# Unravelling Player's Insights: A Comparative Analysis of Topic Modelling Techniques on Game Reviews and Video Game Developers' Perspectives

Xinge Tong, Ian Willcock and Yi Sun

***Abstract*—Game reviews function as an important customer-created resource for game studies as they allow practitioners and developers to analyse players' opinions. Despite this, there are few studies that undertake comparative evaluations of topic modelling approaches in the context of video game data analysis or assess the results' practical efficacy. Accordingly, this paper aims to evaluate the performance of three topic modelling algorithms - LDA, NMF and BERTopic – as utilised within game reviews study and further to examine the results' reception within the video game industry. This study first uses the game No Man's Sky as a case study to evaluate the performance of different models in the same game context. According to our experiments based on Steam game reviews, the topic's Uci coherence score as identified by the BERTopic model can reach 0.279, which is higher than the other two models, with the extracted keywords allowing humans to interpret the themes when mapping them to original reviews. Semi-structured interviews with seven developers are then presented which demonstrate that the information we provided is useful to improve their games and track players' opinions.**

***Index Terms*—Topic modelling, Review analysis, Game evaluation**

## I. Introduction

ONLINE game reviews can play an important role in both reflecting game-player experiences and influencing business decisions regarding future video game development. According to Panagiotopoulos et al. [1] user-generated content, like reviews can be processed and analysed to determine how the player feels about a game's quality. Previous research has established that game reviews offer a rich source of information, which can be used to understand player opinions and concerns [2], the game's acceptability [4], bugs fixing [3], and the satisfaction a player derives [4].

However, the sheer volume of game-player comments poses developers and game researchers with challenges in analysing reviews about their games [5]. The ever-increasing amount of data means that conventional analysis methods are no longer effective, and hence the rich resource provided by game reviews are underexploited [6]. Analysis requires computational tools to automatically aggregate documents archived thematically [7], [8]. In this context, natural language processing (NLP) techniques[1] provide practical methods for video game practitioners and researchers to extract user feedback and comprehend players' opinions quickly.

[1]A full description of NLP techniques is beyond the scope of this paper. However, readers who are unfamiliar with this field may wish to consult R. S. T. Lee's Natural Language Processing [9] for further understanding.

The possibility of adopting NLP into the field of game studies, especially for review analysis, was established and validated by multiple studies [10]–[12]. Recently, there has been a focus on NLP-based systems that produce topic discovery from a huge amount of player-generated content [13], including for example, the consideration of game reviews on Steam [14]–[16]. Kwak et al. [17] argue that topic modelling (TM) can reveal topics among texts in reviews in the form of listed words. Traditional TM methods like Latent Dirichlet Allocation (LDA) and non-negative matrix factorization (NMF) have exploded in popularity in the study of game reviews using word processing techniques [17]–[19]. Despite the crucial importance of optimised hyperparameter settings [20], the effectiveness of such unsupervised techniques in textual data analysis has been heavily questioned [21]. With the recent development of TM techniques, emerging modelling approaches with advanced algorithms such as BERTopic [22] and Top2Vec [23] were released and rapidly applied to text analysis of big data [24], [25], and especially for analysing tweets [26], [27]. However, up to this point, far too little attention has been paid to the application and performance evaluation of embedding models for analysing game-related content.

In particular, there is scant research concerning the effectiveness of review analysis in the video game industry. For example, most studies in the field of user review analysis and its impact on the industry have only focused on software evaluation [28]–[32]. Such studies, however, have failed to explain directly how players view games or explore game developers' perceptions of reviews. Furthermore, although Lin et al. [2] and Kosmopoulos et al. [33] attempted to extract useful information from Steam game reviews, they fell short in investigating the adaptability of their approach in practice. The accessibility and comprehension of these NLP-based techniques' findings are challenging for game practitioners.

Owing to the lack of information regarding which TM algorithms can enhance player-generated reviews, this study initially utilises a case study approach to evaluate and compare the performance of three TM techniques: LDA, NMF and BERTopic, within the same game context. Subsequently, the model's coherence score is evaluated by quantitative assessment, three separate TM metrics are compared to calculate its performance. By employing derived visualisation algorithms, this study seeks to obtain advantageous insights that will support understanding topic similarity within the practical video game field. Consequently, we conducted semi-structured

interviews with video game practitioners to assess the efficacy of the approach in the video game industry.

**Hence, this study has four objectives:**

1) To perform a comparative study of various TM techniques, including LDA, NMF and BERTopic, to aid thematic analysis in video game reviews.
2) To identify and categorise prominent topics in video game reviews using the selected TM methods, thereby gaining insights into players.
3) To discuss the potential applications and implications of the research findings for optimising game design and enhancing player satisfaction.
4) To verify the usefulness of the TM technique by conducting interviews with video game developers, ultimately elucidating their views on player reviews.

The contribution to the study of online game reviews is twofold. Firstly, this study evaluates and compares the performance of three TM techniques for text analysis of video games reviews. Secondly, by investigating video game developers' perspectives on the topic identification findings of TM models, this study explores how useful they might be for game developers and how NLP-generated results can be most effectively presented. However, it is beyond the scope of this study to examine the feasibility of adopting TM results in actual game development.

## II. Related works

### A. Studies on community and reviews in practice

A substantial body of literature exists pertaining to online community management, while research specifically examining the practical analysis of reviews remains rather limited. However, Ruggles et al. [34] and Parker and Perks [35] explored the importance of game-related online communities for developers by focusing on correlations between the success of market games and the role of streaming in fostering community-building, respectively.

Studies of the analysis of game reviews produced by online communities and discussion of their practical role in game development have two primary foci: player experience investigation and future game improvement. Cheung et al. [36] employed a qualitative approach to examine players' initial gaming experience by integrating the analysis of reviews with developer reflection. Furthermore, McAllister and White [37], in the book Game User Experience Evaluation, noted that developer teams are motivated by observing real players' experiences to improve the game and reading reviews to identify problematic issues. By interviewing several game studios, McAllister et al. [38] made a similar point that some studios began to examine reviews in the post-launch phase and included them for regular game updates and improvements. Furthermore, Ulf [39] suggests that, to some extent, game developers could obtain potential design ideas from reviews. These studies explore how reviews can assist video game developers in practice; however, they also highlight that current methods for gathering opinions from reviews primarily involve manual classification or reading of a limited dataset, thereby establishing the need for review analysis using technical procedures.

To further investigate discussions concerning game problems and reflect the results to developers, Kamienski and Bezemer [40] employed the LDA model to identify topics existing in Question & Answer communities, alongside inquiring about developer's opinions. Their study confirmed that the issue of bug reporting was significant, with a higher level of interaction between developers and users. Although this study has demonstrated the unquestionable merit of employing TM methodologies in video game research, it could have enhanced its relevance by encompassing a wider spectrum of player communities. Consequently, this suggests examination of the extent to which the TM approaches are applied to game review analysis and how practically useful they are for developers.

### B. Topic modelling and Online (game) reviews

Before the introduction of embedding methods, matrix factorization-based and probabilistic models were the main approaches employed to solve the task of NLP topic discovery in a large corpus. Lee and Sebastian [41] proposed the concept of the NMF model based on matrix factorisation to perform semantic analysis. Although the stability and reliability of NMF are criticised [42], [43], existing research has shown that NMF and its variations were used for biological data analysis [44] and text classification [45].

Latent Dirichlet Allocation (LDA), as the first consummate generative probabilistic TM approach, is one of the most employed methods across the community at large [46], [47]. LDA has been particularly widely applied in the analysis of online customer reviews, including movie reviews [48], mobile app reviews, as well as in investigating mental health [49] and resource management [50].

With respect to topic discovery in the video game field, using the LDA method, game designers and researchers can thus explore and evaluate multiple aspects of games based on player reviews. Faisal and Peltoniemi [51] adopted LDA for video game genre classification using the user-generated tag on Steam, while Wang and Goh [6] introduced research on the components of player satisfaction using 9,333 reviews in the video games category on Amazon. Yu et al. [15] identified topics of interest to e-sports players based on community on Steam. Additionally, Kwak et al. [17] employed topic modelling to identify the critical factors that contribute to the success of games. One criticism of much existing research using LDA on video game data is the indeterministic nature of the final categories; the results can thus be challenged as new games are added and the context changes.

To process the game-related context well, Li et al. [16] took a case study approach to topic discovery using LDA in game reviews, using the game *No Man's Sky* [52] as an example and collecting reviews on Steam with the aim of understanding players' opinions and discussing the significant issues noted by reviewers to outline specific directions for improving the game to support the developers. In that work, 21 sub-topics were detected across the three main classifications of functionality, gameplay, and usability. Although the applicability of LDA

has been demonstrated, especially in the video game field, its efficacy in analysing text data has been criticised as other unsupervised learning methods [21], [53], [54].

Apart from the LDA model, the generalised linear model framework is another approach for researchers to identify topics at the document level. For example, Structural Topic Model (STM) was specifically designed for social science research [55]. Lu et al. [56] utilised the STM approach within a large dataset that contained over 85,000 reviews related to *No Man's Sky*, as a way to explore changes in players' opinions over time. This work inferred the existence of 55 sub-topics, which were summarised into 12 main topics based on review content.

Compared the two studies by Li et al. [16] and Lu et al. [56], although they differed in their TM techniques, the results in terms of topic discovery were similar across both studies, highlighting issues such as Functionality (crashing and update), Gameplay (repetitive, narrative and exploration), and Usability (Control system and graphic). A noteworthy aspect of using STM to identify topics is that it required no data preprocessing on the corpus, a peculiarity that might cause the accuracy of the word input of the model to be questioned.

To address the criticisms among traditional topic modelling approaches in the current state, the concept of transformer-based machine learning techniques, such as Bidirectional Encoder Representations from Transformers (BERT) and Pre-trained Language Models (PLMs), were introduced into the topic modelling [57]. One representative model is BERTopic, which uses a class-based version of term frequency–inverse document frequency (cTF-IDF) to extract topics based on the contextual representation [22]. While recent research has examined the availability of BERTopic, these studies have focused on social science research in general. Examples can be noted from a growing amount of literature analysing Instagram messages [58], news articles [59], travel blog journals [60] and Twitter posts [25], [27], [53]. Apart from social media content analysis, the existing study of customer-generated reviews and complaints focuses on the financial business field [61] and healthcare [62], [63]; however, very few studies have investigated the application of BERTopic techniques in video game research.

Recently, rather than relying solely on a single model for analysis, researchers have shown an increased interest in comparing and combining models together to evaluate performance [64]–[70]. Egger and Yu [53] addressed the performance comparison of four TM algorithms, namely LDA, NMF, Top2Vex and BERTopic, using the dataset from the social domain. Ultimately evaluating each model's advantages and disadvantages, their recent work establishes the importance of interpreting results for keyword formation within a particular research domain. Our contribution will build on such works by conducting a qualitative investigation with developers, not only to evaluate the performance of different models in the video game context but also to reflect on how relevant these are for video game practitioners in practice.

## III. METHOD

The present study comprises two distinct methodologies. Firstly, a comparative analysis was conducted to assess the efficacy of various TM techniques. Subsequently, qualitative interviews conducted with video game industry professionals provide insights into these techniques' practical effectiveness.

The themes occurring in the gaming community include player complaints, overall gaming experience, and discussion focusing on specific game contextual information such as core game mechanics, game genre and attitudes toward its developers, which are complex and diverse to observe and study. Accordingly, we decided to pursue a comparative study first to access the three TM: LDA, NMF, and BERTopic within the same game context – *No Man's Sky (NMS)* on Steam[2].

NMS was deemed suitable for discovery and analysis of the topics identification as it observed the topics regarding their diversity and relative differences [71], word-of-mouth changes caused by long-term development following game releases. Previous researchers have used NMS as a typical case to reflect the archaeology in digital environments [72], [73], promises marketing [74] alongside measuring changes in game quality [16], [56]; however, current works on topic identification analyse data using a single model or manual analysis, thereby highlighting the inaccuracy of the findings. Therefore, our method is an extension of Lu et al. [56]'s and Li et al. [16] investigation that further extends the application of TM and practical interpretation in the context of NMS.

Subsequent to the evaluation of various models, we utilised the best-performing model to generate topic recognition results and conduct interviews with video game practitioners to measure both the technique and usability of such results qualitatively. Figure 1 outlines the workflow of the method in this study.

### A. Topic identification

#### 1) Data preparation:

**Data Collection:** This study collected 192,114 reviews from the NMS Steam community spanning October 2016 to May 2023, including 114,533 reviews in English. A discrepancy may exist between the actual data volume of the reviews collected and the number of reviews announced on the store page, since reviews may be subject to deletion either owing to Steam censorship or by the players themselves.

This research adopted an effective and harmless customised crawler programme based on the guideline reported by Fiesler et al. [75] to obtain review data from Steam. As contents within the Steam Community are classified as publicly available information [76], these user-generated comments are suitable for use in academic research where all personally identifying information (PII) is removed. Data were anonymised before analysis, with all PII defined by Valve, including Steam ID, username, and avatar details, were removed, to maximise the anonymity, confidentiality, and safety of the dataset.

[2]https://store.steampowered.com/app/275850/No_Mans_Sky/

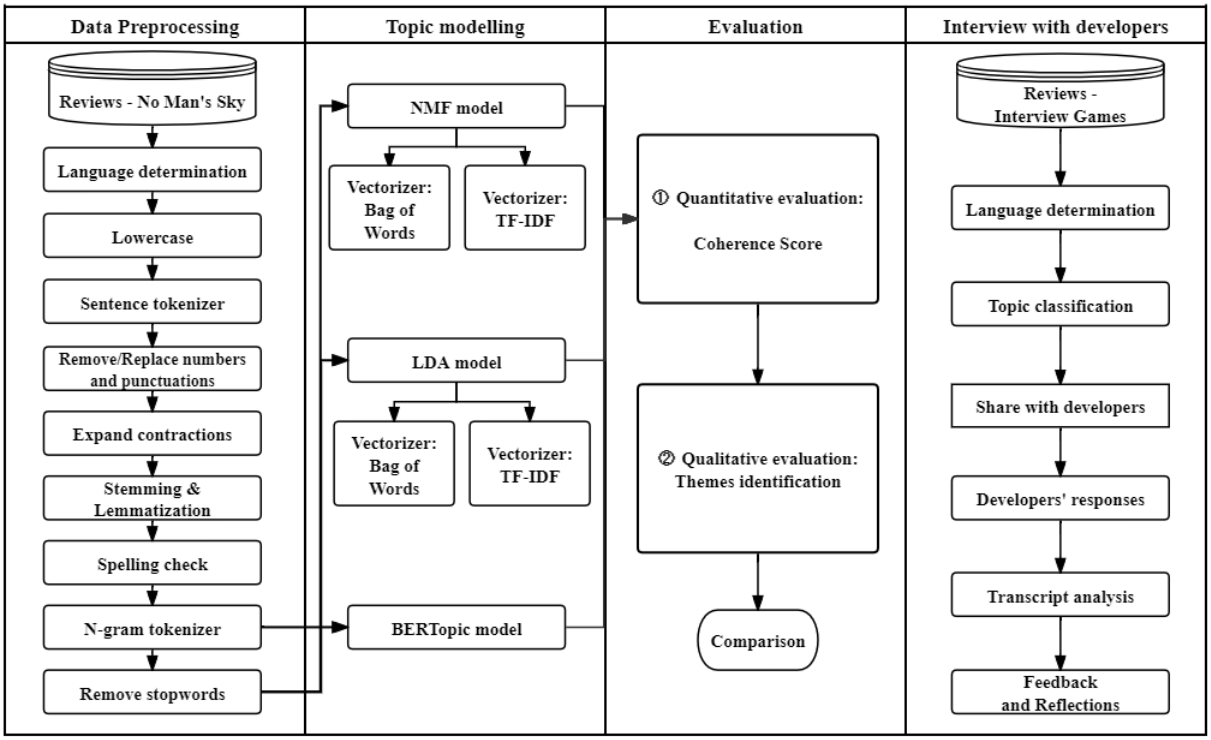


Fig. 1. Overview of methodology.

This study uses player-generated reviews, particularly from the Steam Community[3], for the following reasons: Firstly, the Playtime Tracking function on Steam not only indicates the extent of a player's engagement with a game but also verifies their authentic player identification. Secondly, there is a plentiful data resource on that community. Moreover, Steam provides friendly services for academics; its open application programming interfaces (APIs) allow researchers to dynamically access the metadata on Steam [77].

**Data Pre-processing:** Before commencing the TM, two types of pre-processing programs were used to clear the initial dataset. The first type of program was designed for LDA and NMF. The text was cleaned for each review using a standard preprocessing procedure (Figure 1). The most and least frequent words, such as *'game'*, were also removed during this process to optimise the valuable information contained in the data. Last, Bag of Words (BoW) and TF-IDF were used for feature collection on converting texts [78], [79]. They differ in their technique of creating features; the former produces word mapping based on their raw frequencies in documents. The latter considers the words frequently occurring in multiple documents irrelevant and decreases their relevance score. Using these two feature extraction techniques is a well-accepted methodology in TM comparison studies [66], [80].

To maintain the original structure of the textual data in the transformer models while employing the embedding approach, data utilised in the BERTopic model was not used in the majority of the pre-processing stages outlined above. However, where abbreviations and words identified as jargon, like *'E3'*, are featured in reviews, these words were expanded or replaced.

*2) Topic Models Implementation:* We have employed three topic-model algorithms. Given the constraints on space, it is challenging to delve into the finer details of each algorithm within this paper. However, we highly encourage readers with a keen interest in the methodologies employed to consult the provided references [22], [46], [81].

**NMF:** It employs a linear algebraic strategy for topic extraction [64]. In this study, we used Gensim[4] to develop the NMF model. According to the coherence metric, the NMF (BoW) and NMF (TF-IDF) models get the greatest Cv coherence scores of 0.734 and 0.574, respectively, when the number of topics is set to 5.

**LDA:** It assumes a topic comprises a set of words from a document. Each word in the document is given a topical assignment with a probability determined by this distribution. For a fixed number of random seeds when undertaking an experiment, the number of topics (K) and iterations (NOI) are vital parameters influencing the LDA model's performance. In this research, we consider controlling the hyperparameter settings of the LDA models by using the perplexity and log-likelihood metrics. Perplexity and log-likelihood were used as objective and generalised measures as they are commonly used metrics to evaluate probabilistic models [82], [83]. This study employed the GridSearchCV[5] to obtain the optimal parameter K and NOI values based on dynamic TM. By

[3]https://steamcommunity.com/app/275850/reviews/?browsefilter=toprated&snr=1_5_100010_

[4]https://radimrehurek.com/gensim/models/nmf.html

[5]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

simultaneously computing the log-likelihood score and model perplexity, the optimal K was determined as 5; meanwhile, 20 and 50 NOI were determined in LDA+BoW and LDA+TF-IDF, respectively.

**BERTopic:** Unlike NMF and LDA, the stochastic nature of the BERTopic model yields different outcomes with repeated modelling [53]. As a TM algorithm that supports unsupervised learning, BERTopic can be finetuned by three methods: pre-trained model, dataset pre-processing, and optimising hyper-parameters [84]. This study uses the all-MiniLM-L6-v2 model as the pre-trained sentence embedding model. All-*models are widely used, with the computing speed of the "all-MiniLM-L6-v2" model being five times faster than the all-mpnet-base-v2 model, which provides satisfactory quality. As mentioned above, the data have been adequately cleaned, thereby maximising the preservation of sentence structure. Regarding hyperparameters, this study first utilises "auto" to automatically decrease topics, merging topics with a similarity greater than 0.915 and minimising the need for unnecessary manual involvement. The method's usability and threshold values have been validated in prior research [23] and the BERTopic documentation. We also adopted the "reduce_outliers" function to reduce outliers [53] using the probability-based soft-clustering strategy employed by HDBSCAN. As a result, the study reduced the number of topics from 716 to 521. Although the number of topics obtained through the BERTopic model is typically large and difficult to display, BERtopic allows researchers to enter keywords and search for the most relevant subjects.

**Evaluation Metrics:** For objective evaluation of models, this study uses coherences values to compare model performance quantitatively. Topic coherence was chosen because it provides a practical method to quantitatively measure the semantic interpretability of topics [85], particularly for analysing online corpora [86]. Four different coherence algorithms in the Palmetto library [85] were employed: coherence measure (Cv), normalized pointwise mutual information (NPMI), UMass, and Uci, to quantify the performance of the different TM algorithms, with higher values indicating better performance. Cv is widely used to evaluate topic models in existing research [87]–[89]. However, CV's accuracy is criticised when applied to randomly generated datasets [90] and has been rarely used in recent analyses. In contrast, the UMass metric is also widely used in studies of topic classification [91], [92], which assesses word co-occurrence with not affected by bi-grams and lemmatization and focuses on document-topic distribution. In addition, Uci [86], [93] and NPMI [94] indicators are also suitable for establishing standard measures; the latter especially has been used to compare the performance of unsupervised models and BERTopic [24]. Notably, while these coherence metrics provide a quantitative evaluation of TM performances, they are not infallible, and human intervention and interpretation are sometimes necessary to make sure the identified topics make sense to researchers.

### *B. Interviews with video game developers*

Semi-structured interviews were conducted in the second study to contextualise the TM techniques and offer in-depth reflections on their usability in the video game industry. Firstly, we used the appropriate modelling technique to analyse reviews in the Steam dataset and then generated TM findings reports. Subsequently, we organised interviews to consult developers about their opinions and feedback.

*1) Data collection:* Interviews were held online with a total of 7 practitioners from 5 distinct game companies following a strict ethics protocol approved by the authors' university ethics committee[6]. The practitioners we conducted interviews with have extensive expertise in game development, project management, and player support, with the cumulative length of all the meetings amounting to 187.41 minutes (M=37.48, SD=9.74). The interviews contained semi-structured questions focusing on three key themes: Game review perspectives, the interpretation of the results of topic identification and the degree to which they regarded these technologies as being beneficial to their needs. Notably, the experimental procedure remains constant among all our participants.

Table I contains information about the participants and their respective organisations. In order to minimise biases, developers were chosen for interviews to ensure representation from a range of game companies of different sizes. Crucially, this study prioritises the utmost anonymity and confidentiality for everyone involved, whether directly or indirectly. Thus, in the subsequent passage, the identities of the individuals have been rendered anonymous. The letter "P" and the ordinal number assigned to each interview serve as their representations. The term "Case" and the ordinal designation of the interview serve as identifiers for their organisation.

TABLE I
DESCRIPTION OF INTERVIEWEES.

| Round | Organisation | Team size* | Number of released games | Participants |
|---|---|---|---|---|
| 1 | Case A | Small | More than 20 released products | P1 |
| 1 | Case B | Small | Less than 5 released products | P2, P3, P4 |
| 2 | Case C | Micro | Less than 10 released products | P5 |
| 2 | Case D | Small | More than 15 released products | P6 |
| 3 | Case E | Micro | More than 20 released products | P7 |

*The production team sizes are as follows: Micro$<$ 10, Small 10-50 [95]

The interview research was conducted in three rounds, each focusing on different representations of the topic identification findings. During the initial round of interviews, we presented the ten most prevalent keywords derived straight from the algorithm as outcomes for topic identification. Then, during the second round of interviews, the outcomes were substituted with easily comprehensible topics, such as music, game genres, mood, and so forth. These themes were carefully interpreted using keywords, original reviews and game context. In the final round of interviews, alongside the topics, representative reviews included in each topic are also presented.

*2) Data Analysis:* The conversations conducted during the interviews were digitally recorded. The audio was transcribed using the auto-transcription function offered by communication platforms, followed by interviews coded using NVivo 12. Moreover, the qualitative interviews were analysed using the coding procedures of a grounded theory approach, thereby

[6]The protocol approval number: aCTA/PGR/UH/05469(1)

following the guidelines established by Strauss and Corbin [96]. Furthermore, the principles of theory formation and result interpretation as described by Eisenhardt [97] were also adhered to.

Following the methodology outlined by Corbin and Strauss [98], the interviews underwent inductive coding for analysis. The codebook finally consisted of a total of 18 codes. During the axial coding stage, the codes were organised into four sub-categories and three major categories based on their connections.

## IV. Results

### A. Evaluation of models

We evaluated three topic modelling algorithms with both bag of words and TF-IDF. Table II clearly shows that the overall performance of BERTopic is significantly higher than LDA and NMF, especially according to the NPMI (0.093), Umass (-0.84) and Uci (0.279) metrics. In contrast, The LDA+TF-IDF model gives the lowest coherence values with most of the metrics. The coherence scores of NMF+TF-IDF are also lower than BoW with the same modelling approach. These results suggest that unsupervised models using TF-IDF for feature extraction may not be suitable for this study's data.

Interestingly, the performance evaluation using the Cv metric is not consistent with that obtained using other metrics. Specifically, using the Cv metric results in better performance with NMF+BoW than both BERTopic and LDA. This inconsistency may be due to the different focus of the metrics used to assess TM quality and the possible instability of the Cv metric, since the data preprocessing steps are the same in all metrics. This observation also underscores the need to employ multiple measures to establish agreement in unsupervised TM, rather than depending on a single metric.

TABLE II
PERFORMANCE EVALUATION OF THREE TOPIC MODELLING METHODS USING TWO FEATURE EXTRACTION APPROACHES. FOR EACH METRIC, THE COLOUR SCALE (BLUE-WHITE-RED) DEPICTS HIGH, MEDIUM AND LOW VALUES. THE BEST SCORE FOR EACH METRIC IS SHOWN IN BOLD.

| Models | Cv | NPMI | Umass | Uci |
|---|---|---|---|---|
| NMF+BoW | **0.734** | 0.036 | –3.165 | 0.161 |
| NMF+TF-IDF | 0.574 | 0.015 | –3.814 | 0.006 |
| LDA+BoW | 0.613 | 0.034 | -3.232 | 0.158 |
| LDA+TF-IDF | 0.634 | -0.425 | -13.435 | -11.726 |
| BERTopic | 0.595 | **0.093** | **-0.840** | **0.279** |

### B. Topics identified by NMF and LDA

Table III shows the topics identified using NMF and LDA on the NMS's review dataset with the top 10 words. The researcher conducted a comparative analysis of the main contributing factors of 10 cluster terms. These terms were then cross-referenced with the game's external contextual knowledge. The purpose of this step was to manually summarise and assign themes to the topics generated by the TM models, as well as to confirm that the same approach to topic labelling was applied to the different methods. For example, the first and fifth topics, identified by the BoW variations of the NMF and LDA models have a shared underlying theme relating to in-game content. This is because terms like *'planet', 'ship'* and *'resource'* mainly refer to the game's core components or fundamental constituents, and seldom occur together in other topics simultaneously.

This joint table reveals several interesting co-occurrence patterns. First, some specific words always appear together in certain topics. For example, the terms *'explore'* and *'exploration'* often appear in pairs, with a high probability of being accompanied by the word *'space'*, which would be appropriate descriptors for the gaming world that the players were exploring. The simultaneous occurrence of all these three terms in a given topic may reflect NMS's exploration mechanism, such as Topic 1 in NMF+TD-IDF and LDA+BoW. Similar word combinations include *'hour'* and *'play(ed)'*, which clearly refers to the number of hours a player has spent playing a game. Unlike the co-occurrences of word pairs centring around the features of a particular game, this combination applies to games in general. Identifying general co-occurrence patterns gives the themes distinct meanings, allowing them to refer to different aspects of the specific game, as well as to aspects of the actions and processes common to most video games, such as expressing game progression (Topic 2 in both NMF+BoW and LDA+BoW), all of which enhances the reviews' credibility (Topic 4 in NMF+BoW and TF-IDF).

Second, high concentrations of emotionally charged words, such as *'fun', 'love', 'boring',* and *'angry'*, in some topics, identifies those topics as about players' subjective feelings, moods and emotional responses. Such topics were subsumed under the theme "Reactions" and included Topics 2, 3 and 5 in NMF+BoW, Topic 3 and 5 in NMF+TF-IDF, Topic 2 in LDA+BoW and Topic 3 in LDA+TF-IDF. One intriguing aspect of this theme is that it can be viewed as a self-contained theme and also integrated with other thematic elements in a separate theme. Stand-along themes seldom include words that have game-specific significance, such as Topic 5 in NMF+BoW, but instead include those that describe aspects common to all or most games and game playing, such as *'story'* and *'gameplay'* in Topic 3. This allows them to integrate human emotions while articulating distinct facets of the game. Importantly, the usage of terms such as *'grind'* and *'chore'* in the context of NMS, typically conveys the sentiment associated with performing a necessary but tedious action to complete a task. It is worth emphasising that the perception of a task as being grinding is highly subjective, and different from one player to the next. While certain individuals may get pleasure from a gameplay's repetitive and laborious nature, others may perceive it as a cause of annoyance or dissatisfaction. In this instance, Topic 3 in NMF+BoW and LDA+TF-IDF are, to some extent, categorised as the reaction theme. Here is an excerpt from one review:

> *'Base decoration may be an appeal to the intrinsically motivated, but be locked behind ten hours of grind and turn it into a chore that punishes experimentation.'*

Apart from the player's emotional reactions, more intuitive information is encompassed in identified themes, which could help developers improve the game. For example, the term *'bug'* appears repeatedly in the findings of the BoW variants

TABLE III
TOPICS IDENTIFIED FROM NMS'S REVIEW DATASET

| NMF+BoW | | | | | NMF+TF-IDF | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Topic 1 In-game content | Topic 2 Evaluation and Reactions | Topic 3 Gameplay and Reactions | Topic 4 Evaluation | Topic 5 Reactions | Topic 1 Exploration mechanic | Topic 2 Support for developer | Topic 3 Reactions (mixed) | Topic 4 Evaluation | Topic 5 Reactions |
| planet | get | good | like | thing | great | good | love | play | fun |
| go | hour | grind | play | still | like | pretty | 10/10 | no | space |
| ship | back | bad | update | even | get | job | absolutely | update | pretty |
| resource | start | long | time | no | much | actually | explore | go | much |
| another | first | story | one | much | amaze | work | exploration | still | friend |
| collect | money | average | see | lot | explore | hello_games | labor | hour | cool |
| new | try | bug | hour | fun | lot | luck | far | time | super |
| find | boring | easy | give | great | exploration | story | want | say | relax |
| explore | lose | gameplay | review | space | space | virtual_reality | hello_games | recommend | explore |
| space | take | price | say | many | thing | bad | every | see | Minecraft |
| LDA+BoW | | | | | LDA+TF-IDF | | | | |
| Topic 1 Exploration mechanic | Topic 2 Evaluation and Reactions | Topic 3 Support for developer | Topic 4 Price-quality (PQ) ratio | Topic 5 In-game content | Topic 1 Other games | Topic 2 Improvement | Topic 3 Reactions | Topic 4 Game Maintenance | Topic 5 Unspecific |
| like | hour | update | promise | planet | hang | insane | chore | stage | nope |
| great | play | good | run | ship | flat | upset | genuinely | brilliant | everybody |
| fun | time | review | people | different | skyrim | pain | diverse | developed | scenery |
| space | worth | release | buy | thing | mislead | polish | screw | euro | dedication |
| exploration | bad | hello$_games$ | issue | base | creator | majority | brings | five | pile |
| explore | start | 10/10 | like | resource | tire | balance | angry | smoothly | period |
| want | bug | developer | $60 | new | fault | introduce | plot | involve | regular |
| sky | boring | recommend | refund | space | spoiler | funny | pure | get_well | desire |
| experience | good | year | lie | need | kudos | nowhere | proper | playing | upgraded |
| amaze | try | launch | steam | like | mainly | sight | freeze | exocraft | hour_played |

of TM, which directly reveal the disadvantages that the computational and optimisation aspects of NMS have on the game quality. Furthermore, Topic 2 of the LDA+TF-IDF detailed players' dissatisfaction with the overall quality and the degree of refinement of the game (e.g. *'polish'*). This information provides developers with valuable insights that can be used to enhance the game, such as making adjustments to the level of difficulty (e.g. *'balance'*) and scene visibility (e.g. *'sight'*).

Turning to another theme, "Support for developers", both models produced this topic (Table III). What is most significant about this theme is that the topics all directly relate to the game's development team, such as *'hello_games'* and *'developers'*, along with words denoting their actions, such as *'update'*, *'work'*, and *'job'*. Interestingly, the topic identified by LDA within this theme may be more related to the marketing of NMS, as it contains terms such as *'release'*, *'year'* and *'launch'*, terms which reflect the pre-release schedule and marketing, including references to disappointment in its ability to fulfil players' expectations. In contrast, the topic identified by NMF is more related to new features the developers added. For example, the term *'virtual_reality'* occurs, which was an additional mode only introduced years after NMS was released.

Another notable topic of interest among players is the price-quality (PQ) ratio of the game. Table III shows that the terms that appeared in identified topics, such as *'money', 'price', 'refund', 'buy'*and *'$60'* (NMS's retail price), clearly relate to the evaluations of the game's value for money (e.g. *'worth'*). The themes frequently associated with these terms typically reflect the players' assessment of NMS, particularly about whether the quality of the game justifies its price. Topic 4, however, which was identified by LDA+BoW, is slightly different as it focuses on gaming issues and includes terms such as *'run'* and *'issues'* to evaluate the value of NMS and purchase policy on Steam (e.g. *'refund'* and *'steam'*). These particular aspects are not commonly observed in the other subjects under consideration.

Despite the performance of both models being similar up to this point, the LDA+TF-IDF model generates topics that include terms that rarely occur in other models. For example, there are references to the in-game landscapes, with words like *'freeze', 'scenery',* and *'flat'*, were seldom found in other models. Here is an example:

> *'Some planets are fairly flat and barren; others have big mountains; others are riddled with narrow ravines; some are mostly covered in water.'*

Additionally, the discussion of other games (such as *'Skyrim'* and *'Minecraft'*) was only observed in TF-IDF variants of TMs. While such games have some similarities to NMS in terms of gameplay, no comparative words were found, and further research would be necessary to determine the reason for this.

What emerges from the results is that some common themes emerged from the NMS reviews, such as Exploration mechanics, In-game content, experience evaluation, improvement suggestions and developer-related, as well as some interesting

TABLE IV
TOPIC IDENTIFIED BY BERTOPIC FOR SELECTED KEYWORDS.

| Keywords | Similar topics (Top 3) | Cluster words |
|---|---|---|
| 'explore' and 'exploration' | Topic 2: Exploration mechanic | exploration, space, explore, adventure, if, love, survival, like, best, you |
| | Topic 183: Reactions to exploration mechanic | adventure, relax, thrill, exploration, explore, adrenaline, own, mystery, story, enjoy |
| | Topic 125: Core mechanics | craft, survival, crafting, exploration, combat, explore, grind, building, gathering, element |
| 'crash' and 'bugs' | Topic 1: Errors in different Player Mode | multiplayer, bug, crash, buggy, glitch, fix, player, crashed, break, single |
| | Topic 149: Errors in saving progress | save, freighter, ship, bug, exit, stuck, lose, inside, autosave, crash |
| | Topic 511: Errors in specific star system | broken, broke, elate, terrify, piece, dorothy, allison, still, af, mead |
| 'grind' | Topic 36: Reactions | grind, grindy, fest, reward, bit, grinder, material, boring, repetitive, but |
| | Topic 230: Reaction to gathering and crafting | grind, resource, material, gathering, grindy, gear, resources, collect, stuff, craft |
| | Topic 104: Resource gathering | mining, mine, element, miner, resource, material, sell, mineral, stuff, periodic |
| 'chore' | Topic 100: Reactions and Motivation | gratification, confuse, excitement, pointless, instant, over, simple, easy, attention, chore |
| | Topic 176: Motivations | pointless, progress, purpose, complicate, life, distraction, distract, lose, useless, strategy |
| | Topic 44: Reactions | boring, repetitive, dull, bland, fast, tedious, quickly, very, boredom, get |
| 'developer' | Topic 13: Support for developer | developer, developers, team, their, development, work, they, promise, community, respect |
| | Topic 27: Game maintenance | content, developer, add, update, they, release, support, lack, new, continue |
| | Topic 130: Team type | indie, aaa, title, studio, triple, an, price, team, small, company |
| 'hello games' | Topic 6: Support for developer | hello, games, their, thank, job, work, update, they, promise, deliver |
| | Topic 410: Praise to developer | redeem, themselves, games, hello, redemption, arc, yourselves, itself, their, villainous |
| | Topic 250: Followship | gamer, gamers, appeal, old, type, fellow, people, who, everyone, year |
| 'worth' and 'buy' | Topic 3: PQ ratio (considering discount) | sale, price, worth, buy, money, full, buying, wait, purchase, bought |
| | Topic 5: PQ ratio (considering price) | 60, worth, dollar, 30, 20, price, 40, sale, tag, buck |
| | Topic 311: PQ ratio (considering updates) | stand, price, content, sale, picked, tag, worth, update, next, bought |
| Other games: 'Minecraft', 'Skyrim' | Topic 334: Elder Scrolls, Fallout, The Witcher, Cave Quest | skyrim, fallout, oblivion, elder, scroll, witcher, like, quest, cave, space |
| | Topic 55: Minecraft; Elite Dangerous | minecraft, space, like, terrarium, elite, basically, kind, if, in, similar |
| | Topic 399: Call of Duty | cod, codders, duty, call, doom, battlefield, who, action, mod, someone |

topics, including value for money and other games discussions. However, the identified themes could only be expressed in obscure ways due to a lack of contextual information processing. Additionally, some topics overlapped, which called into question the efficacy of applying LDA and NMF to the field of video games.

### C. Topics Discovered by BERTopic

Due to limited space, we cannot display the complete list of 521 identified topics taken by BERtopic. However, we used the search functions on BERTopic to filter the topic results based on the observed representative terms in the previous observations. Table IV gives the filtered topic outcomes based on 10 representative words or phrases; the topics were then named by the same method used in the previous models. The table displays the three most pertinent topics for each selected keyword. To construct a comparative study with NMF and LDA, the discussion of the BERTopic results focuses on five aspects: game mechanics, program issues, developer-related, PQ ratio, and other games.

The primary focus of NMS is the exploration of the universe, during which several gameplay features are incorporated, including survival, combat, trading, and base building. NMS's core game mechanics are reflected in the topics related to the terms *'explore'* and *'exploration'*, the retrieval of which generates a few clusters related to the concept of exploration. As shown in Table IV, Topic 2 mostly reflects the basic gameplay mechanics, focusing on space travel and survival (such as *'adventure'* and *'survival'*). Topic 183 also contains the main terms of the exploration theme but includes several outlier terms which relate to personal reactions, like *'relax'*, *'thrill'*, and *'adrenaline'*. In contrast, Topic 125 incorporates a broader range of game activities in addition to exploration, including combat, resource gathering, crafting, and construction. It is clear that BERTopic can not only identify the particular theme cluster but also connect the player's personal feelings to specific elements to form a detailed and progressive player feedback analysis of various aspects of the game.

To find topics related to bug reporting, the terms *'bug'* and *'crash'* were chosen as filters for the topic selection. What is notable in the table is that the filtered topics not only reflect the detection of bugs, but also identify the precise areas of the game affected by them. Specifically, Topic 1 focuses on the game's player mode and associated glitches, with terms such as *'multiplayer'*, *'single'*, *'player'*, *'glitch'* and *'break'*. Topic 149 explores NMS's save functions and their integration within the game, using terms like *'save'*, *'autosave'*, *'freighter'*, and *'ship'*. In NMS, ships could serve as designated save points, where the player's progress can be preserved by embarking. Consequently, this topic includes words associated with problems that occur when individuals attempt to save their game but experience issues, such as being unable to proceed (e.g. *'stuck'*) or save their progress (e.g. *'lose'*). In contrast, the issues reflected in Topic 104 correspond to specific in-game content, since it pertains to the star system 'Dorothy' within the NMS universe - a name rarely found in other topics. The findings demonstrate the capability of the BERTopic model to detect potential flaws inside distinct components of the game. This information is clearly helpful for developers in locating and addressing glitches and implementing efficient

game maintenance strategies.

To further clarify a player's experience of playing NMS, we extracted topics related to the terms *'grind'* and *'chore'*, which relate to player preferences. For the theme related to grinding, Topic 36 mainly expressed the sense of grinding by resource-outcome imbalance (*'reward'*) and repetitive actions (*'repetitive'*). Topic 230 adds to this by including game-specific elements, such as *'material'*. In contrast, Topic 104 focuses more on the specific actions of mining and selling resources. Additionally, compared with the grinding theme, the identified topics related to *'chore'* during playing highlight the player's motivation (*'gratification', 'pointless', 'distraction'*) and frequently combine with more negative adjectives, such as *'boring', 'confuse'* and *'tedious'*. Hence, one of the reasons for poor motivation when playing NMS could be the lack of clear objectives. It would help developers have more detailed information about the complaints expressed by players to enable them to improve the game accordingly. For example:

> *'Almost the entire time I played the game, I was trying to find the purpose of the game.'* (Topic 176)

The results also reflect the player's awareness of the development team. For example, Table IV shows how the term *'developer'* primarily revolves around topics like encouragement (Topic 13), game maintenance (Topic 27), and types of teams (Topic 130). Furthermore, it is clear from the game's background information that Hello Games is widely acknowledged as the developer of NMS, making the term *'hello games'* a suitable grouped search keyword for filtering topics related to the development team, including players' appreciation of their work (Topic 6). Such topics were characterised by the focus on the developer's efforts and the presence of lexemes denoting affirmative sentiments, such as *'respect', 'thank'* and *'promise'*.

In the secondary and tertiary topics, more gaps are found in cluster words between the two terms. In other words, the terms found in Topics 27 and 130 under the *'developer'* classification do not intersect with the terms found inside Topics 410 and 250 under the *'hello game'* classification. These topics differ in their emphasis on the development team and the expression of attitudes. The former has objective terms of appraisal and evaluation of the team and their efforts, while the latter includes sentiment words that convey admiration towards the developer and the connection between the player and developer. Additionally, when players address developers by their names instead of employing a collective noun, they may be expressing a sense of reverence (Topic 410) and intimacy (Topic 250).

BERTopic provides a more comprehensive analysis of PQ ratio judgements than obtained in previous analyses. For instance, Topic 3 revealed the players' focus on sales and discounts, and Topic 5 specifically addressed the precise pricing of the game, including its purchase prices. In comparison to the other topics under this thematic category, Topic 311 places more emphasis on assessing whether the game's price is justified by its updated content. This topic includes a range of terms associated with updates, such as *'content', 'update',* and *'next'* (name of NMS's patch), and reveals the factors that influence players' assessments of the game's value for money. It also offers novel perspectives for gamers about other players' concerns about one-time purchase games.

In addition to the discussions of NMS itself, there has been considerable discussion of other video games. In contrast to the overlapping categories obtained through NMF and LDA, the games reflected in the BERTopic results are well classified based on player opinions. Firstly, the game *The Elder Scrolls V: Skyrim* and *Fallout* noticed in Topic 334 can both be referenced for their gameplay pacing and compared to NMS. Secondly, games such as *Minecraft* and *Elite Dangerous*, as discussed in Topic 55, have similar basic gameplay elements as NMS, as indicated by the use of the word *'similar'* itself. These elements include an open-world sandbox environment, building mechanics, space exploration, trading, resource acquisition, and crafting. Finally, Topic 399 specifically highlights the game *Call of Duty (COD)*, which has an entirely different gameplay to NMS. It might be useful to explore the differences between whole game genres in addition to those between particular games. In general, player mentioned such well-known games to provide comprising with their experience of playing NMS. These references help both players and developers (including the creators) understand how NMS is perceived by players and how the gaming experience it offers compares to other games. Here are some examples of reviews referring to other games:

> *'I feel too many people expected it to have the same pace and feel as Skyrim or Fallout, but it is much more chill and slow-paced than that.'* (Topic 334)
>
> *'Minecraft meets Elite Dangerous, then it had a baby with Evochron Mercenary.'* (Topic 55)
>
> *'if you're looking for a cod type run and gun in space, then this isn't your cup of tea.'* (Topic 399)

Overall, many of the BERTopic results from our analysis are consistent with those derived using the NMF and LDA models. However, the BERTopic results offer a more comprehensive and nuanced understanding by incorporating additional details and contextual factors. Despite the redundancies in the generated topics, users can still extract relevant and helpful information from a vast corpus by retrieving keywords.

### *D. Semi-structured interview*

*1) Categories:* The core category, Perception and Practicality of Topic Identification, represents the practitioners' subjective perspectives and reflections on game reviews and TM outcomes in the video game industry. This is divided into three different focus areas: Viewpoints of Game Reviews, Viewpoints of Topic Identification results and Developer Satisfaction & Practice Reflection.

The category *Viewpoints of Game Reviews* pertains to the perspectives expressed by interviewees in online game reviews. It reveals interviewees' attitudes towards existing reviews and identifies the content they consider valuable in practice. The category *Viewpoints of Topic Identification* describes interviewees' reception of the TM results. Besides their overall opinions, we also focus on the comprehensibility of the result representation, attention to specific topics, and thoughts on how the topic changes over time. The category

*Developer Satisfaction & Practice Reflection* describes how useful the technologies are to the interviewees. During this phase, the helpfulness and usability of TM techniques are well examined.

*2) Observation:*

**1. Useful reviews usually contain detailed descriptions about the precise content of the experience**

Undoubtedly, reviews are useful for developers to improve the quality of their games. According to P4, players would like their opinions to be heard, and they give considerable feedback. Their reviews can be both positive and negative. Positive reviews can be beneficial since they help boost developers' moods (P2) and determine people's preferences when producing DLCs or similar games (P7). In contrast, some respondents are more interested in negative reviews because they normally contain game elements that need fixing, typically bugs (P2 & P7).

> *"Any negative review because of a bug is helpful because you can try to fix that."* – P7, Case E

Furthermore, most respondents evaluated the usefulness of reviews based on the descriptive content rather than the emotion expressed. The evaluation of the useful information by most of the respondents gives a good understanding of the specific game experience, including a description of what players enjoyed (P3) and what went wrong (P1) while playing. For example, P1 mentioned that some reviews are not helpful since they only record the presence of an issue, such as game crashes, without providing specific details regarding the location, timing, or overall characteristics of what went wrong.

> *"The most helpful reviews . . . are giving you the descriptions in detail about what it is exactly that they enjoy so much about."* – P3, Case B

**2. The efficacy of TM in accurately identifying and categorising subjects within a given dataset could be deemed satisfactory within the domain of video games.**

The study's most significant finding was that most participants were pleased with the TM outcomes or at least considered the identified topics as reasonable. All respondents provided favourable feedback on the TM outcomes and remarked that they had not seen this kind of analysis before. According to P1 & P4, the TM results effectively highlighted problems with their game, such as performance and localisation issues. P5 also confirmed the accuracy of the identified topics based on their professional understanding of the game and personal reading of existing reviews. Additionally, P5 referred to their observation of specific themes related to developers and updates and how these reflected the effectiveness of game maintenance and community management.

**3. TM results could serve as indicators of the right direction, guiding attention to key areas, and informing future decisions.**

All respondents affirmed TM's usefulness at different levels in their discussions of the findings. Identifying what the players were discussing helps developers know whether they are on the right track for game development (P1) and facilitates the marketing department's recognition of compelling sales features (P3). P5 and P6 supposed that by identifying what people care about most, the findings could benefit the development of sequels or similar games in the future. Even though P7 thought that the themes in their data analysis did not vary greatly over time, tracking what people were talking about still provided useful insights.

**4. More contextual information and steps are needed to process and interpret the TM results.**

In the initial round of interviews (Case A and B), where respondents were provided with lists of keywords, they experienced confusion, afterwards indicating they needed more contextual information to understand better and interpret the results. After the interviewer explained the TM technology to them, they could combine their development logs to find commonalities, such as bug reports (P1) and game updates (P3).

Next, we provided generalised themes to represent the topics for the second round of interviews with Case C and D. Participants were quick to recognise more specific topics, such as discussions about free weekends, developers and updates, and game genres. P5 noticed that it was interesting to see some descriptions of game mechanics which are out of Steam tags. At this point, respondents also offered some suggestions for clarifying the meaning of the results, including filtering out meaningless topics (e.g., game titles, according to P5) and identifying trends in players' sentiments for each topic (P6).

In the final round, and to help participants interpret the results, we provided the overarching themes together with examples of the original sentences or paragraphs for Case E. One positive point that P7 made is that this kind of analysis is enjoyable and more useful than simply searching. Using sentences to aid comprehension of the TM outcomes was not always seen as an advantage, with P7 suggesting that an accurate summary of a review's overall context would be more useful than a specific text presented in isolation. This is quite different to previous research on TM, such as those conducted by Gao et al. [32] and Albalawi et al. [70], where splitting longer texts into short sentences or paragraphs is preferred. Thus, it is demonstrated that there still is a gap between human comprehension and TM techniques in the requirement and processing of contextual information.

## V. Discussion and Limitations

### A. Comparison of NMF, LDA and BERTopic

The proliferation of player-generated material has had a substantial impact on game evaluation. More information about how players perceive a game and how they engage with it is available and accessible online, and offers a rich source of data to be further explored.

Earlier studies already pointed out the importance of introducing machine-learning approaches to video game analysis, but the existing literature has been restricted to traditional modelling techniques, such as LDA, K means, SNM, and Naive Bayes [6], [33], [51], [80], and even then usually employing only one method. This study examined how different TM techniques facilitate the thematic analysis of video game

reviews. We evaluated three TM algorithms with two feature selection representations.

The quantitative evaluation found BERTopic the most effective in analysing Steam game reviews with respect to topic coherence. NMF showed slightly superior performance compared to LDA in terms of topic coherence, which is consistent with the results reported by Ray et al. [89].

One interesting finding is that neither the NMF nor the LDA model performs well with TF-IDF compared to the BoW. Anantharaman et al. [66] also found the LDA model performed worse with TF-IDF compared to BoW, attributing this to the probabilistic nature of the LDA model. Another possible explanation for this finding is that topic models like LDA and NMF were designed to analyse longer documents [99] than online reviews, making results based on the latter an inappropriate test of their efficacy. Given the nature of TF-IDF, it is not surprising that word frequency extraction algorithms suffer more from document size.

Next, humans identified some common themes with different TM, including core game mechanics, game evaluation, development team, players' responses, and game experience. The outcomes of the topic classifications in this study are consistent with previous studies on game reviews for NMS. For example, topics related to evaluating gameplay, updates, promotions, quality levels, and glitches, were also mentioned in studies conducted by Lu et al. [56] and Li et al. [16]. However, unlike those studies, this study enhanced the clarity and specificity of themes, as well as providing a comparative analysis of the distinct emphases associated with each topic using BERTopic.

Both the LDA and NMF models have the advantage of not requiring prior domain knowledge and as well as being computationally efficient. However, as mentioned earlier, one problem the models share is the hyperparameter settings [20], [53]. In this study, we used objective evaluation metrics to help reduce the impact of this on the optimal number of topics.

The topics discovered by NMF are more ambiguous and characterised by more terms relating to the player's responses and subjective experiences, as well as lacking coherence between the delivered topics - also seen by Egger and Yu [53]. What is more, very few of the themes identified by NMF+TFIDF provide information that could conceivably help developers improve their games.

In contrast, LDA-based generated topics are straightforward for humans to summarise by comparing co-occurrence and exclusion patterns across cluster terms, especially those produced with the BoW variant. Although overlapping clusters in the LDA model were also observed in this study, similar to Passos et al. [47], topics identified by LDA+BoW are relatively distinct. However, the same cluster words generated with the LDA+TFIDF model are rarely observed in the other models, and the themes are difficult to summarise, at least from the perspective of human cognition. One possible explanation is that the TF-IDF algorithm assigns greater significance to infrequent words [79], which could result in categories emerging that relate to a highly specific technical area with its own specialised terminology.

The current investigation found that BERTopic generates a large number of topics, a finding also observed in study made by Egger and Yu [53], where having to inspect and reduce outcomes was emphasised, as well as Sharifian-Attar et al. [84], where value intervals for topics was set manually. BERTopic may not provide a particularly concise overview of players' concerns due to the redundancy of some generated topics and terms. However, unlike NMF and LDA, BERTopic is very good at capturing the nuances and semantics of textual materials through the generation of contextual word embeddings [57]. For example, identifying and categorising bug reports and glitches within reviews can be subdivided into distinct subjects that centre around different aspects of a game, an observation that has received little attention in previous TM studies on NMS. BERTopic's ability to do this when analysing game reviews is an advantage that could help developers to optimise game design and enhance player satisfaction. Hence, adopting a TM algorithm that identifies which aspect of a game is problematic, and extracts what players commonly reflect and care about in the way BERTopic does, could be very useful for the video game industry.

Therefore, this study proposed using BERTopic as the preferred method over NMS and LDA for extracting detailed and innovative themes in the analysis of game reviews. To verify this proposal, developers' views on TM will be examined together with the specific features of the TM results that align with their requirements.

### *B. Developer perspectives*

This study confirmed previous findings that game reviews are generally considered of value to video game practitioners. It further provided strong evidence for previous statements about the definition of helpful game reviews [2], [29]. The most obvious finding to emerge from the analysis was that players' game reviews need to be able to pinpoint exactly where any problems with a game occur. This could have been anticipated from previous studies, in which, for example, Lewis et al. [3] provided a detailed summary of more than 11 subcategories of player-reported video game failures. This suggests that BERTopic could be of more use to developers than other TM algorithms since it has the ability to segment subjects.

In this study, BERTopic was used to generate TM outcomes from various game datasets. The findings demonstrate the system's efficacy in handling the diversity of topics included in different datasets and of generating valuable insights about each game. The semi-structured interviews with developers verified the efficacy of our technique in practice. The participants considered our TM technique to have the potential to enhance the quality of existing games and to provide valuable insights for developers in future game development.

While the current TM system appears to function well at this point, participants noted certain weaknesses in the presentation of results, including the lack of theme interpretation, data filtering, and reflection on user reviews. This was not surprising, as non-mathematical explanations and experiments have rarely been provided in previous studies, nor applied in practice. Moreover, the uncertain nature of unsupervised learning has

meant that most studies comparing TM methods have been based on their evaluation on statistical measures, including precision, recall, and F-score, such as Albalawi et al. [70]. To gain a comprehensive understanding of TM performance in practice, additional studies will be needed to address the gap between computer-generated results and results acceptable and readable for non-computer professionals.

### C. Limitations

This research has several limitations that need to be acknowledged. Firstly, the comparison study focused on reviews collected from a specific game, with the unsupervised TM methods evaluated on coherent contextual information we collected. The limitation to a single game inevitably reduces the generalisability of the findings. However, subsequent analyses with additional games and their reviews helped mitigate this shortcoming. Secondly, the hyperparameter tuning of this study is limited, particularly for building the BERTopic model. For example, customising the appropriate value for 'min topic size' in BERTopic could improve the model's performance. However, such parameter tuning in BERTopic is rarely found in the existing literature, an exception being Sharifian-Attar et al. [84]. For further research, adding additional evaluation metrics, such as topic diversity, would attain better reliability. Thirdly, the current study did not provide an in-depth discussion about how players' opinions might evolve over time. This is an important aspect of review analysis and further work is needed to provide more insights into player feedback in future. Finally, players' attitudes about the different topics were not addressed in this study, although the developers recommended this. The scope of this study was also limited in theme identification, preventing an in-depth investigation of players' sentiments. This would be an important area to examine in future research.

## VI. Conclusion

In conclusion, this study further builds on previous research on the topic classification of game reviews by evaluating the three different topic models. The study improved the automatic discovery of topics from game reviews by using unsupervised machine-learning methods with two BoW and TF-IDF features. Meanwhile, we used BERTopic to provide a novel way of presenting and interpreting the topics extracted from online players' reviews within the game development context. This paper suggests a rationale for using an advanced embedding model (BERTopic) to understand the output of topic classifications in video game research.

This study also contributes to video game industry practice by demonstrating that TM techniques can be valuable in the development of video games. It provides a novel insight for game developers to collect suggestions for improvement from real players based on their games' reviews. It could serve as the starting point for more longitudinal studies on topic extraction and may inspire new research directions for investigating NLP techniques applied to the video game industry. We believe the conclusions of this paper suggest a good starting point for information extraction and further game user research in the video game domain. Findings from this study need to be validated with more types of games or gaming platforms for future research.

## References

[1] G. Panagiotopoulos, G. Giannakopoulos, A. Liapis, A Study on Video Game Review Summarization, in: Proceedings of the Workshop MultiLing 2019: Summarization Across Languages, Genres and Sources Associated with RANLP 2019, Incoma Ltd., Shoumen, Bulgaria, 2019, pp. 36–43. doi:10.26615/978-954-452-058-8_006.

[2] D. Lin, C.-P. Bezemer, Y. Zou, A. E. Hassan, An empirical study of game reviews on the Steam platform, Empirical Software Engineering 24 (1) (2019) 170–207. doi:10.1007/s10664-018-9627-4.

[3] C. Lewis, J. Whitehead, N. Wardrip-Fruin, What went wrong: A taxonomy of video game bugs, in: Proceedings of the Fifth International Conference on the Foundations of Digital Games, FDG '10, Association for Computing Machinery, New York, NY, USA, 2010, pp. 108–115. doi:10.1145/1822348.1822363.

[4] L. F. S. Britto, L. D. S. Pacífico, Evaluating Video Game Acceptance in Game Reviews using Sentiment Analysis Techniques, in: SBGames 2020, Virtual, 2020, pp. 399–402.

[5] P. Melville, W. Gryc, R. D. Lawrence, Sentiment analysis of blogs by combining lexical knowledge with text classification, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09, Association for Computing Machinery, New York, NY, USA, 2009, pp. 1275–1284. doi:10.1145/1557019.1557156.

[6] X. Wang, D. H.-L. Gohb, Components of game experience: An automatic text analysis of online reviews, Entertainment Computing 33 (2019). doi:10.1016/j.entcom.2019.100338.

[7] A. Asuncion, M. Welling, P. Smyth, Y. W. Teh, On smoothing and inference for topic models, in: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI '09, AUAI Press, Arlington, Virginia, USA, 2009, pp. 27–34.

[8] J. Mcauliffe, D. Blei, Supervised Topic Models, in: Advances in Neural Information Processing Systems, Vol. 20, Curran Associates, Inc., 2007.

[9] R. S. T. Lee, Natural Language Processing: A Textbook with Python Implementation, Springer.
URL https://link.springer.com/book/10.1007/978-981-99-1999-4

[10] J. P. Zagal, A. Ladd, T. Johnson, Characterizing and understanding game reviews, in: Proceedings of the 4th International Conference on Foundations of Digital Games, FDG '09, Association for Computing Machinery, New York, NY, USA, 2009, pp. 215–222. doi:10.1145/1536513.1536553.

[11] J. P. Zagal, N. Tomuro, A. Shepitsen, Natural Language Processing in Game Studies Research: An Overview, Simulation & Gaming 43 (3) (2012) 356–373. doi:10.1177/1046878111422560.

[12] J. P. Zagal, N. Tomuro, Cultural differences in game appreciation: A study of player game reviews, in: Proceedings of the 8th International Conference on the Foundations of Digital Games, Chania, Greece, 2013.

[13] J. Pelletier-Gagnon, 'Very much like any other Japanese RPG you've ever played': Using undirected topic modelling to examine the evolution of JRPGs' presence in anglophone web publications, Journal of Gaming & Virtual Worlds 10 (2) (2018) 135–148. doi:10.1386/jgvw.10.2.135_1.

[14] I. Busurkina, V. Karpenko, E. Tulubenskaya, D. Bulygin, Game Experience Evaluation. A Study of Game Reviews on the Steam Platform, in: D. A. Alexandrov, A. V. Boukhanovsky, A. V. Chugunov, Y. Kabanov, O. Koltsova, I. Musabirov (Eds.), Digital Transformation and Global Society, Communications in Computer and Information Science, Springer International Publishing, Cham, 2020, pp. 117–127. doi:10.1007/978-3-030-65218-0_9.

[15] Y. Yu, B.-H. Nguyen, F. Yu, V.-N. Huynh, Discovering Topics of Interest on Steam Community Using an LDA Approach, in: C. Leitner, W. Ganz, D. Satterfield, C. Bassano (Eds.), Advances in the Human Side of Service Engineering, Lecture Notes in Networks and Systems, Springer International Publishing, Cham, 2021, pp. 510–517. doi:10.1007/978-3-030-80840-2_59.

[16] X. Li, Z. Zhang, K. Stefanidis, A Data-Driven Approach for Video Game Playability Analysis Based on Players' Reviews, Information 12 (3) (2021) 129. doi:10.3390/info12030129.

[17] M. Kwak, J. S. Park, J. G. Shon, Identifying Critical Topics for Successful Games in Game Reviews by Applying Latent Dirichlet Allocation, in: J. J. Park, V. Loia, Y. Pan, Y. Sung (Eds.), Advanced Multimedia and Ubiquitous Engineering, Lecture Notes in Electrical

Engineering, Springer, Singapore, 2021, pp. 41–48. doi:10.1007/978-981-15-9309-3_6.

[18] D. Youm, J. Kim, Text Mining Approach to Improve Mobile Role Playing Games Using Users' Reviews, Applied Sciences 12 (12) (2022) 6243. doi:10.3390/app12126243.

[19] R. Bunga, F. Batista, R. Ribeiro, From implicit preferences to ratings: Video games recommendation based on collaborative filtering, From implicit preferences to ratings: Video games recommendation based on collaborative filtering (2021) 209–216doi:10.5220/0010655900003064.

[20] L. Zhou, S. Pan, J. Wang, A. V. Vasilakos, Machine learning on big data: Opportunities and challenges, Neurocomputing 237 (2017) 350–361. doi:10.1016/j.neucom.2017.01.026.

[21] M. J. Sánchez-Franco, M. Rey-Moreno, Do travelers' reviews depend on the destination? An analysis in coastal and urban peer-to-peer lodgings, Psychology & Marketing 39 (2) (2022) 441–459. doi:10.1002/mar.21608.

[22] M. Grootendorst, BERTopic: Neural topic modeling with a class-based TF-IDF procedure (Mar. 2022). arXiv:2203.05794, doi:10.48550/arXiv.2203.05794.

[23] D. Angelov, Top2Vec: Distributed Representations of Topics (Aug. 2020). arXiv:2008.09470, doi:10.48550/arXiv.2008.09470.

[24] A. Abuzayed, H. Al-Khalifa, BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique, Procedia Computer Science 189 (2021) 191–194. doi:10.1016/j.procs.2021.05.096.

[25] F. Alhaj, A. Al-Haj, A. Sharieh, R. Jabri, Improving Arabic cognitive distortion classification in Twitter using BERTopic, International Journal of Advanced Computer Science and Applications 13 (1) (2022) 854–860.

[26] M. Chong, H. Chen, Racist Framing through Stigmatized Naming: A Topical and Geo-locational Analysis of #Chinavirus and #Chinesevirus on Twitter, Proceedings of the Association for Information Science and Technology 58 (1) (2021) 70–79. doi:10.1002/pra2.437.

[27] A. Baird, Y. Xia, Y. Cheng, Consumer perceptions of telehealth for mental health or substance abuse: A Twitter-based topic modeling analysis, JAMIA Open 5 (2) (2022) ooac028. doi:10.1093/jamiaopen/ooac028.

[28] C. Gao, B. Wang, P. He, J. Zhu, Y. Zhou, M. R. Lyu, PAID: Prioritizing app issues for developers by tracking user reviews over versions, in: 2015 IEEE 26th International Symposium on Software Reliability Engineering (ISSRE), 2015, pp. 35–45. doi:10.1109/ISSRE.2015.7381797.

[29] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, H. C. Gall, How can i improve my app? Classifying user reviews for software maintenance and evolution, in: 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME), 2015, pp. 281–290. doi:10.1109/ICSM.2015.7332474.

[30] Y. Liu, L. Liu, H. Liu, X. Wang, Analyzing reviews guided by App descriptions for the software development and evolution, Journal of Software: Evolution and Process 30 (12) (2018) e2112. doi:10.1002/smr.2112.

[31] Y. Tan, J. Chen, W. Shang, T. Zhang, S. Fang, X. Luo, Z. Chen, S. Qi, STRE: An Automated Approach to Suggesting App Developers When to Stop Reading Reviews, IEEE Transactions on Software Engineering (2023) 1–18doi:10.1109/TSE.2023.3285743.

[32] S. Gao, L. Liu, Y. Liu, H. Liu, Y. Wang, Updating the goal model with user reviews for the evolution of an app, Journal of Software: Evolution and Process 32 (8) (2020) e2257. doi:10.1002/smr.2257.

[33] A. Kosmopoulos, A. Liapis, G. Giannakopoulos, N. Pittaras, Summarizing Game Reviews: First Contact, in: SETN Workshops, Athens, Greece, 2020, p. 10.

[34] C. Ruggles, G. Wadley, M. R. Gibbs, Online Community Building Techniques Used by Video Game Developers, in: F. Kishino, Y. Kitamura, H. Kato, N. Nagata (Eds.), Entertainment Computing - ICEC 2005, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2005, pp. 114–125. doi:10.1007/11558651_12.

[35] F. Parker, M. E. Perks, Streaming ambivalence: Livestreaming and indie game development, Convergence 27 (6) (2021) 1735–1752. doi:10.1177/13548565211027809.

[36] G. K. Cheung, T. Zimmermann, N. Nagappan, The first hour experience: How the initial play can engage (or lose) new players, in: Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play, CHI PLAY '14, Association for Computing Machinery, New York, NY, USA, 2014, pp. 57–66. doi:10.1145/2658537.2658540.

[37] G. McAllister, G. R. White, Video Game Development and User Experience, in: R. Bernhaupt (Ed.), Game User Experience Evaluation, Human–Computer Interaction Series, Springer International Publishing, Cham, 2015, pp. 11–35. doi:10.1007/978-3-319-15985-0_2.

[38] G. McAllister, P. Mirza-Babaei, J. Avent, Improving Gameplay with Game Metrics and Player Metrics, in: M. Seif El-Nasr, A. Drachen, A. Canossa (Eds.), Game Analytics: Maximizing the Value of Player Data, Springer, London, 2013, pp. 621–638. doi:10.1007/978-1-4471-4769-5_27.

[39] H. Ulf, Where Do Game Design Ideas Come From? Invention and Recycling in Games Developed in Sweden (2009).

[40] A. Kamienski, C.-P. Bezemer, An empirical study of Q&A websites for game developers, Empirical Software Engineering 26 (6) (2021) 115. doi:10.1007/s10664-021-10014-4.

[41] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791. doi:10.1038/44565.

[42] S. Athukorala, W. Mohotti, An effective short-text topic modelling with neighbourhood assistance-driven NMF in Twitter, Social Network Analysis and Mining 12 (1) (2022) 89. doi:10.1007/s13278-022-00898-5.

[43] T. Virtanen, A. Taylan Cemgil, S. Godsill, Bayesian extensions to non-negative matrix factorisation for audio signal modelling, in: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, pp. 1825–1828. doi:10.1109/ICASSP.2008.4517987.

[44] A. Pascual-Montano, J. Carazo, K. Kochi, D. Lehmann, R. Pascual-Marqui, Nonsmooth nonnegative matrix factorization (nsNMF), IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (3) (2006) 403–415. doi:10.1109/TPAMI.2006.60.

[45] J. Buciu, Non-negative Matrix Factorization, A New Tool for Feature Extraction: Theory and Applications., International Journal of Computers, Communications & Control 3 (3) (2008) 67–74.

[46] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, The Journal of Machine Learning Research 3 (null) (2003) 993–1022.

[47] A. Passos, H. M. Wallach, A. McCallum, Correlations and Anticorrelations in LDA Inference, in: Proceedings of the 2011 Workshop on Challenges in Learning Hierarchical Models: Transfer Learning and Optimization, Granada, 2011, pp. 1–5.

[48] A. Xu, T. Qi, X. Dong, Analysis of the Douban online review of the MCU: Based on LDA topic model, Journal of Physics: Conference Series 1437 (1) (2020) 012102. doi:10.1088/1742-6596/1437/1/012102.

[49] O. Oyebode, F. Alqahtani, R. Orji, Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews, IEEE Access 8 (2020) 111141–111158. doi:10.1109/ACCESS.2020.3002176.

[50] C. Iacob, R. Harrison, Retrieving and analyzing mobile apps feature requests from online reviews, in: 2013 10th Working Conference on Mining Software Repositories (MSR), 2013, pp. 41–44. doi:10.1109/MSR.2013.6624001.

[51] A. Faisal, M. Peltoniemi, Establishing Video Game Genres Using Data-Driven Modeling and Product Databases, Games and Culture 13 (1) (2018) 20–43. doi:10.1177/1555412015601541.

[52] G. Hello, No Man's Sky on Steam, https://store.steampowered.com/app/275850/No_Mans_Sky/.

[53] R. Egger, J. Yu, A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts, Frontiers in Sociology 7 (2022) 886498. doi:10.3389/fsoc.2022.886498.

[54] Lupton, The thirteen Ps of big data (2015).

[55] M. E. Roberts, B. M. Stewart, D. Tingley, Stm: An R Package for Structural Topic Models, Journal of Statistical Software 91 (2019) 1–40. doi:10.18637/jss.v091.i02.

[56] C. Lu, X. Li, T. Nummenmaa, Z. Zhang, J. Peltonen, Patches and Player Community Perceptions: Analysis of No Man's Sky Steam Reviews, in: DiGRA '20, Tampere, Finland, 2020, p. 24.

[57] I. Ali, M. A. Naeem, Identifying and Profiling User Interest over time using Social Data, in: 2022 24th International Multitopic Conference (INMIC), 2022, pp. 1–6. doi:10.1109/INMIC56986.2022.9972955.

[58] T. Verbeij, I. Beyens, D. Trilling, P. M. Valkenburg, Happiness and Sadness in Adolescents' Instagram Direct Messaging: A Neural Topic Modelling Approach (Dec. 2022). doi:10.31234/osf.io/5pgdb.

[59] U. Bayram, Revealing the Reflections of the Pandemic by Investigating COVID-19 Related News Articles Using Machine Learning and Network Analysis, Bilişim Teknolojileri Dergisi 15 (2) (2022) 209–220. doi:10.17671/gazibtd.949599.

[60] Y. Jin, Travel Guide Using Text Mining and BERTopic, Ph.D. thesis, University of California, Los Angeles (2022).

[61] S. Vasudeva Raju, B. Kumar Bolla, D. K. Nayak, J. Kh, Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings, in: 2022 IEEE 7th International Conference for Convergence in Technology (I2CT), 2022, pp. 1–6. doi:10.1109/I2CT54291.2022.9824873.

[62] C. Meaney, M. Escobar, T. A. Stukel, P. C. Austin, L. Jaakkimainen, Comparison of Methods for Estimating Temporal Topic Models From

Primary Care Clinical Text Data: Retrospective Closed Cohort Study, JMIR Medical Informatics 10 (12) (2022) e40102. doi:10.2196/40102.
[63] P. Tueschen, Customer Review Analysis, Master's thesis (Oct. 2022).
[64] A. Krishnan, Exploring the Power of Topic Modeling Techniques in Analyzing Customer Reviews: A Comparative Analysis (Aug. 2023). arXiv:2308.11520, doi:10.48550/arXiv.2308.11520.
[65] D. Tsarev, M. Petrovskiy, I. Mashechkin, Using NMF-based text summarization to improve supervised and unsupervised classification, in: 2011 11th International Conference on Hybrid Intelligent Systems (HIS), 2011, pp. 185–189. doi:10.1109/HIS.2011.6122102.
[66] A. Anantharaman, A. Jadiya, C. T. S. Siri, B. N. Adikar, B. Mohan, Performance Evaluation of Topic Modeling Algorithms for Text Classification, in: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 704–708. doi:10.1109/ICOEI.2019.8862599.
[67] D. Mensouri, A. Azmani, M. Azmani, Combining Roberta Pre-Trained Language Model and NMF Topic Modeling Technique to Learn from Customer Reviews Analysis, International Journal of Intelligent Systems and Applications in Engineering 11 (1) (2023) 39–49.
[68] T. Rajasundari, P. Subathra, P. Kumar, Performance analysis of topic modeling algorithms for news articles, Journal of Advanced Research in Dynamical and Control Systems 2017 (Special Issue 11) (2017) 175–183.
[69] X. Li, C. Lu, J. Peltonen, Z. Zhang, A statistical analysis of Steam user profiles towards personalized gamification, in: The 3nd International GamiFIN Conferenc, Levi, Finland, 2019, pp. 217–228.
[70] R. Albalawi, T. H. Yeap, M. Benyoucef, Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis, Frontiers in Artificial Intelligence 3 (2020).
[71] D. Santos, N. Zagalo, C. Morais, Players Perception of the Chemistry in the Video Game No Man's Sky, Simulation & Gaming 54 (3) (2023) 375–394. doi:10.1177/10468781231169301.
[72] A. Reinhard, Archeology of Abandoned Human Settlements in No Man's Sky: A New Approach to Recording and Preserving User-Generated Content in Digital Games, Games and Culture 16 (7) (2021) 855–884. doi:10.1177/15554120211005236.
[73] E. R. Tait, I. L. Nelson, Nonscalability and generating digital outer space natures in No Man's Sky, Environment and Planning E: Nature and Space 5 (2) (2022) 694–718. doi:10.1177/25148486211000746.
[74] P. Siuda, D. Reguła, J. Majewski, A. Kwapiszewska, Broken Promises Marketing. Relations, Communication Strategies, and Ethics of Video Game Journalists and Developers: The Case of Cyberpunk 2077, Games and Culture (2023) 15554120231173479doi:10.1177/15554120231173479.
[75] C. Fiesler, N. Beard, B. C. Keegan, No Robots, Spiders, or Scrapers: Legal and Ethical Regulation of Data Collection Methods in Social Media Terms of Service, Proceedings of the International AAAI Conference on Web and Social Media 14 (2020) 187–196. doi:10.1609/icwsm.v14i1.7290.
[76] K. Stepien, Topic Modelling and Data Analysis: Impact of using topic modelling on game review, Mres, University of Lincoln (Apr. 2021).
[77] F. Baumann, D. Emmert, H. Baumgartl, R. Buettner, Hardcore Gamer Profiling: Results from an unsupervised learning approach to playing behavior on the Steam platform, Procedia Computer Science 126 (2018) 1289–1297. doi:10.1016/j.procs.2018.08.078.
[78] G. Al-Talib, H. S. Hassan, A Study on Analysis of SMS Classification Using TF-IDF Weighting, International Journal of Computer Networks and Communications Security 1 (5) (2013) 189–194.
[79] H. Baker, M. R. Hallowell, A. J. P. Tixier, Automatically learning construction injury precursors from text, Automation in Construction 118 (2020) 103145. doi:10.1016/j.autcon.2020.103145.
[80] D. S. Sisodia, S. Bhandari, N. K. Reddy, A. Pujahari, A Comparative Performance Study of Machine Learning Algorithms for Sentiment Analysis of Movie Viewers Using Open Reviews, in: M. Pant, T. K. Sharma, S. Basterrech, C. Banerjee (Eds.), Performance Management of Integrated Systems and Its Applications in Software Engineering, Asset Analytics, Springer, Singapore, 2020, pp. 107–117. doi:10.1007/978-981-13-8253-6_10.
[81] D. Lee, H. S. Seung, Algorithms for Non-negative Matrix Factorization, in: Advances in Neural Information Processing Systems, Vol. 13, MIT Press, 2000.
[82] H. M. Wallach, I. Murray, R. Salakhutdinov, D. Mimno, Evaluation methods for topic models, in: Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09, Association for Computing Machinery, New York, NY, USA, 2009, pp. 1105–1112. doi:10.1145/1553374.1553515.
[83] I. Korshunova, H. Xiong, M. Fedoryszak, L. Theis, Discriminative Topic Modeling with Logistic LDA, in: Advances in Neural Information Processing Systems, Vol. 32, Curran Associates, Inc., 2019.
[84] V. Sharifian-Attar, S. De, S. Jabbari, J. Li, H. Moss, J. Johnson, Analysing Longitudinal Social Science Questionnaires: Topic modelling with BERT-based Embeddings, in: 2022 IEEE International Conference on Big Data (Big Data), 2022, pp. 5558–5567. doi:10.1109/BigData55660.2022.10020678.
[85] M. Röder, A. Both, A. Hinneburg, Exploring the Space of Topic Coherence Measures, in: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM '15, Association for Computing Machinery, New York, NY, USA, 2015, pp. 399–408. doi:10.1145/2684822.2685324.
[86] D. Newman, J. H. Lau, K. Grieser, T. Baldwin, Automatic evaluation of topic coherence, in: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10, Association for Computational Linguistics, USA, 2010, pp. 100–108.
[87] E. Rijcken, F. Scheepers, P. Mosteiro, K. Zervanou, M. Spruit, U. Kaymak, A Comparative Study of Fuzzy Topic Models and LDA in terms of Interpretability, in: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), 2021, pp. 1–8. doi:10.1109/SSCI50451.2021.9660139.
[88] S. Bellaouar, M. M. Bellaouar, I. E. GHADA, Topic Modeling: Comparison of LSA and LDA on Scientific Publications, in: 2021 4th International Conference on Data Storage and Data Engineering, DSDE '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 59–64. doi:10.1145/3456146.3456156.
[89] S. K. Ray, A. Ahmad, C. A. Kumar, Review and Implementation of Topic Modeling in Hindi, Applied Artificial Intelligence 33 (11) (2019) 979–1007. doi:10.1080/08839514.2019.1661576.
[90] A. Hoyle, P. Goel, A. Hian-Cheong, D. Peskov, J. Boyd-Graber, P. Resnik, Is Automated Topic Model Evaluation Broken? The Incoherence of Coherence, in: Advances in Neural Information Processing Systems, Vol. 34, Curran Associates, Inc., 2021, pp. 2018–2033.
[91] B. Dahal, S. A. P. Kumar, Z. Li, Topic modeling and sentiment analysis of global climate change tweets, Social Network Analysis and Mining 9 (1) (2019) 24. doi:10.1007/s13278-019-0568-8.
[92] Y. Meng, Y. Zhang, J. Huang, Y. Zhang, J. Han, Topic Discovery via Latent Space Clustering of Pretrained Language Model Representations, in: Proceedings of the ACM Web Conference 2022, WWW '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 3143–3152. doi:10.1145/3485447.3512034.
[93] Y. Zuo, J. Wu, H. Zhang, H. Lin, F. Wang, K. Xu, H. Xiong, Topic Modeling of Short Texts: A Pseudo-Document View, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 2105–2114. doi:10.1145/2939672.2939880.
[94] N. Aletras, M. Stevenson, Evaluating Topic Coherence Using Distributional Semantics, in: Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers, Association for Computational Linguistics, Potsdam, Germany, 2013, pp. 13–22.
[95] SME definition, https://single-market-economy.ec.europa.eu/smes/sme-definition_en.
[96] J. M. Corbin, A. Strauss, Grounded theory research: Procedures, canons, and evaluative criteria, Qualitative Sociology 13 (1) (1990) 3–21. doi:10.1007/BF00988593.
[97] K. M. Eisenhardt, Building Theories from Case Study Research, The Academy of Management Review 14 (4) (1989) 532–550. arXiv:258557, doi:10.2307/258557.
[98] J. Corbin, A. Strauss, Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, SAGE Publications, 2014.
[99] O. Gencoglu, Deep Representation Learning for Clustering of Health Tweets (Dec. 2018). arXiv:1901.00439, doi:10.48550/arXiv.1901.00439.