Investigating the Causes of Player Frustration in Modern Games: A Large-Scale Review Analysis

Jiacheng Xia
Computer Science, NYUAD
jx2467@nyu.edu

Advised by: Mohamad Kassab

ABSTRACT

Player frustration is a critical but underexplored consideration in game design and software engineering. This study presents a large-scale, data-driven investigation into the sources of player frustration by mining over two million user reviews across platforms, including Steam, Reddit, and Metacritic. Using a multi-stage natural language processing pipeline that incorporates sentiment filtering, BERTopic clustering, and qualitative interpretation, we categorize player complaints through the lens of software architecture and non-functional requirements (NFRs). The resulting topic clusters reveal recurring frustration triggers such as technical instability, poor matchmaking fairness, toxic social environments, restrictive access policies, and grind-heavy progression systems. Surprisingly, difficulty-related complaints—often theorized as a central frustration factor—did not emerge as distinct clusters, suggesting that challenging gameplay may not provoke widespread dissatisfaction among invested players. These findings inform a software-oriented framework for understanding user frustration and highlight potential avenues for integrating user sentiment into game requirements engineering.

KEYWORDS

frustration, user reviews, topic modeling, sentiment analysis, software quality attributes, non-functional requirements, game requirements engineering

This report is submitted to NYUAD's capstone repository in fulfillment of NYUAD's Computer Science major graduation requirements.

جامعة نيويورك أبوظبي
NYU | ABU DHABI

Capstone Seminar, Spring 2026, Abu Dhabi, UAE

Reference Format:
Jiacheng Xia. 2026. Investigating the Causes of Player Frustration in Modern Games: A Large-Scale Review Analysis. In *NYUAD Capstone Seminar Reports, Spring 2026, Abu Dhabi, UAE.* 10 pages.

1 INTRODUCTION

Frustration is a central emotional experience in gameplay, closely tied to user retention, satisfaction, and long-term engagement. While it is a well-documented construct in psychology, its integration into game development practice—particularly from a software engineering perspective—remains limited. Existing literature has characterized frustration as arising from perceived obstruction, lack of competence, or loss of control, all of which are highly relevant in interactive digital systems like games. However, operationalizing this concept at scale remains challenging.

To bridge this gap, this study investigates the causes of player frustration by leveraging large-scale user-generated data. We hypothesize that frustration is not solely a product of game difficulty but can emerge from various breakdowns across technical, design, and social dimensions. Through automated scraping and preprocessing of reviews from major platforms, we apply unsupervised topic modeling to detect latent themes of dissatisfaction. By framing our analysis in terms of software architecture and non-functional requirements—such as usability, availability, accessibility, and fairness—we aim to identify which software qualities most frequently correlate with negative user sentiment.

This approach allows us not only to validate existing theories of frustration, but also to generate actionable insights for game developers seeking to improve player experience. Moreover, the study contributes a methodological blueprint for integrating natural language feedback into the requirements engineering process, providing a foundation for future tools and frameworks tailored to emotion-aware game development.

Research Questions

- What causes frustrations?

- Can player frustration in games be categorized?
- Is there a correlation between certain types of player frustration and the game features?
- Can we predict certain types of player frustration?

2 RELATED WORKS

2.1 Psychological Definitions of Frustration

Frustration is broadly defined in psychology as a state that arises when a desired goal or expected outcome is thwarted. The APA Dictionary of Psychology describes frustration as "the thwarting of impulses or actions that prevents individuals from obtaining something they have been led to expect based on past experience," whether the obstruction comes from internal conflicts or external obstacles (e.g. other people or rules of society), and as "the emotional state an individual experiences when such thwarting occurs". In other words, frustration encompasses both the situational blockage of a goal and the resulting negative affective state. Historically, one of the earliest formalizations of frustration was the frustration–aggression hypothesis. Dollard et al. (1939) famously proposed that if a goal-directed action is blocked, the resulting frustration invariably leads to aggression. Berkowitz (1989) later refined this view, defining a frustration as "an interference with the occurrence of an instigated goal-response at its proper time in the behavior sequence" (Berkowitz, 1989). In essence, whenever a planned or motivated behavior toward a goal is interrupted or prevented, a state of frustration occurs. This notion aligns with Lewin's field theory account: "Frustration arises when a person's movement toward a goal-directed behavior is blocked by an obstacle" (Lewin, 1959). Lewin emphasized that blocking an individual's path to a goal creates an unresolved psychological tension (a "quasi-need") that manifests as frustration until the goal can be achieved. Other theorists have approached frustration from a learning and motivation perspective. Amsel (1992), in his frustration theory of learning, defined frustration "simply as a temporary state that results when a response is nonreinforced (…in the presence of a reward expectancy)". In this view, frustration is an internal state triggered by the omission of an expected reward. Notably, Amsel argued that this aversive state becomes a learned motivational force or "conditioned drive" that has behavioral consequences. In other words, the shock of not getting an anticipated outcome (a violation of expectancy) generates frustration, which can then energize the organism to persist or try new behaviors (a phenomenon known in learning theory as the frustration effect). Similarly, Brown and Farber (1951) conceptualized frustration as arising from conflicts within the organism's response tendencies. They suggested that frustration is the consequence of either "the simultaneous activation of two competing excitatory tendencies" or a single goal-directed impulse being opposed by an inhibitory force. Under this conflict model—rooted in Hullian learning theory—frustration is an intervening state that occurs when an active drive is blocked by some internal or external constraint. Importantly, Brown and Farber noted that such frustration can amplify motivation, releasing characteristic "frustration cues" and altering which behavior the organism ultimately performs. This implies that while frustration is unpleasant, it may also intensify one's effort to overcome the barrier or find alternative routes to the goal. Contemporary psychological theories highlight that frustration can result not only from tangible goal blocks but also from the thwarting of basic psychological needs. Self-Determination Theory, for example, differentiates between mere need non-fulfillment and active need frustration. Deci and Ryan (2000) (as cited by Bartholomew et al., 2011) described need frustration as a distinct state in which essential needs are actively undermined, rather than just unmet. Bartholomew et al. (2011) elaborate that the "negative experiential state of need thwarting" arises when individuals perceive that their autonomy, competence, or relatedness needs are being actively subverted by external forces. This state of need-focused frustration is associated with more intense negative outcomes and emotional ill-being, compared to the milder feelings of dissatisfaction that come from passive need deprivation. Specific sub-types of need frustration have been identified: autonomy frustration, the feeling of being controlled or coerced (e.g. "I feel forced to follow decisions made for me"); competence frustration, the feeling of ineffectiveness and failure (e.g. "I am told things that make me feel incompetent"); and relatedness frustration, the feeling of being rejected or excluded (e.g. "I feel I am rejected by those around me"). Each of these reflects the general principle that when a person's fundamental goals or needs (to feel volitional, effective, and connected) are blocked or actively counteracted, the result is a pronounced state of frustration.

3 METHODOLOGY

3.1 Phase One: Data Collection

3.1.1 Main Game Selection Criteria. To find out sources of player frustration, we chose to obtain game-related comments or reviews for 46 video games released after 2010. Games were chosen based on the above research on psychological definitions of frustration. Given these theoretical perspectives on frustration, the present study established specific criteria for selecting video games that are likely to provoke the kinds of frustration outlined above. The game selection was explicitly guided by the psychological literature, ensuring that each chosen title exemplifies a distinct frustration-inducing factor identified in prior research. In particular, four key sources of player frustration were targeted: (1) challenges to mastery and competence, (2) technical or external interruptions, (3) game design and mechanics issues, and (4) negative social or multiplayer dynamics. Each of these selection criteria was grounded in the frustration theories discussed, as justified below.

First, games with steep learning curves, high difficulty, or "high-stakes" failure situations were prioritized to capture frustration related to mastery and competence. When players struggle to overcome in-game challenges or repeatedly fail to achieve important goals, they experience what SDT describes as competence frustration, a feeling of ineffectiveness or inadequacy in performance (Bartholomew et al., 2011). For example, a game that forces the player to restart a level after a costly failure (creating a sense of lost progress and forced repetition) directly blocks an instigated goal-response, which is precisely the condition Berkowitz (1989) identified as producing frustration. According to the frustration-aggression framework, such goal obstruction tends to elicit anger and tension, underscoring why high-difficulty games can be so infuriating. On the positive side, overcoming this type of frustration can be rewarding; Amsel's theory suggests that the frustration from an unfulfilled expected reward can intensify motivation to persist. In practice, this means a game that initially frustrates the player's sense of mastery (by blocking success) may also encourage the player to try harder and eventually feel a greater sense of achievement once the obstacle is overcome.

Second, the selection included games known for technical issues or external hindrances (such as control lag, bugs/glitches, or network connectivity problems) to represent sources of frustration outside the core gameplay design. Frustration arising from technical failures aligns with the classic definition of frustration as an outcome of external forces thwarting one's expected outcomes. If a server crash or input lag prevents a player from accomplishing an in-game objective, this external barrier is essentially an environmental constraint blocking goal attainment. In Lewin's terms, an impediment in the environment (here, a technical malfunction rather than an in-game challenge) can halt the player's locomotion toward their goal, leading to frustration. Such "at-game" frustration is typically unproductive and anger-inducing, since the player has no direct control over these external interferences. By including games with known technical frustrations, the study ensures that the frustration observed is not only due to game difficulty but also due to unpredictable external goal-blocks, reflecting the full scope of the APA's definition of frustration (from internal conflicts to outside obstacles).

Third, games featuring punishing or restrictive design elements were selected to examine frustration caused by game mechanics and rules. Design choices like unclear objectives, unresponsive controls, unfair mechanics, or forced repetitive tasks can directly interfere with the player's intended actions and goals. From a theoretical standpoint, these situations create exactly the kind of interference in goal-response sequences that Berkowitz (1989) highlighted in his definition of frustration. For instance, if a game's poor control responsiveness causes the player's desired action to fail (despite the player knowing what to do), the game is effectively interrupting an "instigated goal-response," leading to frustration at the design rather than at the player's own skill. Moreover, certain design constraints thwart players' psychological needs: a game that heavily restricts player choice or forces a particular playstyle can induce autonomy frustration (the feeling of being controlled), while a game that makes the player feel constant failure with little guidance can induce competence frustration (feeling incapable). In line with Amsel's expectancy-based model, when a game violates the player's expectations of making steady progress or mastering a task – for example, by resetting progress repeatedly without clear justification – it creates a strong frustration response due to the unfulfilled expectation of reward or success. Thus, the games chosen for this criterion exemplify how poor or punitive design can thwart both concrete goals and basic autonomy/competence needs, leading to the kind of frustration documented in the psychological literature.

Finally, multiplayer titles with toxic social environments or perceived unfairness were included to capture frustration arising from social dynamics. Negative interactions with other players – such as harassment, griefing, or cheating – can severely frustrate a player by undermining their social needs and fair-play expectations. In self-determination terms, a toxic multiplayer experience can create relatedness frustration, the emotional pain of feeling rejected, targeted, or isolated by others (Bartholomew et al., 2011). For example, a player who is bullied by teammates or thwarted by an unscrupulous opponent will likely feel a sense of social exclusion or injustice, reflecting a direct frustration of the need for respect and belonging. Such situations also fit with the general definition of frustration as goal blockage by external agents: another person's actions (e.g. using a cheating exploit or engaging in sabotage) become the obstacle that prevents the focal player from attaining victory or enjoyment. According to the frustration–aggression hypothesis, these socially caused frustrations are particularly important to consider, as a player whose goals are blocked by perceived unfairness or hostility may react with anger or aggressive behavior (Berkowitz, 1989). By selecting games notable for toxic or unfair multiplayer elements, the study addresses frustration that is interpersonal in nature – a form of frustration grounded in the thwarting of social expectations and norms, which the literature indicates is a potent trigger for negative affect.

In summary, the game selection criteria were deliberately tied to established psychological theories of frustration. Each chosen game scenario (whether it involves mastering a difficult challenge, dealing with technical failures, enduring harsh game design, or facing toxic players) corresponds to a specific frustration mechanism identified in prior research. This literature-guided approach ensures that the study can investigate frustration in a comprehensive manner – from the classic case of blocked goals and unmet expectations to the more nuanced cases of psychological need thwarting – using examples that resonate with both theoretical and practical significance.

3.1.2 Extended Game Selection Criteria. Building on the frustration theory–based criteria, we also considered several non-mechanical factors to justify including certain games in our analysis. These factors—a large player base, high player expectations, extensive review volume, and a high negative review ratio—are grounded in psychological and methodological reasoning from related work. Frustration has long been studied in psychology, commonly defined as the emotional state that occurs when an individual's goals or expected outcomes are blocked or thwarted [11]. Classic theories conceptualize frustration as arising from the obstruction of goal-directed behavior or the prevention of an expected reward [6, 9, 10]. For example, the seminal frustration–aggression hypothesis posits that frustration (the blocking of an anticipated goal) inevitably instigates an aggressive drive [6]. Later work by Berkowitz re-examined this hypothesis and suggested that frustration produces an aversive emotional state (e.g., anger) which can lead to aggression under certain conditions [4]. More recent perspectives highlight that frustration can also result from the thwarting of fundamental psychological needs or other anticipated rewards [2, 5]. In the context of games, even if a title does not neatly exemplify one of the core "frustration types" identified in prior literature, it can still provide valuable insight into player frustration via these broader mechanisms. Below, we outline each supplementary selection factor and its rationale in light of prior research.

Large Player Base

Games with an expansive player base expose any potential frustration triggers to a huge audience, thereby statistically increasing the likelihood that those triggers will be encountered and reported. In essence, sheer numbers act as a stress-test: given enough players, even low-probability issues (e.g., a rare bug or an edge-case design flaw) are almost guaranteed to surface. This is analogous to an exposure effect or a "law of large numbers" principle in sampling—a larger sample heightens the chance of observing infrequent phenomena. By selecting games with large communities, we ensure our dataset captures such emergent frustration scenarios. Even minor design or technical issues will impact a substantial absolute number of players, making those problems more likely to appear in player feedback simply due to the size of the population.

For instance, the widely-played *Deathloop* initially received stellar critical reviews (Metascore 88, including multiple 9/10 and 10/10 scores from major outlets). However, once millions of players got their hands on it, numerous frustration-inducing problems came to light [13]. A great number of players discovered that *Deathloop* "was not as good as the press described," encountering frequent freezes, endless bugs, poor performance on PC, and ineffective anti-cheat measures that prevented meaningful engagement with its level design—after all, no one could appreciate the level design if they could not even enter the game [13]. These issues, largely missed in limited pre-release tests, illustrate how a vast player population can reveal frustration points that a smaller beta might overlook.

Thus, including games with very large player bases increases the probability of capturing such frustration incidents in our analysis; even subtle or low-frequency problems will surface when exposed to a sufficiently large number of players. This approach aligns with purposive sampling logic in qualitative research: cases with rich and numerous user experiences maximize the observed range of the phenomenon of interest (player frustration).

High Player Expectations (Hype and Brand Recognition)

Prior literature on expectancy violation suggests that when outcomes fall short of prior expectations, the resulting negative affect (e.g., disappointment and frustration) is amplified. In the context of games, a well-known franchise, beloved IP, or heavily marketed title sets a high expectation bar for players. If the delivered experience fails to meet these elevated hopes, players experience a classic expectancy-disconfirmation effect that can manifest as acute frustration. The American Psychological Association's Dictionary of Psychology defines frustration as "the thwarting of impulses or actions that prevents an individual from obtaining something they have been led to expect based on past experience" [1]. In other words, frustration fundamentally involves the blocking of an anticipated outcome.

Thus, a highly anticipated game that does not live up to its hype creates a textbook frustration scenario—the player's goal (a satisfying experience) is thwarted by unmet expectations. This dynamic is well documented in consumer psychology: according to expectancy-disconfirmation theory, the gap between what was expected and what was obtained drives dissatisfaction or frustration [11]. Oliver's model of customer satisfaction posits that unmet expectations (i.e., negative disconfirmation) directly lead to dissatisfaction and negative emotions [11].

Moreover, psychological need-thwarting in games can exacerbate this effect. Research on player experience finds that when games undermine players' basic psychological needs or fail to deliver on a promised experience, players report intense frustration and disengagement [2, 5]. Ballou and Deterding's recent grounded theory analysis of need frustration in games, for instance, noted that players often become frustrated when a game fails to fulfill the autonomy, competence, or relatedness experiences they were led to anticipate [2].

Empirically, we have seen this pattern in several high-profile releases. Players were deeply disappointed with *Deathloop* not just due to its technical flaws, but because the expectation of a masterpiece (fueled by glowing early reviews) was shattered by the reality of a buggy product. Similarly, over-hyped titles such as *Cyberpunk 2077* and *No Man's Sky* demonstrate that the greater the pre-release hype, the more intense the backlash and frustration when those expectations go unfulfilled. When a cherished outcome fails to materialize, the resulting sense of frustration is stronger and tends to transform initial disappointment into anger and public criticism. Therefore, we included games with strong fan anticipation or renowned IP status, as expectancy-based frustration is likely to be especially pronounced in these cases.

Large Volume of Player Reviews

We also prioritize games with an abundance of player reviews or comments, since a large user-review corpus provides a rich source of qualitative data on frustration experiences. Methodologically, analyzing titles with thousands of reviews increases the likelihood of capturing diverse and representative examples of player frustration, which ties into principles of saturation in qualitative research. Empirical studies have shown that even relatively modest sample sizes (on the order of a dozen interviews) can yield a very high percentage of the thematic content in a given domain. For example, Hennink and Kaiser found that 9–17 interviews were often sufficient to reach saturation (uncovering around 90% of themes) in studies with relatively homogeneous participant groups [8]. By contrast, a game with 5,000+ user reviews offers an immensely broader sample—virtually guaranteeing that all the major frustration themes will have surfaced.

In other words, a high volume of user feedback allows us to approach thematic saturation: beyond a certain point, additional reviews are unlikely to reveal entirely new types of frustration. This strengthens the reliability of our analysis, as we can be confident we are not missing key frustration issues that players experience.

Moreover, the prominence of certain complaints in a large review set benefits from the availability heuristic in cognitive psychology. According to Tversky and Kahneman [12], people tend to judge an issue as more significant if examples of that issue come readily to mind. In the context of game reviews, problems repeatedly reported by many players become highly salient and easily recalled as defining characteristics of the game. Sheer repetition makes specific frustration issues "top of mind" for the community. Thus, by examining games with voluminous player feedback, we effectively leverage this cognitive bias: the most frequently cited frustrations stand out in our analysis as the key problem areas.

High Ratio of Negative User Reviews

Finally, we gave special consideration to games with an unusually high proportion of negative user reviews or abnormally low user ratings (especially when those ratings sharply contrast with critics' scores). A high negative-review ratio is a strong signal of widespread player dissatisfaction—essentially a proxy indicator that the game has frustrated a large segment of its audience. Psychological research on negativity bias shows that people tend to give more weight and attention to negative experiences than to positive ones [3]. Classic frustration–aggression theory also notes that frustration often induces feelings of anger and an urge to vent or retaliate [4, 6].

In consumer domains, negative word-of-mouth is disproportionately driven by feelings of frustration and unmet expectations. People are far more likely to speak up about bad experiences than to praise good ones, precisely because aversive outcomes spur stronger emotional responses [3]. Industry reports echo this phenomenon, noting that "disappointment and frustration are often strong drivers, prompting people to vent," and that around 95% of customers share bad experiences with others [7].

Including such titles in our study is methodologically prudent: these games provide clear, emphatic examples of what has gone wrong from the player's perspective, since so many players felt strongly enough to document their anger. Moreover, a high negative-review ratio helps zero in on the specific elements of a game that are dissatisfying. The accompanying review comments usually make it obvious why players are frustrated—whether it is due to a buggy launch, unbalanced or unfair design choices, broken promises, or a toxic in-game community.

These prevalent complaints align with established frustration triggers. For instance, a game-breaking bug or technical failure can be seen as a classic blocked goal—the player's progress is directly impeded. Likewise, aspects of a game that undermine player autonomy or fairness represent psychological need thwarting, which is known to generate strong negative affect [2, 5].

Summary

In summary, these supplementary selection criteria complement the core frustration-type approach by incorporating broader psychological and community factors identified in related work. A large player base ensures that even subtle frustration triggers are exposed through sheer user volume; high expectations leverage expectancy-violation dynamics that can turn otherwise minor flaws into major episodes of disappointment and anger; a large volume of reviews provides ample qualitative evidence (approaching data saturation) of players' frustration experiences; and a high negative review ratio flags games with pronounced, widespread dissatisfaction. Grounding our game inclusion strategy in these considerations allows us to bridge established theories of frustration with modern player-community dynamics. This ensures our analysis captures not only theoretically derived frustration categories, but also the real-world contexts in which player frustration manifests most visibly and intensely.

3.1.3 Game Review Sources. To capture how players articulate frustrations, the project harvested user-generated reviews from several platforms.

Steam (for PC releases) — User reviews were downloaded through Valve's public `appreviews` API. A command-line tool prompts the user for a game name, the Steam App ID, and a start and end date. If the user enters "y" for both dates, the script automatically selects a two-year window ending at the current date; otherwise it parses specific `YYYY-MM-DD` dates and converts them to Unix timestamps. The script then repeatedly requests pages of up to 100 reviews, ordered by recency (`filter=recent`) and filtered to English (`language=english`) and all purchase types (`purchase_type=all`). Reviews outside the time window are skipped, and encountering a review older than the start date terminates the loop. Each in-range review's ID, author metadata (Steam ID, number of games owned, total playtime), review text, creation and update timestamps, helpful votes, and flags (purchase / early access / Steam Deck use) are inserted into a SQLite table; primary keys prevent duplicates.

Reddit — To collect discussion threads and comments resembling reviews, an ingestion pipeline was built on top of the PRAW library. A YAML configuration file specifies the time window, list of games and their aliases, subreddits to search, and heuristic parameters (minimum words, regular expressions). For each game, the agent constructs a set of subreddits and performs listing pulls (`recent`, `top_day`, `top_week`) and search pulls using Lucene queries such as `title:"<game>"` (review OR impressions OR "my thoughts" ...). Posts older than the target window are discarded and the text (title + body) is normalised (lower-cased and punctuation removed). The agent attempts to match a post to the current game by exact, alias, and fuzzy matching, and requires that posts satisfy English-language and review heuristics: the ASCII character ratio must exceed a threshold, and the title or flair must contain review-like keywords, or the body must exceed a minimum word count. For accepted posts, the pipeline extracts the platform (PC / console) and playtime using regular expressions, and stores rows in a SQLite table with fields such as Reddit ID, kind (post or comment), subreddit, author, timestamps, title, body, URL, score, number of comments, matched game, platform, playtime, and flags for "looks like review" and "passes English heuristics". Comments are processed recursively in a similar manner; accepted comments are upserted into the same table with their parent ID.

Metacritic — The `metacritic_scraper.py` script uses Selenium to handle Metacritic's infinite scroll and collects user reviews across all available platforms (PC, PlayStation, Xbox, etc.). The user supplies a Metacritic game URL and a start and end date; entering "y" for both uses a two-year window ending today. The script derives a slug from the URL path and constructs a table name `<slug>_<start>_<end>_metacritic`. It loads the user-reviews page in a headless Chrome browser and scrolls in 1000-pixel increments until either the last-loaded review's date is earlier than the start date or the bottom of the page is reached. For each platform, the script parses each review block, extracting the rating (0–10), username, review date, platform, and review text. It discards spoiler blocks, reviews without usernames or valid dates, and non-English reviews (less than 90% ASCII characters). Reviews are deduplicated in memory and inserted into a SQLite table with a unique constraint on `(username, platform, rating, review_date, review_text)` to avoid duplicates.

Official forums and other sources — Additional scripts within the archive collected posts from official developer forums (e.g., Escape From Tarkov, EA, Overwatch 2, Baldur's Gate 3) and patch-note comments from the League of Legends subreddit. These scrapers follow a similar pattern: they take a URL or forum API endpoint, iterate through pages or threads within a specified time window, extract the main post text, author, timestamps, and metadata, apply simple English detection (e.g., ASCII ratio and minimum words), and write results into SQLite with per-source schemas. Fields recorded include thread or post IDs, titles, authors, dates, body text, and tags or platform information.

3.1.4 Data Storage. All harvested data were stored in a SQLite database to simplify joins across sources. Steam tables contain recommendation IDs, game names, app IDs, Steam IDs, number of games owned, number of reviews written, playtime statistics, last played, language, review text, timestamps, vote counts, purchase flags, early-access flag and whether the player primarily used a Steam Deck. Reddit tables record comment/post IDs, kind, subreddit, author, creation time, title, body, URL, score, number of comments, parent ID, inferred game title, platform and playtime, booleans for whether it looks like a review and whether it is English, a timestamp when the agent first saw it and the fetcher's version. Metacritic and official-forum tables include analogous fields for user ratings, review date, platform and main text. Primary keys or unique constraints on identifiers ensure idempotent insertion.

3.2 Phase Two: Pre-processing

For this phase, we need to identify and extract requirement-related data from the game projects selected in the last phase.

3.2.1 Time-Window and Language Filtering. To ensure a consistent temporal scope across all platforms, every data source was constrained to a two-year window measured from the extraction date. Although each scraper permits users to specify custom start and end dates, the default two-year range was used in most cases to maintain comparability between games and to manage dataset size. This constraint is enforced at the scraping stage: each tool compares the timestamp of a review against the defined window and discards entries outside it. For example, the Steam scraper terminates pagination as soon as it encounters a review older than the start date.

Language filtering is applied throughout the pipeline. Steam reviews are restricted to English through the API parameter `language=english`. Reddit comments are evaluated using a heuristic requiring an ASCII-character ratio of at least 0.85 alongside a minimum word count, while Metacritic reviews are excluded if fewer than 90% of their characters are ASCII.

Once initial collection is complete, the full database undergoes an additional unified filtering stage. All entries are re-evaluated, and any review falling outside the two-year window relative to the extraction date is removed.

3.2.2 Data merging and categorizing. After this temporal pass, the platform-specific tables are merged into a single combined dataset with shared columns (game_name, main_text, comment_platform, time) to prepare for further analysis. This unified table then undergoes a final language-detection step using langdetect, ensuring a consistent English-language standard across all sources. Duplicate entries are removed using primary keys or unique constraints. To identify which reviews are likely to contain frustration-related content—and to further reduce dataset size—we applied a sentiment-analysis stage.

We employed a hybrid sentiment pipeline that assigns each review a label of positive, negative, neutral, or mixed. The pipeline integrates: (a) a fine-tuned DistilBERT model generating a probability distribution over positive and negative sentiment; (b) the VADER lexicon for capturing sentiment polarity in social-media text; and (c) rule-based adjustments tailored to domain-specific expressions and aspect-based heuristics (e.g., phrases like "too difficult" or "low FPS" signal negative sentiment). The scores from (a) and (b) are averaged, and heuristic rules are applied to refine borderline cases. Finally, all entries whose final sentiment label was negative were extracted into a separate table, forming the subset used for downstream frustration-focused topic modeling.

3.3 Phase Three: Topic Clustering

3.3.1 Latent Dirichlet Allocation (LDA)). An initial unsupervised analysis used LDA to uncover themes in the reviews. The LDA pipeline operates as a multi-stage, streaming workflow that processes large text corpora stored in SQLite without loading the entire dataset into memory. It begins by reading documents in batches and performing an initial sampling pass in which tokenized texts are collected to compute coherence measures and, when enabled, to train a bigram phraser. A second full pass constructs a Gensim dictionary, applying tokenization, optional lemmatization, stopword removal, bigram detection, and vocabulary pruning based on document frequencies and highest-frequency terms. A third pass reprocesses all texts and converts them into bag-of-words representations, which are serialized to disk using the Matrix Market format, alongside a record mapping each internal index back to its original database identifier. The serialized corpus is then randomly split into training and held-out test partitions, allowing the model to be evaluated using perplexity.

With the corpus prepared, the pipeline trains an LDA model using user-specified hyperparameters, including the number of topics, the Dirichlet priors α and η, the number of passes and iterations, and the number of parallel workers. After training, the script computes a wide range of diagnostic metrics, including u_mass, c_v, and c_npmi coherence; perplexity on the test set; measures of topic diversity such as unique-word ratios and pairwise Jaccard similarity; and inter-topic similarity matrices based on Jensen–Shannon divergence and cosine similarity. Topic size distributions are also assessed using entropy and Gini coefficients. The model's outputs are then exported in multiple formats: top-word lists for each topic, dominant topics and full topic-weight vectors for every document, summary JSON files containing all evaluation metrics, and several visualizations such as dendrograms, heatmaps, UMAP or PCA document projections, and pyLDAvis interactive displays. Through this structured sequence of streaming passes, dictionary construction, corpus serialization, model training, quantitative evaluation, and export, the pipeline produces a comprehensive and reproducible LDA modeling workflow suited for large-scale text analysis.

We performed LDA using K (number of topics) values from 2 to 20, toggled the use of bigrams, experimented with different settings of the topic–word Dirichlet prior η, and varied how many of the highest-frequency words were removed from the dictionary. We initially attempted a hybrid strategy for selecting K, grouping runs into low-, mid-, and high-K ranges, identifying local maxima of c_v coherence within each range, and then manually inspecting representative topics to judge interpretability. However, this approach proved overly subjective. To maintain methodological objectivity, we adopted c_v coherence alone as the selection criterion and chose the run with the highest c_v value. In our dataset, the highest c_v occurred at very small K values (K = 2), indicating that LDA tended to collapse different forms of frustration into only a few broad themes. This limitation motivated the exploration of an alternative modelling approach.

3.3.2 BERTopic. Because HDBSCAN automatically determines the number of clusters based on data density, BERTopic can capture both broad themes and fine-grained subtopics without requiring a pre-defined K. To improve topic quality, we ran BERTopic iteratively, inspecting the generated topic clusters after each run. During this manual inspection, we observed that certain high-frequency words appeared across many topics without contributing meaningful semantic distinctions. These noisy terms diluted topic interpretability and masked underlying thematic patterns. To address this, we developed a custom stopword list tailored specifically to our dataset and use case.

The stopword refinement process was grounded in multiple considerations. First, we excluded function words and soft intensifiers—such as "really," "very," and "actually"—which often dominate informal review text but add little to topic coherence. Second, we removed weak verbs and general-purpose action words like "play," "get," "use," and their inflected forms, which occurred frequently but were too generic to distinguish between topics. Third, we filtered out common profanity and slang (e.g., "fuck," "shit," "lol," "lmao") that, while expressive, offered low interpretive value for topic modeling. Fourth, and crucially, we identified platform-specific spam or meme noise, such as words like "snail," "democracy," and "Planet," which appeared disproportionately due to community-driven review bombing or inside jokes (e.g., Helldivers 2 discourse). Finally, we excluded domain-specific but overly broad terms such as "game," "games," "playing," and "played," which tended to dominate topic keywords without pointing to specific sources of frustration.

This curated stopword set combined traditional English stopwords with a wide range of filler language, informal internet expressions, weak verbs, profanity, and dataset-specific artifacts. After each round of refinement, we reran BERTopic to evaluate improvements in topic coherence, uniqueness, and interpretability. This iterative tuning substantially reduced noise, yielding cleaner topic clusters that reflected distinct sources of player frustration. The final BERTopic output included a diverse and nuanced set of topics addressing the limitations of LDA, whose coherence metric tended to favor overly simplistic topic structures at very low K values.

4 EVALUATION

4.1 Results

For the results we obtained, we experimented with reducing the number of topic clusters generated by BERTopic. After exploring various configurations, we selected 30 topics as the most interpretable and semantically distinct outcome. These were labeled from −1 to 28, and we organize our interpretation below according to key software architectural qualities and non-functional requirements (NFRs) that relate to player frustration.

Before delving into specific categories, certain topics must be excluded from detailed analysis. Notably, Topic −1 comprises 307,188 entries and features high-frequency terms such as 'recommend', 'ea', 'issue', 'server', 'cheat', 'much', 'its', 'it', 'fix', and 'fun'. This cluster represents a broad and heterogeneous collection of reviews that express general dissatisfaction without forming a coherent thematic focus. In HDBSCAN-based topic modeling, such clusters are typically treated as semantic "noise," encompassing vague complaints, overlapping sentiments, or diffuse sources of frustration. While not thematically classifiable, Topic −1 still offers value as a background indicator of widespread discontent. Additional topics omitted from subsequent sections were determined to be unrelated to identifiable causes of player frustration and thus fall outside the scope of this analysis.

Several other topics more precisely reflect frustration related to usability, particularly concerning game balance and player learning. For example, topic 0 with keywords like 'nerf', 'buff', 'tank', 'support', 'balance', and 'patch' suggests players experience frustration due to perceived imbalances in gameplay roles or the dominance of certain characters or strategies. Sample reviews indicate discontent with frequent balance changes, disruptive reworks, and evolving metas that require players to relearn mechanics repeatedly. These issues hinder the game's usability by undermining consistency, fairness, and a player's ability to interact with the system effectively. In this context, balance-related complaints represent clear usability failures that contribute directly to player frustration.

Toxic player behavior also emerges as a key usability issue. Topic 13, for example, includes keywords such as 'ban', 'toxic', 'reporting', 'report', 'abuse', 'toxicity', 'blizzard', 'abusive', 'mute', and 'chat'. This topic reflects concerns about verbal abuse and toxicity in communication channels. Persistent harassment degrades the usability of core communicative features like chat and voice, leading many players to disengage from them entirely. This not only diminishes the player experience but also undermines gameplay that depends on team coordination. Topic 15 extends this concern to identity-based abuse, particularly racial slurs. Here, frustration stems from both competence-related stress (e.g., being insulted for poor performance) and relatedness frustration (e.g., being socially excluded or attacked by teammates).

Finally, Topic 23 captures frustration related to grind-based progression systems. With keywords like 'grind', 'grinding', 'grindy', 'level', 'need', 'get', 'pain', 'progress', 'recommend', and 'want', this topic highlights player dissatisfaction with repetitive gameplay loops that offer little meaningful progress. Excessive grinding compromises usability by reducing clarity of progression, undermining the perceived value of effort, and decreasing the efficiency with which players can achieve in-game goals. These factors collectively contribute to a frustrating and demotivating user experience.

Several topics reflect frustration arising from accessibility constraints—specifically when players are prevented from accessing or engaging with the game due to infrastructural or policy decisions. One prominent example, topic 1, originates from Helldivers 2, where the requirement to link a PlayStation Network (PSN) account—regardless of regional availability—prompted widespread backlash. This is evident in the topic characterized by keywords such as 'playstation', 'sony', 'psn', 'account', 'steam', 'update', and 'buy'. In this case, the frustration did not stem from gameplay mechanics, but rather from external access restrictions that rendered the game unplayable for affected users.

Another accessibility-related issue appears in Topic 12, which includes terms like 'linux', 'cheat', 'rootkit', 'kernel', 'secureboot', 'os', 'ea', 'ban', 'anticheat', and 'spyware'. This topic emerged following EA's deployment of a kernel-level anti-cheat system that mandated Secure Boot activation, effectively excluding the Linux user base. These examples illustrate structural accessibility failures, wherein system-level requirements prevent entire segments of the player community from participating, resulting in significant frustration.

Topic 19 reflects a different dimension of accessibility concern—one tied to temporal and economic demands. The topic, marked by keywords such as 'battlepass', 'premium', 'issue', 'pay', and 'update', centers on dissatisfaction with the cost and grind associated with battlepass systems. While the monetary value of such systems varies across games, players frequently report frustration when progression requires continuous and time-intensive engagement. For individuals with limited free time or external commitments, these designs exhibit low accessibility and foster autonomy frustration, as players feel coerced into daily tasks to avoid missing rewards.

Another topic express frustration caused by technical instability and degraded performance, tied to the architectural qualities of availability and performance. Topic 2, a topic of 49,729 entries, includes keywords like 'server', 'crash', 'connection', 'unplayable', 'fps', 'ping', and 'steam'. These reviews convey frustration over lag, disconnects, crashes, and low frame rates—conditions that break immersion or render a game outright unplayable. Technically, these span availability (can the system be reached?) and performance (does it run smoothly?). But from the player's perspective, the distinction is often irrelevant. When a game stutters or drops out, it interrupts the experience—prompting immediate and intense frustration. Linguistically, these breakdowns are linked in players' minds, and therefore, they appear in a shared cluster, reinforcing their emotional equivalence.

A distinct and recurring type of player frustration centers on fairness, which—though not traditionally listed in software quality taxonomies—has become a recognized NFR in player-centered systems. Topic 3 (33,420 entries) addresses this form of frustration, with keywords like 'matchmaking', 'csgo', 'rank', 'dota', 'competitive', and 'queue'. Combined with topic 28, with keywords 'smurf', 'smurfer', 'smurfs', 'smurfe', 'smurfing', 'cheater', 'cheat', 'account', 'level', 'mmr', these reviews express frustration with skill mismatches, perceptions of rigged or manipulated matchmaking, and suspicions of "win-rate balancing" that penalizes success. Even if these perceptions are subjective, they shape the player's emotional response: feeling cheated or set up to fail quickly erodes trust and enjoyment. Thus, matchmaking fairness is a powerful source of frustration in competitive games.

Finally, fairness-related frustration also manifests in complaints about cheating and inadequate enforcement of anti-cheat measures. Topic 5 (14,187 entries) includes terms like 'cheater', 'hack', 'ban', 'aimbot', and 'problem'. Players in this topic are frustrated not just by individual cheaters, but by a perceived systemic failure to uphold fairness. When cheaters go unpunished or anti-cheat mechanisms seem ineffective, players feel the game is illegitimate or not worth playing. This erodes community morale and undermines engagement—making robust cheat prevention essential for maintaining perceived fairness and avoiding mass frustration.

4.2 Future Works

The analysis effectively uncovered multiple sources of player frustration across a dataset of more than two million reviews, mapping them to established software engineering and architectural quality attributes. Overall, the findings were consistent with the game selection criteria. However, it was notable that no topic clusters emerged relating to difficulty or complaints that games were too hard. This absence may suggest that players who are averse to challenge tend to avoid such titles altogether, resulting in limited representation in the review data. Future research could further examine the broader determinants of player frustration and explore the development of tools to support game requirement engineering informed by these insights.

5 CONCLUSION

This study systematically examined the multifaceted sources of player frustration in modern video games by integrating psychological theory, software architecture perspectives, and large-scale user review analysis. By grounding game selection in established frustration frameworks—such as goal obstruction, psychological need thwarting, and expectancy violation—and applying topic modeling to over two million user reviews, we identified key frustration patterns across technical, design, social, and systemic dimensions. The findings underscore that frustration is not merely a consequence of game difficulty or failure, but a complex emotional response to perceived violations of usability, accessibility, fairness, and psychological autonomy. Importantly, this research highlights the value of aligning user feedback analysis with psychological constructs and non-functional software quality attributes, offering a more nuanced understanding of negative player experiences. The absence of difficulty-based frustration topics also suggests that aversion to challenge may manifest more subtly or preemptively, through avoidance rather than voiced complaint. Future work could further investigate underrepresented frustration types and develop proactive design tools that help developers detect and mitigate frustration triggers early in the game development process. Ultimately, a deeper understanding of frustration's underlying mechanisms can support more empathetic, inclusive, and user-centered game design.

REFERENCES

[1] American Psychological Association. 2020. APA Dictionary of Psychology: Frustration. Entry in the APA Dictionary of Psychology (2020) for "frustration.".

[2] Nicholas Ballou and Sebastian Deterding. 2023. "I Just Wanted to Get It Over and Done With": A grounded theory of psychological need frustration in video games. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY, Article 382 (2023), 27 pages.

[3] Roy F. Baumeister, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. 2001. Bad is stronger than good. *Review of General Psychology* 5, 4 (2001), 323–370.

[4] Leonard Berkowitz. 1989. Frustration–aggression hypothesis: Examination and reformulation. *Psychological Bulletin* 106, 1 (1989), 59–73.

[5] Edward L. Deci and Richard M. Ryan. 2000. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry* 11, 4 (2000), 227–268.

[6] John Dollard, Neal E. Miller, Leonard W. Doob, Orval H. Mowrer, and Robert R. Sears. 1939. *Frustration and aggression.* Yale University Press.

[7] Hellopeter. 2020. The Battle of Positive and Negative Reviews. Hellopeter Consumer Blog.

[8] Monique M. Hennink and Bonnie N. Kaiser. 2022. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine* 292 (2022), 114523.

[9] Kurt Lewin. 1959. *Field Theory in Social Science.* Harper & Brothers.

[10] Neal E. Miller. 1941. *The frustration–aggression hypothesis.* Yale University Press.

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

[12] Amos Tversky and Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5, 2 (1973), 207–232.

[13] L. Yongcheng. 2022. Behind the negative reviews of high score games: Why game reviews can't satisfy players? GameDeveloper. Published February 24, 2022.