Voices in Motion: Sentiment Intensity and Thematic Evolution Across Game Release Stages*

Elena Kim and Jaewon Choi†

Soonchunhyang University, Asan, South Korea
{voogiejam22,jaewonchoi}@sch.ac.kr

Abstract

The rapid growth of the gaming industry underscores the importance of understanding player perceptions in user-generated reviews. This study examines Steam reviews across three stages of the game lifecycle—pre-release, release day, and one-month post-release—using sentiment analysis (VADER) and topic modeling (LDA, DTM). Results show clear shifts in sentiment polarity and thematic focus: anticipation dominates pre-release reviews, while unmet expectations and technical flaws drive negativity post-release. Extreme sentiments cluster around gameplay immersion, monetization, and defects, with a notable rise in extreme negative reviews one month after release. Gameplay duration strongly correlates with sentiment intensity, indicating deeper engagement among invested players. Grounded in Expectancy-Confirmation Theory, UGC, and eWOM, the study advances understanding of digital consumer behavior and offers practical guidance for developers and platforms to anticipate and manage player sentiment throughout the game lifecycle.

1 Introduction

The global gaming industry has experienced sustained and vigorous growth over the past decade. Statistics indicate that by the end of 2024 the global gaming market had reached approximately $177.9 billion[1], and the total number of players worldwide exceeded 3.4 billion[1]. Such a vast market entails intense competition, and the success or failure of a game product depends on word-of-mouth dissemination and community influence among player groups. In this context, various digital distribution platforms have rapidly emerged and amassed massive user bases. Among them, Steam, as one of the world's largest digital distribution platforms for PC games, not only boasts hundreds of millions of active players but also aggregates over fifty thousand games [2]. Players on Steam are free to write reviews and share their gaming experiences. Consequently, player reviews have become an

*Proceedings of The 2025 IFIP WG 8.4 International Symposium on E-Business Information Systems Evolution (EBISION 2025), Article No. 17, December 16-18, 2025, Sapporo, Japan.

†Corresponding author

important form of electronic word-of-mouth (eWOM), providing decision-making references for other potential players and valuable data sources for developers to understand user feedback.

The sentiments and opinions expressed in player reviews hold significant importance for all stakeholders in the gaming industry. On one hand, review sentiment directly reflects player satisfaction and game quality, and sentiment analysis can quantitatively characterize the subjective tendencies in reviews [3]. Sentiment polarity is typically divided into two categories—positive and negative [3]and reflects the overall direction of player evaluations of the game. Emotional reactions expressed by players in reviews are also crucial indicators of gaming experience. Research indicates that the intensity of emotional reactions is often closely related to the degree of player involvement and experience quality [4]. When players invest a high level of attention and expectations in a game, if the actual experience greatly exceeds or falls short of those expectations, the Disconfirmation of Expectations Theory suggests this will trigger strong positive or negative emotional reactions [4]. On the other hand, as online word-of-mouth, player reviews significantly influence other consumers' purchasing decisions and shape community public [5]. Studying the polarity and intensity of sentiments in player reviews not only helps in understanding a game's acceptance among players but also provides data support for game marketing and continuous operation. Therefore, performing sentiment analysis and mining on game reviews has important academic value and practical significance.

Despite growing research interest in text analysis of game player reviews in recent years, existing literature still has shortcomings. Research on how review sentiment changes over time remains insufficient. Existing work has found that most user reviews appear soon after an application is released [6], and then review frequency rapidly declines over time. This suggests that we should pay attention to user feedback differences across different stages of the product lifecycle. However, currently few studies delve into comparing how sentiment intensity and types evolve in player reviews for games during the pre-release hype period, at official release, and at a certain time after release (e.g., one month later). The temporal evolution of discussed topics in reviews has not been fully uncovered. Player reviews often cover multiple aspects, such as feature requests, user experience, bug complaints, etc. [6]. Discussion focal points may shift within different time windows, but existing research lacks systematic exploration of this dynamic process. The expression of extreme sentiments in reviews and their association with discussion topics have also not received sufficient attention. Research on mobile apps has shown that user negative reviews often involve specific issues like app crashes or missing features [7], suggesting that certain topics may trigger stronger emotional responses. However, current literature contains little work that quantitatively tests whether extreme positive or negative sentiments in game reviews are significantly associated with specific discussion topics. Finally, both the industry and player communities are concerned whether player sentiment "cools off" or reverses (e.g., word-of-mouth reversal) soon after a game's release. There is a lack of empirical research to verify whether the proportion of extremely negative sentiments in reviews one month after release is significantly higher than that before and at release.

Based on the above research gaps and practical needs, this study adopts a lifecycle perspective combined with sentiment analysis and topic mining to conduct an in-depth analysis of player reviews for different game genres on the Steam platform, focusing on three key stages: before release, at release, and one month after release. Through this study, we will explore how players review sentiment change and content dimensions over time, filling the gap in literature regarding the association between sentiment polarity and discussion topics. Specifically, this paper will address the following four research questions:

RQ1: On the Steam platform, do different game genres exhibit significant differences in the dominant sentiment types and intensities in player reviews before release, at release, and one month after release?

RQ2: Within the time windows before release, at release, and one month after release, how do the main topics discussed in player reviews dynamically change over time?

RQ3: Is there a significant association between the expression of extreme sentiments (extreme positive or extreme negative) in reviews and specific discussion topics?

RQ4: In player reviews one month after game release, is the proportion of extreme negative sentiments significantly higher than at pre-release and release stages?

2 Literature Review

2.1 Sentiment Analysis and Extreme Emotional Expression

Sentiment analysis is one of the key techniques in text mining that has been widely used for studying reviews overall. Two approaches are common: lexicon and rule-based methods, and machine learning or deep learning models. For lexicon-based methods, well-known example is the VADER algorithm [8], which is often applied to social media and review data. It assigns scores from –1 (very negative) to +1 (very positive) and considers features like negations, intensifiers, or punctuation marks. In the gaming context, researchers have also trained classification models to label reviews as positive or negative. For instance, [9] worked with Weibo game data and showed that supervised learning methods are effective. [10] used Naive Bayes and decision trees on Steam reviews to examine sentiment patterns. Later studies compared different algorithms: [11] found that SVM outperformed logistic regression and neural networks on Steam and Metacritic reviews. More recently, deep learning approaches such as BERT have been applied, showing clear improvements in accuracy [12].

A separate issue in sentiment analysis is how extreme reviews are expressed. Players often exaggerate by using very strong adjectives, multiple exclamation marks, or writing in ALL CAPS. Tools like VADER capture these signals and assign higher absolute scores [8]. In the gaming world, this is visible in cases of so-called "review bombing," when large groups of players leave very negative ratings shortly after release. Such protests do not always reflect the actual game quality [13]. For example, The Last of Us 2 received an average user score of only 3.4 out of 10 on the first day, while critics rated it 9.5. Later, as more players shared their experiences, the user score stabilized around 5.8. This suggests that early extreme reviews may often come from a small, motivated group reacting to specific controversies rather than from a wider player base.

2.2 Lifecycle Evolution of Player Reviews

The sentiments and content of player reviews change dynamically across different stages of the product lifecycle. In the pre-release phase of a game, official trailers, beta versions, and media evaluations often create expectations and emotional tones among players for the upcoming release; many core players express high anticipation or concerns about a game not yet officially released via forums or social media. This pre-release public opinion can be viewed as the early-stage emotional state of the product lifecycle. Once the game is released, player reviews flood within a short time, marking the peak phase of the lifecycle. In this period, review sentiment may show large fluctuations and divergence: loyal fans are likely to give enthusiastic positive reviews at launch, while another group of players who feel disappointed by the actual experience may concentrate on negative reviews. For example, the mentioned "review bombing" phenomenon often occurs in the first few days after game release, when a minority of dissatisfied players dominate public opinion with extreme negative reviews, but later as more players delve deeper into the game, the average review score adjusts. [14] also revealed similar trends in their study of smartphone products: early reviewers, being mostly enthusiastic early users, tend to give higher ratings and more positive sentiments, while with more general users participating later, overall evaluations may become more rational and/or slightly decline.

This shows that user reviews have systematic differences in different product lifecycle stages (pre-release, launch, post-launch).

In the gaming domain, the temporal evolution of reviews has drawn particular attention. For example, [15] conducted an in-depth analysis of the case of No Man’s Sky, revealing that the dynamics of player reviews over time are closely related to game version updates. This study used Steam review data from different time periods since the game’s launch in 2016 and applied topic modeling and other methods to compare player focus before and after multiple version updates. Findings show that initial player reviews at release mainly focused on negative topics such as unmet expectations and feature omissions (many comments criticized the game content for not matching pre-release publicity, leading to negative reviews). This was related to discrepancies in promotion and quality issues at launch. Subsequently, as the development team released a series of patches and content updates, the tone of player reviews gradually shifted: initial complaints decreased, replaced by discussions of newly added features and performance improvements, and many players began to give positive feedback. The game’s word-of-mouth gradually changed from “mostly negative” to “mixed” and “mostly positive.” This evolution process shows that player review sentiment is a dynamic process; developer responses (such as timely updates and improvements) can effectively influence community sentiment trends.

Not only can improvements to the game itself bring changes in review sentiment, but the characteristics of player groups in different time windows also lead to changes in review tendencies. [16] studied the PUBG game and showed that before and after major version updates, there are significant differences in sentiment polarity and content focus in player reviews. After major updates, discussions about new features and game mechanic adjustments increase in intensity, while performance complaint discussions frequently mentioned in earlier versions decrease. This indicates that the emotional and topical focus of player reviews shows phase-specific characteristics over time and version changes. Similarly, research in domains such as mobile apps and movies has found that user evaluations often differ between the initial stage after product launch and a later stable period. In summary, considering the lifecycle evolution of user reviews is crucial for a comprehensive understanding of game word-of-mouth: it reminds us that analysis should not focus on data from a single static moment but should pay attention to trends and fluctuations in reviews over time. As the player base shifts from early adopters to the general audience, and as the game content matures, the emotional tone and content focus of reviews continue to change. After understanding how sentiment evolves over time, the further question arises: how exactly do the specific discussion topics (the review content) evolve with time? This is the focus of the next section.

2.3 Dynamic Topic Modeling of Review Content

Topic modeling is widely used to analyze players' reviews and define discussion topics. The Latent Dirichlet Allocation (LDA) model [17] extracts hidden themes and applies to large-scale game sets, revealing problems of players such as gameplay, graphics, and multiplayer. As the LDA does not capture temporal changes, [18] introduced the Dynamic Topic Model (DTM), which tracks how themes evolve over time. Research shows its value in various fields: [19] demonstrated a shift in focus in health reviews before and after COVID-19, while [15] found that 'No Man's Sky' reviews shifted from launch complaints to later approval after updates. Similarly, [16] used subject sensitivity analysis to track changes in PUBG discussions during major updates. Together, LDA and DTM reveal not only what players care about, but how these problems change over time.

Game genres and player characteristics are also one of the important factors, significantly shaping review sentiment and expression. Prior studies show that differences in themes, gameplay, and audiences produce distinct emotional distributions and levels of extremity in reviews. For instance, [20] analyzed over 350 million Steam reviews and found that VR games receive disproportionately positive evaluations, likely due to their novelty and immersive experience. Genre-based differences also emerge in review length: RPG reviews are typically the longest, while racing reviews are shorter; across genres, negative reviews are consistently longer than positive ones (median ~40 vs. ~19 words), suggesting players elaborate more on dissatisfaction than on praise.

Player behavior also further mediates sentiment. Positive reviews are often linked to longer playtimes (median ~6.8 hours), while negative reviews appear after shorter sessions (median ~2.1 hours). There is also proof of genre-specific deviations: tabletop simulations sometimes attract long-playtime negative reviews, while simulation games generate highly positive evaluations after extended engagement. Core players with substantial time or emotional investment may also provide both the most enthusiastic endorsements and harsh criticism, particularly when sequels diverge from expected mechanics.

Emotional tone also varies by genre. Shooter games and games with competitive mode often trigger negativity when balance issues arise, while puzzle games attract broader, more casual audiences, yielding milder sentiment. Horror games stimulate fear and tension, reflected in reviews, when "healing" games induce relaxation and overall positive affect. Player demographics and community dynamics further influence review style—for example, the discourse culture of MMOs differs from that of single-player titles.

In total, genre and player engagement both affect review sentiment and topic emphasis. Analysts must account for these contextual differences: core players often produce polarized reviews, while casual audiences lean toward neutral or moderately positive evaluations. Recognizing these patterns can guide researchers and developers in creating word-of-mouth management strategies for different game communities.

2.4 Supporting Theoretical Frameworks

For more comprehensive understanding of the emotional change and topic evolution mechanisms of Steam player reviews, it is necessary to draw on relevant theoretical frameworks. First, User-Generated Content (UGC) theory provides a macro-background. UGC refers to content created by ordinary users and published on the internet, consisting of reviews, posts, videos, and other forms. Player-written reviews on Steam are a typical form of UGC. UGC theory emphasizes that each user is both a consumer and producer of content, and their shared real experiences and opinions can influence others' perceptions and behaviors [21]. In the context of this study, Steam player reviews as UGC reflect the collective voice of the player community through their quantity and tendency. They have important reference value for potential consumers and game developers.

Electronic Word-of-Mouth (eWOM) theory focuses on the diffusion effect of user reviews on consumer decision-making. eWOM refers to online product or service evaluations published by consumers, which can positively (recommend) or negatively (discourage) influence others' purchase intentions [22]. Numerous studies have confirmed that the valence and intensity of online reviews significantly affect audience attitudes: negative reviews tend to attract more attention than positive ones, a phenomenon known as the "negativity bias." For example, consumers often consider negative reviews more valuable because they contain descriptions of problems that help them avoid possible failures. However, overly extreme reviews (whether positive or negative) can sometimes reduce their effectiveness — readers may question the objectivity of such reviews, suspecting enthusiastic praise might come from fan bias or disguised lobbying, while extreme negative reviews might be emotional venting or isolated cases. Therefore, eWOM theory suggests that we should pay attention to how sentiment intensity influences audience interpretation, which helps explain why some extreme reviews, despite strong emotions, may not be regarded as helpful.

Media Richness Theory, from a communication perspective, can explain differences in emotional expression in player reviews. According to this theory, different media vary in the richness of cues and emotions they can convey [23]. Face-to-face communication is a high-richness medium, capable of conveying rich cues like tone of voice and facial expression; in contrast, text reviews are a low-richness medium lacking nonverbal emotional cues. Steam reviews are primarily presented in text form, which means that players often resort to more extreme and direct language to accurately convey subjective emotions in this medium. For example, users may repeatedly use exclamation points, capital letters, or emoticons to strengthen their tone, ensuring readers grasp their strong likes or dislikes. This explains why exaggerated extreme emotional expressions are often seen in player reviews — the limitations of the text medium prompt users to amplify their language to enhance the “perceptibility” of their emotions. Media Richness Theory thus validates the earlier discussion on mechanisms of extreme emotional expression: when the medium cannot naturally carry rich emotional cues, content creators often strengthen message content to achieve the intended communication effect.

3 Analysis Process

The analysis in this study proceeds as follows. First, review data was collected and preprocessed. Then, sentiment analysis was conducted using VADER [8] to assign polarity scores and measure intensity. Next, we performed Topic Extraction, using Latent Dirichlet Allocation (LDA) [17] to identify major discussion themes and their distributions. To clarify the temporal variation, Dynamic Topic Model (DTM) [18] was applied, enabling observation of how topics evolved across time.

To capture dynamic sentiment–topic relationships, the review corpus was segmented into three periods: pre-release, release day, and one-month post-release. This temporal division allows systematic comparison of shifts in player sentiment and following concerns across the game lifecycle. Based on the outputs of sentiment and topic modeling, we further examined the association between sentiment extremity and specific topics with particular attention to extreme negative reviews. Finally, results were visualized and interpreted to provide both descriptive and explanatory insights.

3.1 Data Collection and Preprocessing

Data fields	Data types	Meaning explanation
language	string	Comment language, fixed as English
review	string	Short text comments on the game by users, used for sentiment analysis
voted_up	Boolean	Whether it is a positive recommendation, True indicates that the user recommends the game
timestamp	date	Comment posting date, in the format “year-month-day”
playtime	float	Total game play time before the comment, in hours
game type	string	Game type, analysis object is a representative game from 11 popular game types

Table 1: Data Schema

The data utilized in this research were systematically collected through automated web scraping from publicly accessible pages of the Steam platform, employing the Python libraries requests and BeautifulSoup. The dataset encompasses user-generated reviews across 11 popular game genres, resulting in a total of 101,030 review entries. During the scraping process, the webpages were individually accessed based on predefined game genre keywords, allowing extraction of the user comments section for each targeted game. The HTML structure of these pages was parsed specifically to retrieve key information, including the textual content of reviews, language indicators, voting outcomes (recommendations), timestamps of reviews, and gameplay durations. Additionally, game titles were recorded to facilitate categorical analysis. To ensure uniformity and accuracy of the dataset, only English language comments were included in the further analysis. Duplicated records and incomplete entries were also removed to prevent noise interference during model training and analysis. The finalized dataset comprises several essential fields: language, indicating the review's language (fixed as English); review, the textual content used for sentiment analysis; voted_up, a Boolean marking indicating whether the review provided a positive recommendation; timestamp, standardized to a "YYYY-MM-DD" format; playtime, representing the total hours of gameplay prior to review submission recorded in floating-point format; and game type, specifying the titles of analyzed games, all of which are high-popularity or high-sales titles.

The raw review data often contain issues such as inconsistent letter casing, embedded URLs, non-ASCII characters, distracting punctuation, meaningless numerical content, and frequent stopwords–all this can hinder the effectiveness of natural language processing (NLP) models. To address these challenges, a systematic text cleaning pipeline was implemented. First, all review texts were converted to lowercase to prevent duplicate counting of words due to case sensitivity. Next, regular expressions were used to remove all web links, including URL strings beginning with “http” or “https”, in order to prevent the model from misinterpreting these meaningless addresses as frequent terms or sentiment cues. Non-ASCII characters were also removed, including special symbols, emojis, or any non-English characters, to ensure a clean and compatible corpus for subsequent processing.

Following this, all punctuation and numerical content was stripped from the text to produce input sequences composed fully of English words. The cleaned text was then tokenized by whitespace into word lists and compared against a predefined list of English stopwords. Common but semantically weak words such as “the,” “is,” and “and” were also filtered out. This step significantly increased the emotional density of the remaining content, enabling the model to detect sentiment shifts, user opinions, and purchase intentions more effectively.

4 Research Methodology

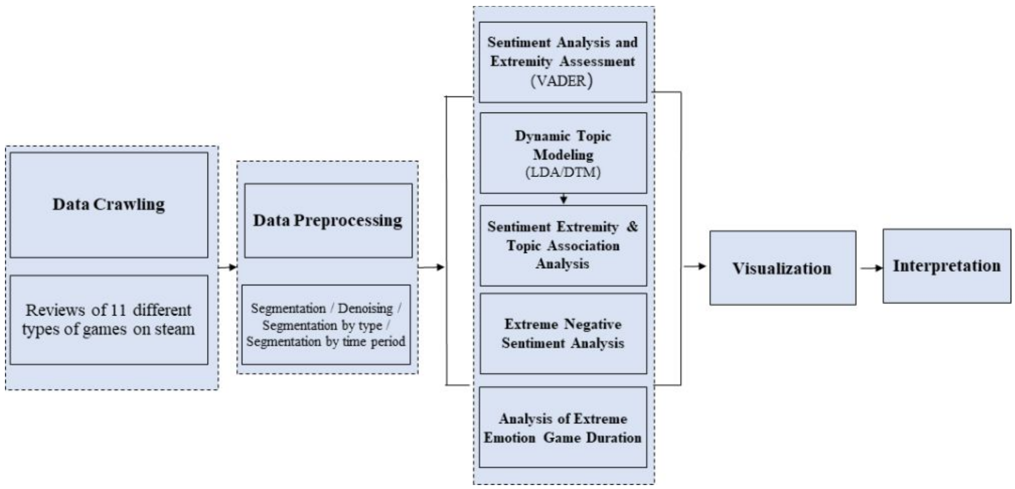

Figure 1: Research Model

The research framework of this study involves multi-stage sentiment and topic analysis, as depicted below. The analysis comprises data crawling and preprocessing, followed by sentiment analysis, topic modeling, dynamic topic modeling, and an extreme sentiment-topic association analysis, with results interpreted and visualized in each step.

4.1 Sentiment Analysis (Vader)

We use VADER (Valence Aware Dictionary and sEntiment Reasoner) to perform sentiment analysis on the English reviews from Steam. VADER is based on a sentiment lexicon and rules and computes for each comment a compound score (ranging from –1 to +1) as well as the proportions of positive, neutral, and negative sentiment. A compound score close to +1 indicates a highly positive comment, whereas a score close to –1 indicates a highly negative comment. We divide reviews into three-time windows relative to the game's release date: before release (Before), on release day (During), and within one month after release (After). For each stage, we calculate the average sentiment value (compound score) and the distribution of sentiment intensity. By comparing the mean compound scores and their distributions across the three stages, we can directly observe the evolution of player sentiment over time. The results show that for most games, review sentiment fluctuates around release time compared to before release. For example, some games have the highest mean sentiment in the "Before" stage, reflecting players' high expectations and positive emotions before launch. In contrast, the mean sentiment tends to decrease in the "After" stage, suggesting that as players experience the game more fully, evaluations become more rational or more negative.

The observed change in player sentiment over time can be explained by User Expectations Theory. According to the Expectancy-Confirmation Theory proposed by [4], the difference between a consumer's expectations before using a product and their actual perceived performance determines satisfaction and emotional outcomes. Applied to the gaming context, players typically hold high expectations before release, leading to generally positive review tendencies pre-launch. At the moment of release, actual experience is compared against prior expectations: if the game quality meets or exceeds expectations, players' emotions are positively confirmed and remain high; if the experience falls short of expectations, it may trigger disappointment and negative reviews. One month after release, as the "honeymoon" period ends, the game's hidden defects (e.g., bugs, balance issues) gradually emerge, and cumulative unmet expectations cause an increase in negative sentiment while the initial positivity wanes. This dynamic aligns with the Disconfirmation of Expectations effect: unmet initial expectations transform into more pronounced emotional responses in the later stage. Therefore, the changes in review sentiment across periods essentially reflect the psychological process of players from expectation to experience to feedback, confirming the influence of expectation management on emotional evolution.

4.2 Dynamic Topic Modeling (DTM)

While LDA effectively extracts main topics from reviews, its static assumption limits the analysis of how topics evolve. To capture temporal dynamics, we apply the Dynamic Topic Model (DTM) [18], which extends LDA by incorporating time-series information. DTM divides the corpus into ordered slices (e.g., Before/During/After) and links topics across periods, allowing each topic's word distribution to evolve coherently over time. Thus, a topic can be viewed as a trajectory rather than a static cluster. We implement DTM using LdaSeqModel in Gensim. For each game, reviews are segmented into three stages, and the same number of topics as in LDA is used for comparability. After training, we extract high-weight words and topic strengths for each slice and visualize temporal trends with heatmaps and line charts.

The advantage of LDA-based methods lies in their ability to cluster frequent discussion themes in an interpretable way. This enables longitudinal comparison of topic keywords across stages. For example, pre-release reviews often highlight anticipation and preorders, release-day reviews emphasize mechanics and graphics, while post-release reviews focus on bugs, optimization, and DLC. These shifts illustrate how player concerns change over time and demonstrate the value of topic modeling for revealing evolving discussion patterns in user-generated content.

4.3 Extreme Sentiment Analysis

For the further examination of emotional dynamics, we isolate reviews with extreme sentiment values. By applying VADER compound scores, reviews with the value of ≥ 0.85 signalize of a presence of a strong sentiment and in this work are classified as Extremely Positive, and those with the value of ≤ -0.50 are marked as Extremely Negative. All stages of the sentiment analysis show the domination of Neutral reviews. However, their proportion gradually decreases over time (e.g., from ~85% pre-release to ~83% post-release). By contrast, the proportion for extreme sentiment in reviews rises significantly (from ~5% pre-release to ~9% post-release), while the proportion for the extremely positive reviews remains relatively stable (~9% - ~8%). This indicates the presence of strong dissatisfaction that grows as gameplay experience accumulates, while strong praise persists yet does not expand.

These findings reinforce Expectation-Disconfirmation dynamics: initial optimism gives way to disconfirmation as flaws emerge, amplifying negative sentiment. Tracking extreme sentiment proportions highlights not only overall shifts but also the intensification during extended play.

4.4 Extreme Sentiment-Topic Analysis

To identify the drivers of strong emotional responses, we apply LDA and DTM to subsets of extreme reviews. Extreme positive and extreme negative corpora are modeled separately, enabling comparison with general review topics. Results show that extreme positive reviews cluster around fun gameplay immersion and storytelling often reflect deep emotional fulfillment and long term and engagement. Extreme negative reviews consistently highlight technical issues, such as bugs, performance bottlenecks, monetization concerns (DLC pricing, pay-to-win) and repetitive or insufficient content. Some topics exhibit polarization: for example, words like ‘Fun’, ‘Grind’, ‘Bugs’ can appear in both highly positive and highly negative contexts, showing the way the same topics can elicit divergent evaluations.

Comparisons reveal that while general reviews cover broad aspects such as gameplay, graphics and story, extreme reviews magnify pain points or highlights. Polarization supports prior findings in eWOM research. As demonstrated in [24], extreme reviews have disproportionate influence on consumer perceptions. In gaming communities, extreme negative topics (e.g., game-breaking bugs, unfair monetization) can trigger waves of backlash, while extreme positive topics (e.g., unforgettable moments, freedom of play) amplify enthusiasm among players.

In summary, extreme sentiment-topic analysis reveals the interaction between emotions and content: intense reviews reflect and reinforce focal concerns.

5 Results

5.1 Sentiment Variation Results Analysis

After performing VADER, we got compound scores based on which we can see the overall distribution of sentiment scores across all reviews. Results are shown in Figure 2.

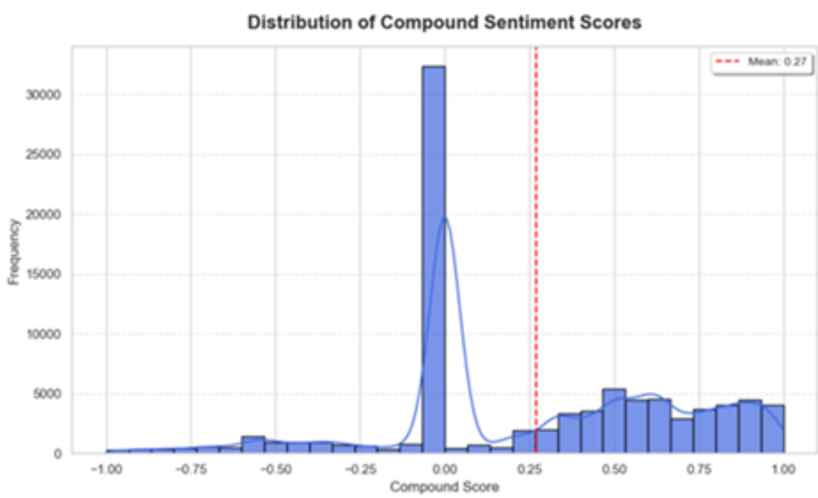

Figure 2: Overall distribution of VADER compound sentiment scores

The distribution is slightly skewed toward the positive sentiment, with most reviews clustering around neutral to moderately positive values. The mean score of 0.27 (red dashed line) confirms that players generally evaluate the games favorably, while strongly negative reviews are rare.

For temporal analysis, reviews were divided into three equal stages: Before, During, and After release. Table 2 and Figure 3 summarize the mean sentiment scores. Patterns differ by genre: RPG and Experimental games maintain consistently high sentiment, peaking in the pre-release phase; Casual and Adventure remain stable at moderately positive levels; Sports titles stay near neutral across all stages, reflecting weaker engagement; and Strategy/Simulation show higher pre-release sentiment that declines slightly after launch, suggesting that expectations were not fully met.

Game Type	Before release	During	After
RPG	0.302952	0.284518	0.28276
Action	0.221381	0.224349	0.229831
SPORTS	0.022841	0.021362	0.004772
Racing	0.220184	0.229538	0.227349
Experimental	0.516942	0.51226	0.459915
Puzzle	0.259086	0.2451	0.246152
Strategy	0.156388	0.139223	0.091733
Casual	0.357499	0.349026	0.340265
Tabletop	0.337154	0.317102	0.30292
Simulation	0.284463	0.256739	0.239653
Adventure	0.259231	0.237652	0.261209

Table 2. Compound score

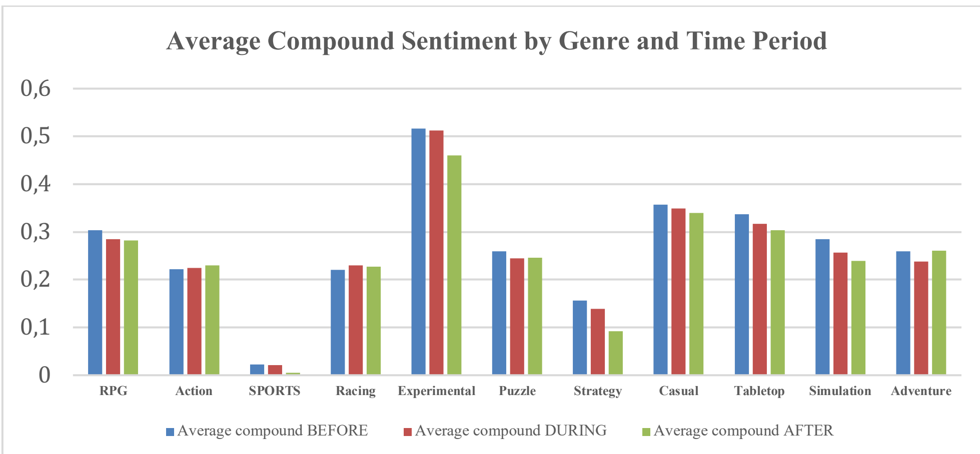

Figure 3: Average VADER compound scores for each game in Before, During, and After stages

To capture both review volume and sentiment variability, we calculated a weighted index combining normalized variance and text count. As shown in Table 3, Simulation ranks highest (1.87), reflecting both a large number of reviews and volatile sentiment. This pattern indicates that Simulation is a “high-attention, high-controversy” genre, while Adventure and Puzzle maintain more stable word-of-mouth.

game_type	sentiment_variance	text_count	norm_variance	norm_text_count	weighted_index
Simulation	0.059811	8822	0.913818	0.958868	1.872685
Experimental	0.057027	8148	0.864287	0.873302	1.737589
Tabletop	0.034233	9032	0.458736	0.985527	1.444264
Adventure	0.023556	8937	0.268772	0.973467	1.242239
RPG	0.020192	9146	0.208912	1	1.208912
Casual	0.017234	9131	0.156281	0.998096	1.154377
Puzzle	0.013986	8985	0.098488	0.979561	1.078049
Strategy	0.064655	1269	1	0	1
Action	0.00845	8977	0	0.978545	0.978545
Racing	0.009354	8415	0.016086	0.907198	0.923284
SPORTS	0.018069	7096	0.171134	0.739749	0.910883

Table 3: Game Weighted Index Table

5.2 Dynamic Topic Evolution (DTM) Results

To capture topic changes finely over time, we applied the Dynamic Topic Model (DTM). DTM is an extension of LDA that incorporates time as a variable in topic generation, allowing us to track how each topic's probability distribution changes over different time segments. This method lets us observe how the prominence of specific discussion topics rises or falls over time, providing stronger temporal sensitivity than static LDA. We applied DTM on the same Simulation game reviews by dividing them into the Before/During/After stages with the same number of topics (five) as before. After training, we obtained the word distributions and evolution of topic intensities for each topic in each stage.

	Gameplay Appraisal		Game Mechanics		Mod-Driven Enjoyment		Pricing Critique		Immersive Engagement	
	word	weight	word	weight	word	weight	word	weight	word	weight
0	sims	0.1660	sims	0.0494	fun	0.2314	game	0.0954	game	0.0886
1	game	0.0257	game	0.0491	game	0.1910	ea	0.0293	hours	0.0457
2	games	0.0249	building	0.0216	love	0.1133	dlc	0.0283	life	0.0335
3	playing	0.0140	life	0.0195	play	0.0414	free	0.0213	amazing	0.0248
4	life	0.0125	create	0.0173	sims	0.0348	packs	0.0181	sim	0.0237
5	bad	0.0116	build	0.0163	mods	0.0196	play	0.0178	played	0.0229
6	ive	0.0107	love	0.0141	playing	0.0183	sims	0.0171	ive	0.0212
7	love	0.0096	gameplay	0.0127	recommend	0.0148	money	0.0170	time	0.0195
8	bugs	0.0093	characters	0.0125	super	0.0133	dlcs	0.0151	cool	0.0187
9	play	0.0090	houses	0.0119	time	0.0112	dont	0.0149	sims	0.0175

Table 4: Before issuance LDA analysis results

	Gameplay Appraisal		Game Mechanics		Mod-Driven Enjoyment		Pricing Critique		Immersive Engagement	
	word	weight	word	weight	word	weight	word	weight	word	weight
0	sims	0.1669	sims	0.0496	fun	0.2320	game	0.0957	game	0.0890
1	game	0.0258	game	0.0492	game	0.1914	ea	0.0294	hours	0.0458
2	games	0.0250	building	0.0216	love	0.1135	dlc	0.0284	life	0.0337
3	playing	0.0140	life	0.0195	play	0.0414	free	0.0214	amazing	0.0249
4	life	0.0126	create	0.0173	sims	0.0348	packs	0.0182	sim	0.0238
5	bad	0.0116	build	0.0164	mods	0.0197	play	0.0178	played	0.0230
6	ive	0.0107	love	0.0141	playing	0.0183	sims	0.0171	ive	0.0212
7	love	0.0096	gameplay	0.0127	recommend	0.0148	money	0.0170	time	0.0195
8	bugs	0.0093	characters	0.0125	super	0.0133	dlcs	0.0151	cool	0.0188
9	play	0.0090	houses	0.0119	nice	0.0112	dont	0.0149	sims	0.0175

Table 5: Release time LDA analysis results

	Gameplay Appraisal		Game Mechanics		Mod-Driven Enjoyment		Pricing Critique		Immersive Engagement	
	word	weight	word	weight	word	weight	word	weight	word	weight
0	sims	0.1660	sims	0.0494	fun	0.2314	game	0.0954	game	0.0886
1	game	0.0257	game	0.0491	game	0.1910	ea	0.0293	hours	0.0457
2	games	0.0249	building	0.0216	love	0.1133	dlc	0.0283	life	0.0335
3	playing	0.0140	life	0.0195	play	0.0414	free	0.0213	amazing	0.0248
4	life	0.0125	create	0.0173	sims	0.0348	packs	0.0181	sim	0.0237
5	bad	0.0116	build	0.0163	mods	0.0196	play	0.0178	played	0.0229
6	ive	0.0107	love	0.0141	playing	0.0183	sims	0.0171	ive	0.0212
7	love	0.0096	gameplay	0.0127	recommend	0.0148	money	0.0170	time	0.0195
8	bugs	0.0093	characters	0.0125	super	0.0133	dlcs	0.0151	cool	0.0187
9	play	0.0090	houses	0.0119	time	0.0112	dont	0.0149	sims	0.0175

Table 6: One month after release LDA analysis results

The DTM results, visualized via line charts, clearly show how topic weights change over time. The main findings are as follows: Topic 2 (Mod-Driven Enjoyment): Representative high-weight words include "fun," "game," "love," "play," etc. This topic has the highest weight in all three stages, remaining the mainstream topic throughout. It peaks during the release day (average weight ~0.077), indicating that players' positive feedback reaches its highest point at launch. During this period, a large number of players are immersed in the fun of the game and give high praise. Topic 3 (Pricing Critique): This topic covers words like "ea," "dlc," "packs," "money," etc., related to paid content. It is active before release, showing that players were discussing the publisher's monetization strategy

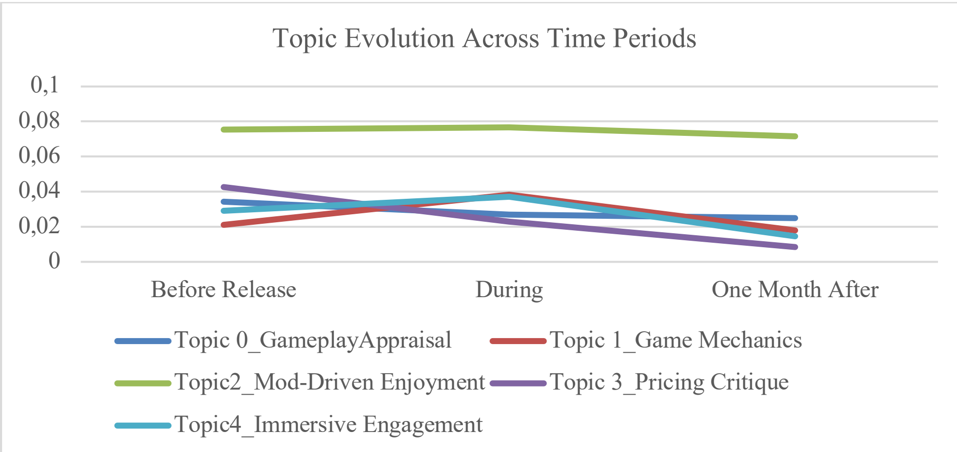

Figure 4: Topic Evolution Across Time Periods

(such as preorder pricing, DLC policy) even before launch; but by one month after release, the weight of this topic has significantly declined. This means that discussions about monetization cool off after launch, possibly because early pricing debates have subsided or subsequent operations did not spark new large-scale discussions. Topic 4 (Immersive Engagement): This topic has a low weight initially, with a secondary peak at release day, then declines to a steady level. This suggests certain minor topics (for example, specific side content or niche features) attract temporary attention at launch, but the heat does not sustain and quickly returns to normal. Topic 1 (Game Mechanics): Contains words like "sims," "life," "building," "create," etc. Its weight is enhanced on release day, related to players actively exploring game systems and creative content early on — many players at launch delve into game mechanics and share their insights, causing this topic's prominence to rise. Topic 0 (Gameplay Appraisal): Composed of words like "bad," "bugs," "ive," "play," etc., this topic's weight gradually decreases from before release to after. This indicates that as game patches and optimizations roll out, complaints about negative experiences (bugs, poor performance) diminish in the later stage, and the importance of this topic decreases.

We see that distinct types of topics exhibit a see-saw trend over the game lifecycle: positive experience topics are most prominent within one month of release, while negative topics like technical issues fade as fixed progress. Monetization discussions show a front-high, back-low pattern, indicating that players' concerns about cost/value are concentrated around launch. This, to some extent, supports [24] point on eWOM diffusion: hot topics of user concern affect the spread of sentiment in the community. Early on, positive topics (such as enjoyable content) spread widely, helping build good word-of-mouth; but if negative topics (like bugs or dissatisfaction with pricing) emerge, their spread can quickly dampen player sentiment and potential consumer confidence. Therefore, companies should be cautious of shifting public opinion: focus and guide discussions about game quality and value around launch, and continuously monitor feedback on technical issues after release, responding quickly to prevent negative sentiment from spreading (Chevalier & Mayzlin, 2006).

5.3 Distribution of Extreme Sentiments and Behavioral Characteristics

We classified reviews into Extreme Positive (compound ≥ 0.85), Extreme Negative (≤ –0.5), and Neutral. Neutral reviews dominate (82–85%) but gradually decrease, while extreme negatives rise from 5.4% before release to 8.7% after, and extreme positives remain stable around 9%.

Recommendation tags on Steam match these polarities, though a few mismatches show that textual context matters. Playtime analysis reveals that extreme positive reviewers typically invest far more hours than extreme negatives—for instance, Simulation players show a gap of 190 hours. This suggests that deep investment amplifies emotional intensity: satisfied players become loyal advocates, while disappointed ones turn into sharp critics. This dynamic aligns with Expectancy-Confirmation Theory [4] and findings by Chen et al. (2022), showing that player involvement drives polarization in online reviews.

5.4 Structural Analysis of Extreme Sentiments Themes

To further investigate the differences in specific discussion topics between extremely positive and extremely negative reviews, we conducted separate topic modeling and comparative analysis for both categories of extreme sentiment. Using LDA and DTM methods like those described earlier, we extracted all extremely positive and extremely negative reviews and incorporated the temporal dimension to construct dynamic topic models. The number of topics was set to five to capture the main themes within the context of extreme sentiment, and we analyzed how these topics evolved across the Before, During, and After release stages.

Extraction of topics from extreme sentiment reviews revealed five polarized themes. Topic 0 (Content Modularity) highlights expansions, DLCs, and mods: players praise the richness and replay value or condemn the lack of content and excessive reliance on paid add-ons. Topic 1 (Gameplay Appraisal) is split between positive comments about fun and engagement and negative ones pointing to grind, bugs, and repetitiveness. Topic 2 (DLC Valuation) centers on dissatisfaction with monetization, where overpriced or pay-to-win content provokes sharp criticism. Topic 3 (Pricing Critique) intensifies these concerns, with users expressing frustration over repetitive gameplay and poor value in harsh language. Finally, Topic 4 (Immersive Engagement) captures the opposite extreme, where players describe long hours of play, strong emotional bond and show enthusiastic intentions.

Overall review of these topics shows that the same game elements—content, gameplay, pricing, and immersion—can spark both intense praise and equally strong criticism, underscoring the polarized nature of extreme reviews.

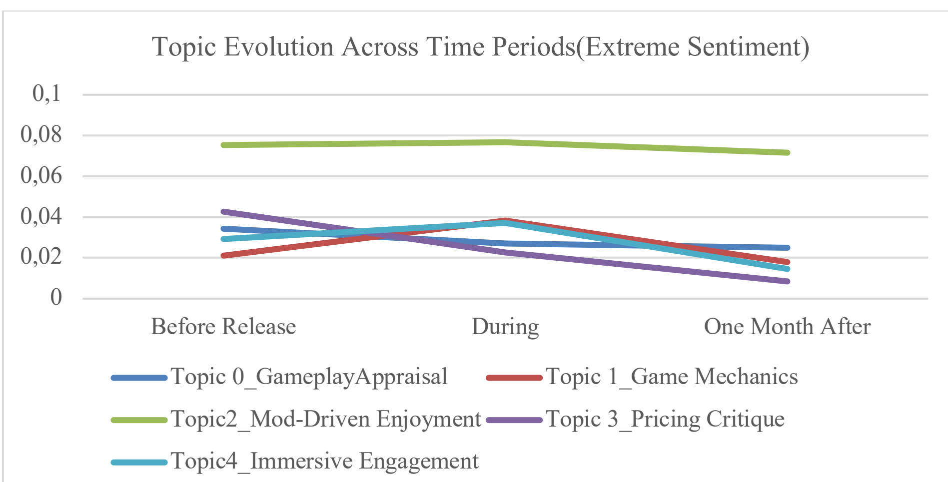

Figure 5: Topic Evolution Across Time Periods(Extreme Sentiment)

As shown in Figure 5, Topic 1 (Gameplay Appraisal) reaches its highest average weight during the release phase (approximately 0.054), indicating that the launch period is the most emotionally volatile stage, with a surge of both highly positive and highly negative reviews and intense opinion clashes. Topic 3 (Pricing Critique) exhibits a clear downward trend–its weight drops from 0.0256 before release to a lower level afterward–suggesting that some of the strong criticisms are alleviated over time, possibly due to subsequent updates addressing the issues or the departure of dissatisfied users who no longer voice their concerns. In contrast, Topic 2 (DLC Valuation) shows stable weights across all three stages, consistently maintaining a certain level of attention. This reflects players' persistent sensitivity to pricing and monetization strategies–whenever the perceived value is low, negative discussions are likely to emerge. Topic 4 (Immersive Engagement) maintains a relatively low overall weight but shows a slight rebound in the later post-release stage, suggesting that although purely extremely positive content constitutes a smaller portion of the reviews, a group of loyal players continues to share their appreciation over time, forming a small but steady stream of positive sentiment.

6 Discussion

Our analysis shows that player sentiment fluctuates significantly across the game lifecycle. Sentiment often peaks on release day, driven by high expectations and initial enthusiasm. However, one-month post-release, sentiment tends to decline, especially in genres like Strategy and Simulation.

This pattern reflects Oliver's (1980) Expectancy-Confirmation Theory: unmet expectations lead to negative disconfirmation, which reduces satisfaction. While early adopters give optimistic feedback, broader audiences later adopt a more critical stance as flaws surface. These results for RQ1 underscore the time-sensitive nature of sentiment in eWOM, revealing both enthusiasm and disappointment as products evolve.

LDA and DTM models reveal that review topics shift from anticipation pre-release to gameplay evaluation at launch, and eventually to technical and monetization concerns post-release. Extreme sentiments correlate strongly with topic content: positive extremes often praise narrative immersion or gameplay freedom, while negative extremes focus on bugs, pricing, and DLC dissatisfaction. Some issues, though minor in general discussions, are magnified in extreme reviews, aligning with Khalid et al. (2014).

These findings for RQ2 and RQ3 suggest that emotional intensity is not only time-dependent but also topic-specific. Identifying emotionally charged topics can help developers manage sentiment proactively. By the one-month mark, extreme negative sentiment rises sharply (from ~5% to nearly 9%), while positive extremes slightly decline. This post-honeymoon effect can be attributed to accumulated user frustration and unmet expectations. According to Oliver (1980) and Chevalier & Mayzlin (2006), dissatisfaction tends to intensify over time and spreads more rapidly than praise. Negative feedback often gains traction in communities, triggering wider dissatisfaction. Notably, this shift may also reflect a changing reviewer demographic—early fans versus later, more critical users.

RQ4 results highlight the importance of monitoring early warning signals in post-launch stages to mitigate reputational risks.

We observe a clear pattern: players who leave extreme positive reviews generally invest more time in gameplay than those with negative extremes, particularly in RPG and Strategy genres. This suggests emotional intensity scales with investment—greater engagement leads to stronger emotional outcomes. If experiences fall short after significant time input, disappointment is magnified. Chen et al. (2022) confirm this link between playtime and sentiment extremity. Highly invested users can act as both advocates and critics. For RQ5, these findings reveal how behavioral metrics like playtime can enhance sentiment analysis and support more nuanced eWOM research and product management.

7 Conclusions

This study offers a multi-dimensional framework for analyzing user reviews in the gaming domain by integrating sentiment analysis, topic modeling, and behavioral metrics over time. From a theoretical perspective, our findings validate the dynamic nature of player sentiment through the lens of Expectancy-Confirmation Theory [4], showing that emotional responses fluctuate across the game lifecycle, especially in the post-launch phase. The application of LDA and DTM models fills a methodological gap in UGC and eWOM research by illustrating how player discussion topics evolve over time, rather than remaining static. Furthermore, our focus on extreme sentiment reveals strong associations between emotional polarity, content specificity, and user investment levels—suggesting that highly engaged players often produce the most polarized reviews. This insight extends existing work on opinion leadership and digital word-of-mouth [24] emphasizing the dual role of core users as both influential advocates and potential detractors. Practically, these findings offer valuable guidance for game developers and platform managers. Developers should pay close attention to post-release feedback patterns, especially within the first month, and proactively address technical or content-related concerns to prevent growing dissatisfaction. Community managers are encouraged to engage high-investment users constructively, as their feedback—positive or negative—carries disproportionate influence. Digital platforms, in turn, can apply sentiment and topic-trend monitoring to improve algorithmic recommendations, detect emerging crises, and support more transparent review systems. Ultimately, this research provides data-driven insights that stakeholders across the gaming ecosystem can leverage to enhance product strategy, community interaction, and long-term user satisfaction.

Despite the valuable insights offered by this study, several limitations suggest areas for future enhancement. First, the dataset—comprising only English-language reviews from Steam—limits cultural and platform generalizability. Diverse linguistic and regional expressions of sentiment remain underexplored, restricting the applicability of findings to a broader global gaming audience. Second, while lexicon-based sentiment tools and standard topic models were effective, they may overlook nuanced sentiment, sarcasm, or semantic complexity. The chosen time windows—pre-release, release, and one-month post-release—also provide a short-term lens on emotional dynamics, leaving longer-term sentiment trends unexamined. Moreover, the study focused exclusively on textual reviews, omitting multimodal content (e.g., screenshots, videos, or social media narratives) that now play a central role in how players share experiences. Lastly, uneven sample sizes across game genres may bias results toward mainstream titles, limiting insights into niche categories. Future research should address these issues by diversifying platforms, languages, and modalities, incorporating visual and audio data, and applying advanced machine learning techniques such as transformer-based models or multimodal learning frameworks. Extending the temporal scope of analysis and integrating network-based approaches could also illuminate the mechanisms of sentiment diffusion and the influence of opinion leaders. These efforts would deepen our understanding of digital word-of-mouth in gaming and support the refinement of both theoretical models and practical applications in user experience and product development.

References

[1] Newzoo. (2024). *The global games market will generate $187.7 billion in 2024*. (Newzoo Industry Report). Newzoo Insights.

[2] Dean, D. (2023, February 12). *Steam just reached 50,000 total games listed*. PCGamesN. (Archived April 28, 2021)en.wikipedia.org

[3] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167.

[4] Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research, 17(4), 460–469.

[5] Andreev, G., Saxena, D., & Verma, J. K. (2021). *Impact of review sentiment and magnitude on customers′ recommendations for video games.* In 2021 International Conference on Computational Performance Evaluation (ComPE) (pp. 992–995). IEEE.

[6] Pagano, D., & Maalej, W. (2013). *User feedback in the App Store: An empirical study.* In 2013 21st IEEE International Requirements Engineering Conference (RE) (pp. 125–134). IEEE.

[7] Khalid, H., Shihab, E., Nagappan, M., & Hassan, A. E. (2015, May-June). What do mobile app users complain about? IEEE Software, 32(3), 70–77.

[8] Hutto, C. J., & Gilbert, E. (2014). *VADER: A parsimonious rule-based model for sentiment analysis of social media text.* In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM) (pp. 216–225).

[9] Chakraborty, S., Mobin, I., Roy, A., & Khan, M. H. (2018, December). *Rating generation of video games using sentiment analysis and contextual polarity from microblog.* In 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS) (pp. 90–95). IEEE.

[10] Zuo, Z. (2018). Sentiment analysis of Steam review datasets using Naive Bayes and Decision Tree classifier. (Master′s thesis, University of Illinois at Urbana-Champaign)

[11] Tan, J. Y., Chow, A. S. K., & Tan, C. W. (2021). *A comparative study of machine learning algorithms for sentiment analysis of game reviews.* In Proceedings of the International Conference on Digital Transformation and Applications (ICDXA 2021) (pp. 63–68). IEM.

[12] Fadhlurrahman, J. A. M., Herawati, N. A., Aulya, H. R. W., Puspasari, I., & Utama, N. P. (2023, October). *Sentiment analysis of game reviews on Steam using BERT, BiLSTM, and CRF.* In 2023 International Conference on Electrical Engineering and Informatics (ICEEI). IEEE.

[13] Coronado-Blázquez, J. (2024). *A NLP approach to ″review bombing″ in Metacritic PC videogames.* arXiv preprint arXiv:2405.06306.

[14] Park, G., & Kwak, M. (2020). *The life cycle of online smartphone reviews: Investigating dynamic change in customer opinion using sentiment analysis.* ICIC Express Letters, Part B: Applications, 11(5), 509–516.

[15] Lu, C., Li, X., Nummenmaa, T., Zhang, Z., & Peltonen, J. (2020). *Patches and player community perceptions: Analysis of No Man's Sky Steam reviews.* In Proceedings of the 2020 Digital Games Research Association (DiGRA) Conference. (Case study paper).

[16] Yu, Y., Nguyen, B. H., Yu, F., & Huynh, V. N. (2021, November). Esports game updates and player perception: Data analysis of PUBG Steam reviews. In 2021 13th International Conference on Knowledge and Systems Engineering (KSE) (pp. 1–6). IEEE.

[17] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent Dirichlet allocation.* Journal of Machine Learning Research, 3, 993–1022.

[18] Blei, D. M., & Lafferty, J. D. (2006). *Dynamic topic models.* In Proceedings of the 23rd InternationalLi, C., Li, S., Yang, J., Wang, J., & Lv, Y. (2023). *Topic evolution and sentiment comparison of user reviews on an online medical platform in response to COVID-19: Taking review data of Haodf.com as an example.* Frontiers in Public Health, 11, 1088119.

[19] Conference on Machine Learning (pp. 113–120). ACM.

[20] Guzsvinecz, T., & Szűcs, J. (2023). *Length and sentiment analysis of reviews about top-level video game genres on the Steam platform.* Computers in Human Behavior, 149, 107955.

[21] Kaplan, A. M., & Haenlein, M. (2010). *Users of the world, unite! The challenges and opportunities of social media.* Business Horizons, 53(1), 59–68.

[22] King, R. A., Racherla, P., & Bush, V. D. (2014). *What we know and don't know about online word-of-mouth: A review and synthesis of the literature.* Journal of Interactive Marketing, 28(3), 167–183.

[23] Lengel, R. H., & Daft, R. L. (1984). An exploratory analysis of the relationship between media richness and managerial information processing (No. TRDG08ONR).

[24] Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. Journal of Marketing Research, 43(3), 345–354.