.3 % accuracy in classifying the sentiment of tweets. We applied this model to the entire dataset. As a result, the positive and negative categories accounted for 75.21 % and 24.77 % of the total review texts. According to previous research (Nigam and Hurst, 2004; Arafat et al., 2014), when assessing the sentiments of a given text, human analysts tend agree around 80–85 % of time. To provide more accurate answers, we decide to use 90 % prediction probability to classify reviews' sentiment.

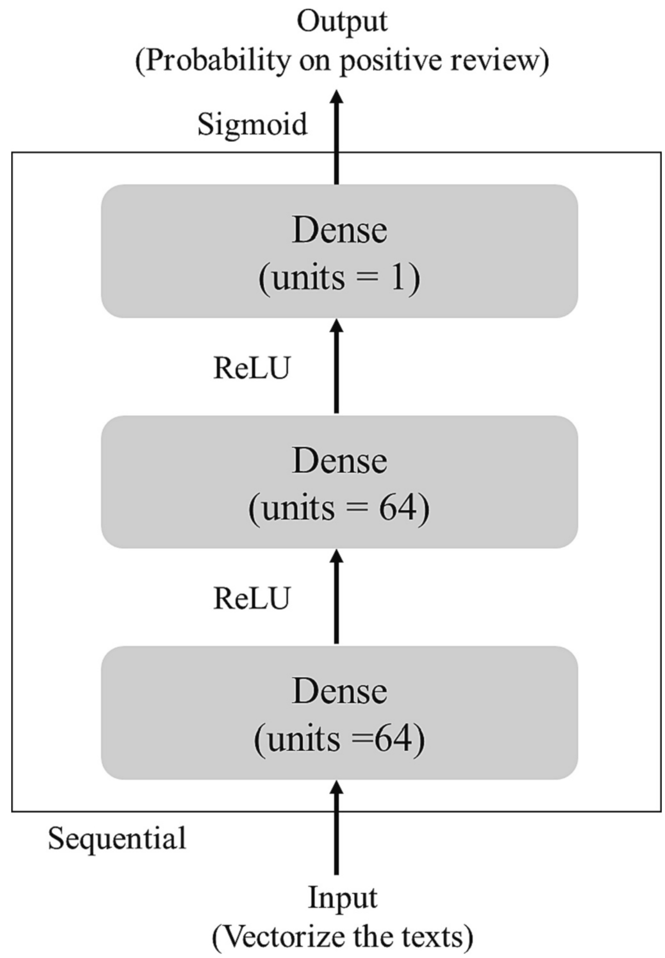

Fig. 2. Sentiment model structure.

4.4. Topic modeling with latent Dirichlet allocation (LDA)

The Latent Dirichlet Allocation (LDA) modeling technique is the most commonly used topic modeling method (Calheiros et al., 2017). It allows for determining the probability of the chosen review belonging to each topic, grouping reviews according to their proximity regarding each considered term. In this research, we adopt LDA in the context of analyzing review manipulation of the textual content of viewer-type vs. netizen-type reviews. Each reviewer may have different underlying subjects to discuss in a review. We expect that promotional reviews will receive differentiated attention from other reviews, and these features are embedded in topic models. As a result, we will explain both authentic and manipulative reviewers' intentions in their reviews.

Before conducting topic modeling analysis with the LDA technique, punctuation, potentially problematic symbols, whitespace, one-letter words, and stop words in the dataset were deleted to ensure that only content words were left as a corpus. Also, topic modeling was applied only to each review that contains more than three nouns as a document. If a review contains a very limited number of words, there are not enough word co-occurrence instances due to word sparsity in such short reviews. Therefore, we put a constraint on the number of nouns contained in a review to cover the potential topics from reviews more effectively.

[4] https://brunch.co.kr/@cmin4411/608.

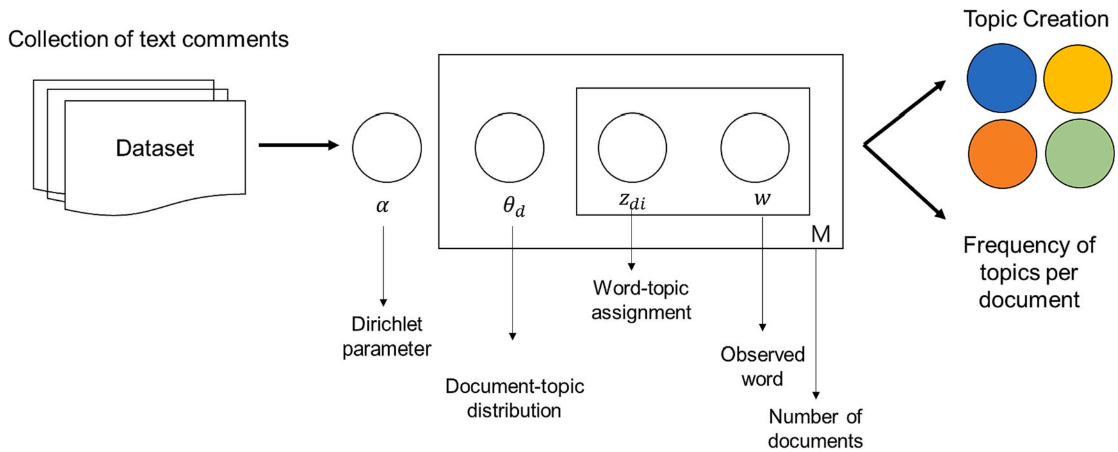

Fig. 3. LDA process.

The LDA model assumes a generative process by which the textual data in each document is generated. Topic proportion θ_d is drawn from a Dirichlet distribution with K-dimensional parameter vector α_d. The topic assignment z_{di} for the word i in document d is drawn from a categorical distribution with parameter θ_d. Given the topic assignment z_{di}, the word i in document d is drawn from a categorical distribution associated with the assigned topic. To exploit conjugacy, each topic distribution is also specified Dirichlet with hyperparameter β. Fig. 3 describes the process of the topic analysis.

This process is repeated for each word in each document. It ignores the order of words in a document. We used one of the traditional metrics for evaluating topic models, coherence score, to find the optimal number of topics. The topic coherence score measures a single topic by measuring the degree of semantic similarity between high-scoring words in the topic. Based on the coherence values, we choose 16 topics as those embedded in the best model.

In Table 3, we present the top 10 words in decreasing order of posterior probability of being in each of the 16 topics, as inferred from the analysis of all reviews of 58 movies. These comprise 15 topics that provide more specific subjects and "general" topic (ID number 2). First, we find that a substantial number of topics focus on recommendation (e.g., ratings, and recommend kids movies). Second, several topics focus on movie content such as characters, direction, story development, acting ability, and particular genre. We call the reviews reflecting the first type of topics as "recommendation" related reviews while calling them reflecting the second type of topics as "movie specific content" related reviews.

4.5. Model specification

Fig. 4 outlines the general steps of our methodology.

To explore the impacts of review manipulation, Mayzlin et al. (2014) used the organizational differences between the online review websites Expedia and TripAdvisor. Expedia enforces a policy allowing only those reviewers who have booked a hotel via the website to leave their reviews (Verified-type policy). Unlike this, TripAdvisor enforces a policy allowing anyone who wants to leave their opinions about a hotel to post their reviews (Unverified-type policy). Therefore, the cost of leaving a promotional review on TripAdvisor is significantly low relative to that on Expedia. Mayzlin et al. (2014) concluded that the differences in online reviews for the same hotels between Expedia and TripAdvisor could be attributable to review manipulation. The setting of their study is very similar to our empirical setting as Naver.com enforces both review policies: viewer-type and netizen-type. The cost of posting a

Table 3
Major topics, associated words, and topic proportions.

ID	Group name	Topic name	Topic proportion (%)	Top 10 words in decreasing order to the posterior probability of belonging to the topic
0	Recommendation	Movie rating	7.5 %	Rating, critic, part-time job, reason, watch, evaluate, release, this movie, reason, today
1	Recommendation	Recommendation	7.9 %	Really, real, absolutely, first-time, oh-my-god, this, sincere, movie theater, highly recommend, creeps
2	No topic group	General topic	5.8 %	Movie, thought, again, once, zombie, Korea, people, if, all, our country
3	Recommendation	Recommend kid movies	7.0 %	Real, fun, impressed, recommend, kid, all, song, video, music, Disney
4	Feelings	Experience of watching alone	4.0 %	Movie, thought, understanding, feeling, first-time, watch, individual, mind, let me, alone
5	Feelings	How they feel about story characters	5.2 %	Just, degree, one, story, content, not taste, strange, individual, not at all, worst
6	Satisfaction	Impression	6.5 %	Deeply moved, tears, heart, family, story, mind, love, during movie, father, parents
7	Movie content	Movie characters	5.3 %	Yi Sunsin, Korea, human being, expression, emotions, circumstances, Choi Minsik, character, heroes, admiral
8	Movie content	Movie materials/direction	5.3 %	Movie, director, work piece, direction, audience, movie materials, self, into a movie, entertainment, restraint
9	Movie content	Acting skills	9.7 %	Acting, actor, Kang Dongwon, acting ability, Hwang Jungmin, Ha Jungwoo, Yoo Ahin, Song Gangho, Lee Byunghyun, Son Yejin
10	Satisfaction	Tension	5.7 %	Scene, throughout the movie, last, twisted, tension, laughter, immersion, thrill, comic, continue
11	Movie content	Story development	6.1 %	Story, sense, little, content, part, development, slightly, ending, somewhat, character
12	Movie content	Action/Marvel series	7.3 %	Action, as expected, expectation, series, disappointed, marvel, this time, next time, X-men, scale
13	Feelings	Connect with history	6.8 %	Our, history, present, time, nation, memory, sacrifice, our nation, appreciate, how
14	Satisfaction	Immersion	6.4 %	Best, movie, time, really, the most, lifetime, this year, during, number one, during movie
15	Feelings	Connect with reality	3.6 %	Movie, person, reality, Korean movie, thought, nowadays, society, different, this movie, what

promotional review in the case of netizen-type reviews is significantly lower than that of viewer-type reviews. Viewer-type reviews can be assumed to be posted by actual viewers of a movie (Type 1). However, netizen-type reviews can include reviews from viewers (Type 1), review manipulators (Type 2), and potential customers who want to share their opinions without watching the movie (Type 3). Table 4 shows the differences in the reviewers between the two types of reviews.

The study by Mayzlin et al. (2014) uses the difference in the two types of reviews as the identification strategy. However, that study displays a limitation: verified-type reviews can be assumed to be posted by customers who stayed at a hotel (Type 1), but unverified-type reviews can be assumed to be posted by reviewers who stayed at a hotel, review manipulators, and potential customers who want to post their opinions about a hotel without staying at the hotel without any financial incentives (Type 1 + Type 2 + Type 3). The problem in the identification stems from the existence of potential customers (Type 3). Therefore, the differences cannot be solely attributed to review manipulation. The differences can be caused by the potential customers who want to post their opinions about a hotel without staying at the hotel and that have no financial incentives (Type 3). To alleviate this limitation, we take a detouring strategy. That is, we include the timing of review manipulation over the promotional period. Accordingly, we incorporate the promotional period as the independent variable.

Considering the empirical approach by Mayzlin et al. (2014), we use a difference-in-differences approach of the dependent variables as the baseline model. Eq. (1) shows the model specification. If there exists an additional difference (Type2) beyond the difference in the constant (Type1 + Type3 – Type1), then this means that the additional difference is solely caused by the review manipulation (Type2 = (Type1 + Type2 + Type3 – Type1) - (Type1 + Type3 – Type1).

$$\frac{N\ Reviews_{it}^{Netizen}}{Total\ Reviews_{it}^{Netizen}} - \frac{N\ Reviews_{it}^{Viewer}}{Total\ Reviews_{it}^{Viewer}} = \alpha + X_{it}{*}\beta + \sum \gamma_i + \delta_i + \varepsilon_{it} \quad (1)$$

where α: a constant, β: the coefficient related to the independent variable (Promotional Period), N: the number of reviews, i: movies, t: days, and γ_i is movel-level heterogeneity. In the model specification, we add monthly fixed and day-of-week fixed effects to control the impacts of seasonality and weekend effects on online review generation. δ_i means movie fixed effects, and ε_{it} is a stochastic error term.

In the case of movies, the financial performance (box office) of movies in South Korea is mostly determined by the performance in the first two weeks after movie release (Ma et al., 2019). Therefore, after two weeks of the movie release, review manipulators would lose the chance to boost movie sales through review manipulation. Hu et al. (2012) also found that there is less review manipulation as time elapsed. Considering this, we use the time elapsing of movies as the independent variable.

Promotional Period: We divide the online reviews into two groups based on the review posting periods. The independent variable, "Promotional Period", represents the first two weeks after movie release when manipulation is most likely. This variable equals 1 for online reviews belonging to the first two weeks and 0 otherwise. In Eq. (1), the constant (α) captures the differences in the dependent variables attributable to someone who wanted to post their opinions about the movie without watching the movie (potential customers). Because review manipulation would concentrate on the first two weeks after the movie release, the estimated coefficient of "Promotional Period" captures the impacts of review manipulation on the differences in the dependent variables.

5. Empirical results

5.1. Review manipulation intensity over time

To test the first hypothesis (1) regarding review volume, we estimate Model (1) using Least Square Dummy Variable (LSDV), where the dependent variable is the difference in the number of reviews between netizens and viewers:

$$\begin{aligned} Model\ (1) : & \ log(n_{1t})^{Netizen} - log(n_{2t})^{Viewer} \\ & = \alpha + X_{it}{*}\beta + \sum \gamma_l{*}Controls_{it}^{l} + \delta_i + \varepsilon_{it} \end{aligned}$$

where n_{1t}: the number of netizens' reviews per each day and n_{2t}: the number of viewers' reviews per day.[5] If there is an increase in the number of netizen type reviews during the promotional period, then it could be attributed to review manipulators. This would be reflected in the estimated positive coefficients of the "Promotional Period".

Table 5 shows the empirical results. In the first column (1), we use the Huber sandwich estimator, while in the second column (2), we use Clustered sandwich estimator for the error term. We use monthly and day-of-week, as the control variables (Controls).[6] We also consider movie fixed effects by including movie-level dummy variables. Both of the estimated coefficients of Promotional Period in (1) and (2) are significantly positive ($\beta_{\text{Promotional Period}} = 1.15$, *p-value* $< .01$, $\beta_{\text{Promotional Period}} = 1.15$, *p-value* $< .01$, respectively). These results show that there is a significant increase in the number of reviews posted by netizens during the Promotional Period, supporting H1. The R^2 is about 0.56 for both columns.

5.2. Review manipulation and average review ratings

To boost movie box office, review manipulation can be conducted by intentionally posting positive reviews including high ratings like "10". If there is an increase in the average rating of netizen-type reviews during the promotional period, then it could be attributed to review manipulators. This would be reflected in the estimated positive coefficient (β) of the "Promotional Period". To test this, we estimate Model (2), where the dependent variable is the difference in the average review ratings between netizen-type reviews and viewer-type reviews:

$$\begin{aligned} Model\ (2) : & \ \frac{\sum (Review\ Rating)_{1t}^{Netizen}}{n_{1t}} - \frac{\sum (Review\ Rating)_{2t}^{Viewer}}{n_{2t}} \\ & = \alpha + X_{it}{*}\beta + \sum \gamma_l{*}Controls_{it}^{l} + \delta_i + \varepsilon_{it} \end{aligned}$$

where n_{1t}: the number of netizens' reviews per each day and n_{2t}: the number of viewers' reviews per day.

Table 6 shows the empirical results. Columns (1) and (2) include the same variables as those of Table 5. The estimated coefficients of the Promotional Period in (1) and (2) are significantly positive ($\beta_{\text{Promotional Period}} = 0.30$, *p-value* $< .01$, $\beta_{\text{Promotional Period}} = 0.30$, *p-value* $< .01$, respectively). These results mean that there is a significant increase in the average ratings of reviews posted by netizens during the first two weeks, supporting H2. The R^2 is about 0.40 for both columns.

5.3. Review manipulation and informativeness

As was expected in H3, review manipulation would focus on providing information-based content. This means that the reviews provided by netizens during the first two weeks are more likely to contain object words (nouns) in the reviews. To test this, we estimate Model (3), where the dependent variable is the difference in the average noun ratio between netizen-type reviews and viewer-type reviews:

[5] These notations (n_{1t} and n_{2t}) are applied the same for all models we used in this research (from Model (1) to Model (6)).

[6] The same control variables are used for all models (from Model (1) to Model (6)).

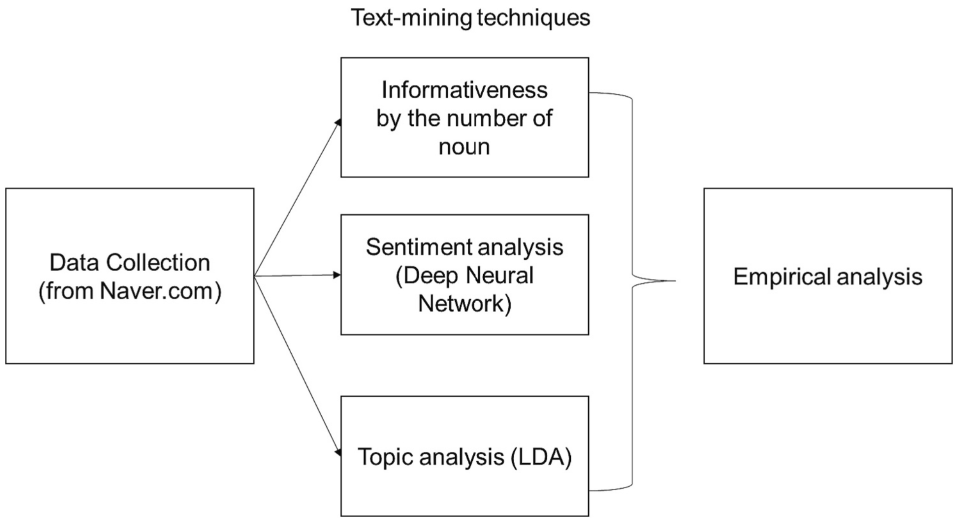

Fig. 4. Methodology steps.

Table 4
Review and reviewer types.

Review types	Reviewer types
Viewer-type reviews	Actual viewers of a movie (Type 1)
Netizen-type reviews	Actual viewers of a movie (Type 1) + Review manipulators (Type 2) + Potential customers (Type 3)

$$Model\ (3): \frac{\sum\left(\frac{Word\ count\ of\ nouns}{Word\ count\ of\ a\ Review}\right)_{1t}^{Netizen}}{n_{1t}} - \frac{\sum\left(\frac{Word\ count\ of\ nouns}{Word\ count\ of\ a\ Review}\right)_{2t}^{Viewer}}{n_{2t}} = \alpha + X_{it}{*}\beta + \sum \gamma_l{*}Controls_{it}^{l} + \delta_i + \varepsilon_{it}$$

where n_{1t}: the number of netizens' reviews per day and n_{2t}: the number of viewers' reviews per each day. If there is an increase in the level of the average noun ratio of netizen-type reviews during the promotional period, then it could be attributed to review manipulators. This would be reflected in the estimated positive coefficient (β) of the promotional period.

Table 7 provides the empirical results. The estimated coefficient of the promotional Period in column (1) is significantly positive ($\beta_{\text{Promotional Period}} = 0.01$, *p-value* $< .01$). The coefficient of column (2) is also significantly positive ($\beta_{\text{Promotional Period}} = 0.01$, *p-value* $< .01$). This shows that the reviews posted by netizens during the first two weeks are more likely to contain object words (nouns), supporting H3. The R^2 is about 0.09 for both columns.

Table 5
The relationship between review manipulation and the number of reviews.

Variable	DV: *Diff in the number of reviews*	
	(1)	(2)
Promotional period	1.15***	1.15***
	(0.03)	(0.06)
Constant	−0.50***	−0.50
	(0.18)	(0.42)
Monthly-FE	Included	Included
Day-of-the-week FE	Included	Included
Movie FE	Included	Included
Huber/white/sandwich estimator	Included	–
Clustered sandwich estimator	–	Included
R-squared	0.56	0.56
# of observations	3350	3350

Notes: ***/** indicates significance at the 1 %/5 % level. Standard errors are in parentheses.

5.4. Review manipulation and sentiment

Because review manipulators tend to promote movies by providing positive reviews, those manipulated reviews are more likely to incorporate positive emotions into the reviews to justify their positive reviews. This would be confirmed by an increase in positive emotional reviews posted by netizens during the first two weeks. To test this, we use the following Model (4):

$$Model\ (4): \frac{(\#of\ \text{positive emotional reviews})_{1t}^{Netizen}}{n_{1t}} - \frac{(\#of\ \text{positive emotional reviews})_{2t}^{Viewer}}{n_{2t}} = \alpha + X_{it}{*}\beta + \sum \gamma_l{*}Controls_{it}^{l} + \delta_i + \varepsilon_{it}$$

where n_{1t}: the number of netizens' reviews per day and n_{2t}: the number of viewers' reviews per day. If there is an increase in the percentage of positive emotional netizen-type reviews during the promotional period, then it could be attributed to review manipulators. This would be reflected in the estimated positive coefficient (β) of the Promotional Period in Model (4).

The empirical results are provided in Table 8. As was expected by H6, the estimated coefficients of the Promotional Period in (1) and (2) are significantly positive ($\beta_{\text{Promotional Period}} = 0.02$, *p-value* $< .01$, $\beta_{\text{Promotional Period}} = 0.02$, *p-value* $< .01$, respectively). The reviews posted by netizens tend to be more positively emotional, supporting H4. The R^2 is about 0.11 for both columns.

5.5. Review manipulation and recommendation type of reviews

After removing the reviews which cannot be classified into one of the topics, we group them into 4 major topics (Group names). Among the 4 major topics, the first topic is related to “recommendation.” If review manipulation exists, it will encourage customers to watch the movie. This could be done by manipulating the textual content to recommend the movie. This means that the reviews provided by netizens during the first two weeks are more likely to include recommendation-related

Table 6
The relationship between review manipulation and the average review ratings.

Variable	DV: *Diff in Average Review Ratings*	
	(1)	(2)
Promotional period	0.30***	0.30***
	(0.02)	(0.05)
Constant	−0.80***	−0.80***
	(0.16)	(0.23)
Monthly-FE	Included	Included
Day-of-the-week FE	Included	Included
Movie FE	Included	Included
Huber/white/sandwich estimator	Included	–
Clustered sandwich estimator	–	Included
R-squared	0.40	0.40
# of observations	3350	3350

Notes: ***/** indicates significance at the 1 %/5 % level. Standard errors are in parentheses.

content in the reviews. To test this, we use Model (5), where the dependent variable is the difference in the average ratio of recommendation-related reviews between netizen-type reviews and viewer-type reviews:

$$Model\ (5): \frac{(\#of\ \ recommendation - related\ \ reviews)_{1t}^{Netizen}}{n_{1t}} - \frac{(\#of\ \ recommendation - related\ \ reviews)_{2t}^{Viewer}}{n_{2t}} = \alpha + X_{it}{*}\beta + \sum \gamma_l{*}Controls_{it}^{l} + \delta_i + \varepsilon_{it}$$

where n_{1t}: the number of netizens' reviews per day and n_{2t}: the number of viewers' reviews per day. If there is an increase in the ratio of recommendation-related reviews during the first two weeks, it would be confirmed by the positive coefficient (β) of the "Promotional Period".

The empirical results are provided in Table 9. Columns (1) and (2) include the same variables as those of previous Tables. Both of the estimated coefficients of Promotional Period in (1) and (2) are significantly positive ($\beta_{\text{Promotional Period}} = 0.03$, *p-value* < .01, $\beta_{\text{Promotional Period}} = 0.03$, *p-value* < .01, respectively). During the first two weeks, the reviews posted by netizens are more likely to contain recommendation-related content in the reviews, supporting H5. The R^2 is about 0.07 for both columns.

5.6. Review manipulation and product-specific content reviews

Unlike recommendation-related reviews, review manipulators would be limited to providing detailed information about the movie because movie-specific information are more likely to be provided by viewers, not by others. This would lead to a decrease in "movie specific content" related reviews posted by netizens during the first two weeks. It would be confirmed by the negative coefficient of the Promotional Period. To test this, we employ Model (6), where the dependent variable is the difference in the average ratio of movie specific content-related reviews between netizen-type reviews and viewer-type reviews:

$$Model\ (6): \frac{(\#of\ \ movie\ specific\ content\ related\ reviews)_{1t}^{Netizen}}{n_{1t}} - \frac{(\#of\ \ movie\ specific\ content\ related\ reviews)_{2t}^{Viewer}}{n_{2t}} = \alpha + X_{it}{*}\beta + \sum \gamma_l{*}Controls_{it}^{l} + \delta_i + \varepsilon_{it}$$

where n_{1t}: the number of netizens' reviews per day and n_{2t}: the number of viewers' reviews per day. If there is a decrease in the ratio of movie-specific content-related reviews during the first two weeks, it is a signal for review manipulation. It would be confirmed by the negative coefficient (β) of the "Promotional Period" in Model 6.

Table 10 shows the empirical results. As opposed to Table 8, the estimated coefficients of columns (1) and (2) are both significantly negative ($\beta_{\text{Promotional Period}} = -0.02$, *p-value* < .01, $\beta_{\text{Promotional Period}} = -0.02$, *p-value* < .01, respectively), supporting H6. During the first two weeks, the reviews posted by netizens are less likely to contain movie-specific content. The R^2 is about 0.06 for both columns.

5.7. Robustness check

As a robustness check, we use alternative estimation approaches (Table 11). Since the daily data used for this research is panel data, we employ fixed effects and random effects models. In the first six columns of the Table, we estimate Models (1)–(6) using the fixed-effects models. In the next six columns, we estimate Model (1)–(6) using the random-effects models. The estimation results are basically the same as those in the previous Tables. In the case of the fixed-effects models, the estimated coefficients are significantly positive except for that of Model (5) (Model (1): $\beta_{\text{Promotional Period}} = 1.15$, *p-value* < .01, Model (2): $\beta_{\text{Promotional Period}} = 0.30$, *p-value* < .01, Model (3): $\beta_{\text{Promotional Period}} = 0.01$, *p-value* < .01, Model (4): $\beta_{\text{Promotional Period}} = 0.03$, *p-value* < .01, Model (5): $\beta_{\text{Promotional Period}} = -0.02$, *p-value* < .05, Model (6): $\beta_{\text{Promotional Period}} = 0.02$, *p-value* < .05, respectively). During the first two weeks, the number of reviews posted by netizens increases, the average review ratings by netizens increases, there is an increase in the noun ratio of the reviews posted by netizens, there is an increase in "recommendation"-related reviews posted by netizens, there is a decrease in "product-content"-related reviews posted by netizens, and there is an increase in positive emotional reviews posted by netizens. In the case of the random-effects models, the same phenomena are confirmed. The estimated coefficients are all significant and positive except for that of Model (5) (Model (1): $\beta_{\text{Promotional Period}} = 1.15$, *p-value* < .01, Model (2): $\beta_{\text{Promotional Period}} = 0.30$, *p-value* < .01, Model (3): $\beta_{\text{Promotional Period}} =$

Table 7
The relationship between review manipulation and the average noun ratios.

Variable	DV: *Diff in the ratios of object words*	
	(1)	(2)
Promotional period	0.01***	0.01***
	(0.00)	(0.00)
Constant	−0.01**	−0.01**
	(0.00)	(0.00)
Monthly-FE	Included	Included
Day-of-the-week FE	Included	Included
Movie FE	Included	Included
Huber/white/sandwich estimator	Included	–
Clustered sandwich estimator	–	Included
R-squared	0.09	0.09
# of observations	3350	3350

Notes: ***/** indicates significance at the 1 %/5 % level. Standard errors are in parentheses.

Table 8
The relationship between review manipulation and positive emotional reviews.

Variable	DV: *Diff in the ratios of positive emotional reviews*	
	(1)	(2)
Promotional period	0.02***	0.02***
	(0.00)	(0.00)
Constant	−0.09	−0.09**
	(0.05)	(0.04)
Monthly-FE	Included	Included
Day-of-the-week FE	Included	Included
Movie FE	Included	Included
Huber/white/sandwich estimator	Included	–
Clustered sandwich estimator	–	Included
R-squared	0.11	0.11
# of observations	3350	3350

Notes: ***/** indicates significance at the 1 %/5 % level. Standard errors are in parentheses.

Table 9
The relationship between review manipulation and recommendation-related reviews.

Variable	DV: *Diff in the ratios of recommendation-related reviews*	
	Model (1)	Model (2)
Promotional period	0.03***	0.03***
	(0.00)	(0.00)
Constant	0.04	0.04
	(0.03)	(0.03)
Monthly-FE	Included	Included
Day-of-the-week FE	Included	Included
Movie FE	Included	Included
Huber/white/sandwich estimator	Included	–
Clustered sandwich estimator	–	Included
R-squared	0.07	0.07
# of observations	3350	3350

Notes: ***/** indicates significance at the 1 %/5 % level. Standard errors are in parentheses.

0.01, *p-value* $< .01$, Model (4): $\beta_{\text{Promotional Period}} = 0.03$, *p-value* $< .01$, Model (5): $\beta_{\text{Promotional Period}} = -0.02$, *p-value* $< .05$, Model (6): $\beta_{\text{Promotional Period}} = 0.02$, *p-value* $< .05$, respectively).

6. Discussion

The development of digital technologies has increased the likelihood of online review manipulation, particularly for short life-cycle products. Since manipulated reviews undermine consumer decision making and the relationship between consumers and firms (Wu et al., 2020), developing effective methods to detect manipulation in online reviews is particularly critical and urgent. To address this issue, this study aims to investigate online reviews' content, thus revealing reviewers' intentions. We analyze both verbal and nonverbal features that proxy reviewers' intentions, reflected in eWOM, which help classify likely fake reviews vs. authentic reviews. Using a unique dataset gathered from the South Korean online platform Naver.com, we identify a series of differences between the two types of reviews (i.e., authentic vs. likely fake reviews). For nonstrategic behaviors (nonverbal cues), we estimate posting time and ratings of reviews. Our findings reveal that review manipulation is concentrated in the early stage of product diffusion (manipulation dynamics). Such manipulation contributes to increasing review ratings. In the perspective of the intentional behavior reflected in textual content, we examine review informativeness, topics, and sentiments reflected in the textual content. We find that likely fake reviews are more informative and positive. Moreover, those reviews are less likely to contain product-specific content-related information but more likely to focus on the recommendation-related topic.

Table 10
The relationship between review manipulation and specific content related reviews.

Variable	DV: *Diff in the ratios of movie specific content-related reviews*	
	(1)	(2)
Promotional period	−0.02***	−0.02***
	(0.00)	(0.00)
Constant	−0.05	−0.05
	(0.04)	(0.03)
Monthly-FE	Included	Included
Day-of-the-week FE	Included	Included
Movie FE	Included	Included
Huber/white/sandwich estimator	Included	–
Clustered sandwich estimator	–	Included
R-squared	0.06	0.06
# of observations	3350	3350

Notes: ***/** indicates significance at the 1 %/5 % level. Standard errors are in parentheses.

6.1. Theoretical contributions

This research generates several theoretical contributions with respect to online review manipulation, fake reviews, and eWOM literature. First, this study innovatively applies interpersonal deception theory (IDT), that emphasizes the importance of both verbal and nonverbal features to reveal reviewers' purposes, to fake review detection. In online environments, nonverbal cues such as facial expression, body gestures, and tone of voice in face-to-face communication are limited to detect deception. Consequently, prior literature mainly focuses on verbal cues to explore online review manipulation (Hajek et al., 2023; Moon et al., 2021; Wang et al., 2012). In this study, we innovatively incorporate both verbal and nonverbal cues that help reveal online reviewers' intentions and enable to detect review manipulation. Our results suggest that both features have strong influence on fake review detection.

Second, to the best of our knowledge, this study is the first to consider topic differences between likely fake vs. authentic reviews to examine in-depth reviewers' intentions. Our research finds that likely fake reviews are more likely to contain recommendation-related topics in their text than authentic reviews, in the context of experience goods. Recommendations of experience goods such as movies highly influence consumers' decision-making (Flavián et al., 2016; Senecal and Nantel, 2004). Therefore, review manipulators are more likely to focus on the recommendation in their reviews for their promotion. On the other hand, since manipulators fabricate stories about the product, they are likely to have less familiarity with the details of the product (Burgoon et al., 2003; DePaulo et al., 2003), and they will less likely include detailed information about movies in their reviews. These findings contribute to extending the line of studies that have examined fake reviews.

Third, this paper provides a different point of view on text informativeness in the context of online review manipulation. Our findings show that the ratio of nouns is higher in likely fake reviews than in authentic reviews. It is a counterintuitive result compared with previous literature: for instance, Banerjee and Chua (2017) show that authentic reviews have more nouns than manipulated reviews. The difference may be related to the type of data collected. While Banerjee and Chua (2017) collect their data by recruiting participants to generate ad hoc manipulated reviews (in an experimental way), our study deploys a large dataset covering a much larger amount of manipulated reviews. However, manual annotation of fake reviews is inherently problematic because participants do not have the same psychological state of mind as real fake reviewers (Mukherjee et al., 2013). Furthermore, among manipulated reviews, some of them have the potential to be produced by automated software that is trained by AI algorithms to avoid fake review detection systems. Therefore, differences in data between our study and previous research generates different findings.

6.2. Practical and managerial implications

Our findings provide implications for online review readers and consumers, reviewers, online platform managers, and companies dealing with eWOM and verified/unverified online reviews. First, online review readers and consumers can determine which reviews are likely to be credible, which could help them make more informed decisions. More specifically, since our findings suggest that review manipulation may be concentrated during the promotional period, customers of short life cycle products need to be more careful when reading eWOM during this period, so they can avoid misleading information before making their purchase decisions. They also need to understand that likely fake reviews tend to be more positively emotional.

From the perspective of reviewers, some customers write reviews to help others make better and more informed decisions (Sundaram et al., 1998). They want to provide credible and helpful eWOM to peers and other individuals, in line with the altruism argument proposed by

Table 11
Robustness check.

Variables	Fixed-effects regression						Random-effects regression					
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
Promotional period	1.15***	0.30***	0.01***	0.03***	−0.02**	0.02**	1.15***	0.30***	0.01***	0.03***	−0.02**	0.02**
	(0.03)	(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.00)	(0.01)	(0.01)	(0.01)
Constant	−0.40***	−0.43***	−0.01***	0.01	−0.05	−0.06**	−0.50***	−0.80***	−0.01***	0.04	−0.06	−0.09
	(0.11)	(0.11)	(0.00)	(0.02)	(0.03)	(0.03)	(0.17)	(0.18)	(0.00)	(0.04)	(0.04)	(0.05)
Monthly-FE	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Day-of-the-week FE	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Movie FE	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
R-squared	0.33	0.05	0.03	0.01	0.01	0.02	–	–	–	–	–	–
Wald chi2	–	–	–	–	–	–	4244.83	2204.82	320.54	255.55	232.79	405.03
# of obs.	3350	3350	3350	3350	3350	3350	3350	3350	3350	3350	3350	3350

Notes: ***/** indicates significance at the 1 %/5 % level. Standard errors are in parentheses.

Hennig-Thurau et al. (2004). Our findings shed light on the specific traits of review manipulation that reviewers should avoid when creating a review. This way, they can differentiate their reviews from likely fake reviews. For instance, to write credible reviews, our study suggests that they should not express too many positive emotions in their reviews. Furthermore, our findings show that it would be desirable for reviewers writing about short life cycle products to include content-related information rather than recommendation-related information in their reviews.

Third, our findings provide online platform managers with some insights to develop an ad hoc index that can proxy review manipulation. Since source credibility is a critical factor for customers to determine the usefulness of eWOM in online platforms (Ayeh, 2015; Ayeh et al., 2013; Mariani and Borghi, 2021), online platform managers can leverage our analysis to label reviews as “mostly authentic” or “mostly suspicious” to increase their usefulness and credibility. This labeling would help prospective customers deviate from information overload and reduce efforts related to the purchase decision-making process. This would contribute to improving the usefulness of reviews for customers. Moreover, platform managers might use our analysis to reduce the number of manipulated reviews on their website, which may increase their review credibility, thus offering less biased information to online customers. Furthermore, platform managers might cross-check if our findings are consistent with their practices and inform their practices in relation to eWOM hosting on their platforms. For example, Yelp.com developed its algorithmic indicator to identify fake reviews and other companies are trying to monitor generative-AI produced content and misinformation (Mariani et al., 2023); we believe that Naver.com could develop its algorithm by incorporating our findings.

Lastly, companies might find it difficult to listen to actual customers' voices due to the explosion of eWOM in the digital realm (Akter et al., 2020). To reduce their burden, they can leverage our analysis to find information on customers who provide authentic evaluations. Subsequently, they may find and connect with the target audience for their products or services. Our findings help companies hear “authentic” reviewers' voices, both praises and complaints, toward their products or services. With a massive number of posted reviews, it would be difficult for them to improve the quality of their products or services and innovate them (Mariani and Wamba, 2020) when they listen to the complaints from manipulated reviews. By incorporating preferences and complaints from actual customers, they may improve the quality of their products and services. Furthermore, companies can select and emphasize authentic reviews to potential customers to gain customer trust in their products or services, which can influence potential customer purchase decisions. For instance, they can use authentic reviews as powerful social proof to enhance trust and loyalty toward their brands.

6.3. Limitations and future research directions

Like other research, this study also has some limitations. First, though this study extensively examines the differences between authentic and likely fake reviews, the approach used for this research is an indirect method (Mayzlin et al., 2014) which uses the different levels of vulnerability to review manipulation. To examine the exact manipulation in eWOM, identifying the underlying mechanism of review manipulation is necessary, so experiential methods that allow direct comparison between fake and authentic reviews are recommended in future research. Second, though this study found systematic differences between authentic and likely fake reviews, it cannot answer how the differences would affect consumer decision-making. Future research may investigate the impact of each type of reviews on customers' decision-making process such as movie revenues with additional datasets. Third, this research provides several important factors that indicate reviewers' intentions. Future research can provide novel factors related to review-writing intentions such as emotions toward products.

CRediT authorship contribution statement

Jong Min Kim: Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Data curation, Formal analysis, Investigation. **Keeyeon Ki-cheon Park:** Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Data curation, Formal analysis, Investigation. **Marcello Mariani:** Writing – original draft, Writing – review & editing, Conceptualization, Validation, Investigation, Supervision. **Samuel Fosso Wamba:** Writing – original draft, Writing – review & editing, Validation.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

References

Akter, S., Motamarri, S., Hani, U., Shams, R., Fernando, M., Babu, M.M., Shen, K.N., 2020. Building dynamic service analytics capabilities for the digital marketplace. J. Bus. Res. 118, 177–188.
Allcott, H., Gentzkow, M., 2017. Social media and fake news in the 2016 election. J. Econ. Perspect. 31 (2), 211–236.
Ananthakrishnan, U.M., Li, B., Smith, M.D., 2020. A tangled web: should online review portals display fraudulent reviews? Inf. Syst. Res. 31 (3), 950–971.
Arafat, H., Elawady, R.M., Barakat, S., Elrashidy, N.M., 2014. Different feature selection for sentiment classification. Int. J. Inf. Sci. Intell. Syst. 1 (3), 137–150.

Ayeh, J.K., 2015. Travellers' acceptance of consumer-generated media: an integrated model of technology acceptance and source credibility theories. Comput. Hum. Behav. 48, 173–180.
Ayeh, J.K., Au, N., Law, R., 2013. "Do we believe in TripAdvisor?" examining credibility perceptions and online travelers' attitude toward using user-generated content. J. Travel Res. 52 (4), 437–452.
Banerjee, S., Chua, A.Y., 2014. A theoretical framework to identify authentic online reviews. Online Inf. Rev. 38 (5), 634–649.
Banerjee, S., Chua, A.Y., 2017. Theorizing the textual differences between authentic and fictitious reviews: validation across positive, negative and moderate polarities. Internet Res. 27 (2), 321–337.
Barbado, R., Araque, O., Iglesias, C.A., 2019. A framework for fake review detection in online consumer electronics retailers. Inf. Process. Manag. 56 (4), 1234–1244.
Buller, D.B., Burgoon, J.K., 1996. Interpersonal deception theory. Commun. Theory 6 (3), 203–242.
Burgoon, J.K., Buller, D.B., Ebesu, A.S., Rockwell, P., 1994. Interpersonal deception: V. Accuracy in deception detection. Commun. Monogr. 61 (4), 303–325.
Burgoon, J.K., Blair, J.P., Qin, T., Nunamaker, J.F., 2003, June. Detecting deception through linguistic analysis. In: International Conference on Intelligence and Security Informatics. Springer, Berlin, Heidelberg, pp. 91–101.
Calheiros, A.C., Moro, S., Rita, P., 2017. Sentiment classification of consumer-generated online reviews using topic modeling. J. Hosp. Mark. Manag. 26 (7), 675–693.
Cano-Marin, E., Mora-Cantallops, M., Sanchez-Alonso, S., 2023. The power of big data analytics over fake news: a scientometric review of Twitter as a predictive system in healthcare. Technol. Forecast. Soc. Chang. 190, 122386.
Casaló, L.V., Flavián, C., Guinalíu, M., Ekinci, Y., 2015. Avoiding the dark side of positive online consumer reviews: enhancing reviews' usefulness for high risk-averse travelers. J. Bus. Res. 68 (9), 1829–1835.
Chen, Y.R., Chen, H.H., 2015. Opinion spam detection in web forum: a real case study. In: Proceedings of the 24th International Conference on World Wide Web, pp. 173–183. May.
Connors, L., Mudambi, S.M., Schuff, D., 2011. Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness. In: 2011 44th Hawaii International Conference on System Sciences. IEEE, pp. 1–10. January.
Dennis, A.R., Fuller, R.M., Valacich, J.S., 2008. Media, tasks, and communication processes: a theory of media synchronicity. MIS Q. 575–600.
DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H., 2003. Cues to deception. Psychol. Bull. 129 (1), 74.
Duncan, B., Zhang, Y., 2015. "Neural networks for sentiment analysis on Twitter," 2015 IEEE 14th Duncan, B., & Zhang, Y. Neural networks for sentiment analysis on Twitter. In: 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC). IEEE, pp. 275–278. July.
Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., Wright, R., 2023. So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int. J. Inf. Manag. 71, 102642.
Flavián, C., Gurrea, R., Orús, C., 2016. The impact of recommendations on the cross-channel shopping behavior. In: Rediscovering the Essentiality of Marketing. Springer, Cham, pp. 295–301.
Gartner, 2019. Gartner identifies Top Digital Experience Trends for 2020. https://www.gartner.com/en/newsroom/press-releases/2019-10-29-gartner-identifies-top-digital-experience-trends-for accessed June 18, 2022.
Goldman, A., 1982. Short product life cycles: implications for the marketing activities of small high-technology companies. R&D Manag. 12 (2), 81–90.
Grazioli, S., Jarvenpaa, S.L., 2003. Consumer and business deception on the internet: content analysis of documentary evidence. Int. J. Electron. Commer. 7 (4), 93–118.
Gurun, U.G., Butler, A.W., 2012. Don't believe the hype: local media slant, local advertising, and firm value. J. Financ. 67 (2), 561–598.
Hajek, P., Hikkerova, L., Sahut, J.M., 2023. Fake review detection in e-commerce platforms using aspect-based sentiment analysis. J. Bus. Res. 167, 114143.
Heinzman, A., 2019. How fake reviews are-manipulating you online. https://www.howtogeek.com/407521/how-fake-reviews-are-manipulating-you-online/.
Hennig-Thurau, T., Gwinner, K.P., Walsh, G., Gremler, D.D., 2004. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? J. Interact. Mark. 18 (1), 38–52.
Hu, N., Bose, I., Koh, N.S., Liu, L., 2012. Manipulation of online reviews: an analysis of ratings, readability, and sentiments. Decis. Support. Syst. 52 (3), 674–684.
Huang, Y.C., Wu, Y.C.J., Wang, Y.C., Boulanger, N.C., 2011. Decision making in online auctions. Manag. Decis. 49 (5), 784–800.
Indriani, F., Nugrahadi, D.T., 2016. Comparison of Naive Bayes smoothing methods for twitter sentiment analysis. In: 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, pp. 287–292. October.
Infante, D.A., Rancer, A.S., Avtgis, T.A., 2010. Contemporary Communication Theory. Kendall Hunt, Dubuque, IA, p. 578.
Jindal, N., Liu, B., Lim, E.P., 2010. Finding unusual review patterns using unexpected rules. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1549–1552. October.
Kahn, K.F., Kenney, P.J., 2002. The slant of the news: how editorial endorsements influence campaign coverage and citizens' views of candidates. Am. Polit. Sci. Rev. 96 (2), 381–394.
Kumar, N., Venugopal, D., Qiu, L., Kumar, S., 2018. Detecting review manipulation on online platforms with hierarchical supervised learning. J. Manag. Inf. Syst. 35 (1), 350–380.
Kumar, N., Venugopal, D., Qiu, L., Kumar, S., 2019. Detecting anomalous online reviewers: an unsupervised approach using mixture models. J. Manag. Inf. Syst. 36 (4), 1313–1346.
Kim, J.M., Lee, E., Mariani, M.M., 2021. The influence of launching mobile channels on online customer reviews. J. Bus. Res. 137, 366–378.
Kim, J.M., Park, K.K.C., Mariani, M.M., 2023. Do online review readers react differently when exposed to credible versus fake online reviews? J. Bus. Res. 154, 113377.
Kumar, A., Shankar, A., Behl, A., Arya, V., Gupta, N., 2023. Should I share it? Factors influencing fake news-sharing behaviour: a behavioural reasoning theory perspective. Technol. Forecast. Soc. Chang. 193, 122647.
Lappas, T., 2012. Fake reviews: the malicious perspective. In: International Conference on Application of Natural Language to Information Systems. Springer, Berlin, Heidelberg, pp. 23–34. June.
Lau, R.Y., Liao, S.Y., Kwok, R.C.W., Xu, K., Xia, Y., Li, Y., 2012. Text mining and probabilistic language modeling for online review spam detection. ACM Trans. Manag. Inf. Syst. (TMIS) 2 (4), 1–30.
Luca, M., Zervas, G., 2016. Fake it till you make it: reputation, competition, and Yelp review fraud. Manag. Sci. 62 (12), 3412–3427.
Ludwig, S., Van Laer, T., De Ruyter, K., Friedman, M., 2016. Untangling a web of lies: exploring automated detection of deception in computer-mediated communication. J. Manag. Inf. Syst. 33 (2), 511–541.
Ma, H., Kim, J.M., Lee, E., 2019. Analyzing dynamic review manipulation and its impact on movie box office revenue. Electron. Commer. Res. Appl. 35, 100840.
Maas, A., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C., 2011. Learning word vectors for sentiment analysis. In: Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, pp. 142–150.
Mariani, M., Borghi, M., 2021. Are environmental-related online reviews more helpful? A big data analytics approach. Int. J. Contemp. Hosp. Manag. 33 (6), 2065–2090.
Mariani, M.M., Machado, I., Magrelli, V., Dwivedi, Y.K., 2023. Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. Technovation 122, 102623.
Mariani, M.M., Wamba, S.F., 2020. Exploring how consumer goods companies innovate in the digital age: the role of big data analytics companies. J. Bus. Res. 121, 338–352.
Mayzlin, D., Dover, Y., Chevalier, J., 2014. Promotional reviews: an empirical investigation of online review manipulation. Am. Econ. Rev. 104 (8), 2421–2455.
McCornack, S.A., 1992. Information manipulation theory. Commun. Monogr. 59 (1), 1–16.
Moon, S., Kim, M.Y., Iacobucci, D., 2021. Content analysis of fake consumer reviews by survey-based text categorization. Int. J. Res. Mark. 38 (2), 343–364.
Mukherjee, A., Venkataraman, V., Liu, B., Glance, N., 2013. What yelp fake review filter might be doing?. In: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 7, No. 1, pp. 409–418.
Nigam, K., Hurst, M., 2004. Towards a robust metric of opinion. In: AAAI Spring Symposium on Exploring Attitude and Affect in Text, Vol. 598603. American Association for Artificial Intelligence, Menlo Park, CA, USA. July.
Ozbay, F.A., Alatas, B., 2020. Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A 540, 123174.
Pang, B., Lee, L., 2004. A Sentimental Education: sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts arXiv preprint cs/0409058.
Rayson, P., Wilson, A., Leech, G., 2002. Grammatical word class variation within the British National Corpus sampler. In: New Frontiers of Corpus Research. Brill, pp. 295–306.
Ren, Y., Ji, D., 2017. Neural networks for deceptive opinion spam detection: an empirical study. Inf. Sci. 385, 213–224.
Salminen, J., Kandpal, C., Kamel, A.M., Jung, S.G., Jansen, B.J., 2022. Creating and detecting fake reviews of online products. J. Retail. Consum. Serv. 64, 102771.
Senecal, S., Nantel, J., 2004. The influence of online product recommendations on consumers' online choices. J. Retail. 80 (2), 159–169.
Shan, G., Zhou, L., Zhang, D., 2021. From conflicts and confusion to doubts: examining review inconsistency for fake review detection. Decis. Support. Syst. 144, 113513.
Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H., 2017. Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor. Newslett. 19 (1), 22–36.
Shukla, A.D., Gao, G., Agarwal, R., 2021. How digital word-of-mouth affects consumer decision making: evidence from doctor appointment booking. Manag. Sci. 67 (3), 1546–1568.
Spence, M., 1978. Job market signaling. In: Uncertainty in Economics. Academic Press, pp. 281–306.
Sundaram, D.S., Mitra, K., Webster, C., 1998. Word-of-mouth Communications: A Motivational Analysis. ACR North American Advances.
Wang, G., Xie, S., Liu, B., Yu, P.S., 2012. Identify online store review spammers via social review graph. ACM Trans. Intell. Syst. Technol. 3 (4), 1–21.
Wang, X., Zhang, X., Jiang, C., Liu, H., 2018. Identification of fake reviews using semantic and behavioral features. In: 2018 4th International Conference on Information Management (ICIM). IEEE, pp. 92–97. May.
Wu, Y., Ngai, E.W., Wu, P., Wu, C., 2020. Fake online reviews: literature review, synthesis, and directions for future research. Decis. Support. Syst. 132, 113280.
Yousaf, S., Kim, J.M., 2023. Did COVID-19 change preferences for hygiene-related service attributes as satisfiers and dissatisfiers? An analysis of textual content of online hotel reviews. J. Hosp. Tour. Manag. 56, 264–271.
Zaman, M., Vo-Thanh, T., Nguyen, C.T., Hasan, R., Akter, S., Mariani, M., Hikkerova, L., 2023. Motives for posting fake reviews: Evidence from a cross-cultural comparison. J. Bus. Res. 154, 113359.
Zhang, X., Ghorbani, A.A., 2020. An overview of online fake news: characterization, detection, and discussion. Inf. Process. Manag. 57 (2), 102025.

Zhang, Z., Gupta, B.B., 2018. Social media security and trustworthiness: overview and new direction. Futur. Gener. Comput. Syst. 86, 914–925.
Zhang, D., Zhou, L., Kehoe, J.L., Kilic, I.Y., 2016. What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. J. Manag. Inf. Syst. 33 (2), 456–481.
Zhou, L., Sung, Y.W., Zhang, D., 2013. Deception performance in online group negotiation and decision making: the effects of deception experience and deception skill. Group Decis. Negot. 22 (1), 153–172.

Jong Min Kim, Ph.D. is an Assistant Professor at the College of Social Science of Konkuk University, South Korea after having worked for College of Business and Public Management, Wenzhou-Kean University, China. His research interests include online word-of-mouth, management response, promotional reviews, mobile marketing, and corporate social responsibility. In terms of methodology, Dr. Kim focuses on quasi-experimental approaches based on Econometrics. His research has appeared in the *Journal of Interactive Marketing*, *Journal of Business Research*, *International Journal of Electronic Commerce*, *Tourism Management, and others*.

Keeyeon Ki-cheon Park is an Assistant Professor of Marketing at the Kedge Business School (France). His research interests lie in how to use digital data and machine learning to uncover consumer preferences. He mainly focuses on applying innovative technique to solve problems in the marketing fields.

Marcello Mariani is a Professor of Management at the University of Reading (UK) and University of Bologna (Italy), member of the Henley Center for Entrepreneurship, IEEE, the Academy of Management and the European Institute for Advanced Studies in Management. His current research interests include big data and analytics, eWOM, digital business models, AI, IoT, automation and coopetition strategies in service industries. His researches have been published in *Technological Forecasting and Social Change*, *Harvard Business Review*, *MIT Sloan Management Review*, *Journal of Business Research*, *IEEE Access*, *Industrial Marketing Management*, *Journal of Advertising*, *Psychology & Marketing*, *Industrial and Corporate Change*, *Long Range Planning*, *International Journal of Electronic Commerce*, *Tourism Management*, *Annals of Tourism Research*, *Journal of Travel Research*, *International Journal of Contemporary Hospitality Management*, *International Journal of Hospitality Management*, *European Management Journal*, *European Accounting Review*, *International Studies in Management and Organizations*, *Journal of Destination Management and Marketing*, and more.

Samuel Fosso Wamba, PhD, is Associate Dean for Research at TBS Education, France, and a Distinguished Visiting Professor at The University of Johannesburg, South Africa. He earned his Ph.D. in industrial engineering at the Polytechnic School of Montreal, Canada. His current research focuses on the business value of information technology, blockchain, artificial intelligence for business, business analytics, and big data. He is among the 2 % of the most influential scholars globally based on the Mendeley database that includes 100,000 top scientists for 2020, 2021 and 2022. He ranks in Clarivate Analytics' Highly Cited Researchers List, which consists of the top 1 % of the “world's most impactful scientific researchers,” for 2020, 2021 and 2022, and in CDO Magazine's Leading Academic Data Leaders 2021.