

LDA와 BERTopic을 활용한 《검은신화: 오공》 영어 Steam 리뷰의 시기별 토픽 및 감성 변화 분석

임 준 형¹ · 배 현 영^{2*}

¹한국폴리텍II대학 인천캠퍼스 반도체공정과 전문학사과정 재학생

²한국기술교육대학교 AI융합교육학과 겸임교수

Temporal Changes in Topics and Sentiment in English-Language Steam Reviews of “Black Myth: Wukong” Using Latent Dirichlet Allocation and BERTopic

June-Hyeong Lim¹ · Hyun-Young Bae^{2*}

¹Student, Department of Semiconductor Process, Korea Polytechnics Incheon Campus, Incheon 21417, Korea

²Adjunct Professor, Department of AI Convergence Education, Korea University of Technology and Education, Cheonan 31253, Korea

[요 약]

국제적 관심과 중국 신화적 배경을 고려해 《검은신화: 오공》을 연구 대상으로 선정하고, 출시 이후 영어 Steam 리뷰에 나타난 이용자 감성과 담론의 시간적 변화를 분석하였다. 2024년 8월 20일부터 2026년 6월 17일까지 수집된 플레이타임 1시간 이상 리뷰 66,981건을 출시 직후, 중기 및 후기로 구분하고 TF-IDF, VADER, LDA 및 BERTopic을 적용하였다. 분석 결과, VADER 분류상 긍정 범주가 전 기간 우세했고, 이용자 담론은 초기의 원숭이 캐릭터·밈과 올해의 게임 평가·시상 화제에서 후기의 보스 전투, 소울라이크 장르 및 게임 품질 평가 중심으로 변화하였다. 중국 신화와 문화적 소재는 두 토픽모형에서 독립 주제로 지속되었다. LDA는 포괄적 담론구조를, BERTopic은 세분화된 문맥 기반 토픽을 제시했으며, 두 모형은 주요 담론과 시기별 변화에서 유사한 경향을 보였다. 본 연구는 장기 게임 리뷰 담론의 변화를 분석하고, 두 토픽모형을 통해 포괄적 구조와 세부 주제를 상호보완적으로 제시했다는 데 의의가 있다.

[Abstract]

Given the international attention surrounding “Black Myth: Wukong” and its Chinese mythological roots, this study examined temporal changes in user sentiment and discourse in English-language Steam reviews. We analyzed 66,981 reviews posted between August 20, 2024, and June 17, 2026, by users with at least one hour of playtime. The reviews were divided into launch, middle, and late periods. We applied term frequency-inverse document frequency (TF-IDF), Valence Aware Dictionary and sEntiment Reasoner (VADER), latent Dirichlet allocation (LDA), and the transformer-based BERTopic model. VADER classification showed that positive sentiment predominated across all periods. User discourse shifted from monkey characters, memes, and game awards to boss combat, Soulslike features, and game quality. Chinese mythology and cultural elements remained distinct topics in both models. LDA captured broad discourse structures, whereas BERTopic identified fine-grained contextual topics; both revealed similar themes and temporal patterns. Together, the models provide complementary views of longitudinal game-review discourse.

색인어 : 검은신화: 오공, Steam 리뷰, 텍스트마이닝, 토픽모델링, 감성분석

Keyword : Black Myth: Wukong, Steam Reviews, Text Mining, Topic Modeling, Sentiment Analysis

<http://dx.doi.org/10.9728/dcs.2026.27.8.2183>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 14 July 2026; **Revised** 31 July 2026

Accepted 06 August 2026

***Corresponding Author; Hyun-Young Bae**

E-mail: youg989@daum.net

I. 서론

게임은 오락을 넘어 서사, 역사, 신화 및 문화적 이미지를 전달하는 디지털 문화콘텐츠로 기능한다. 이용자는 게임을 플레이하며 경험한 전투, 난이도, 그래픽, 성능 및 문화적 소재에 대한 평가와 감정을 온라인 리뷰를 통해 표현한다. 특히 작성 시점이 기록된 플랫폼 리뷰는 게임 출시 이후 이용자의 관심과 평가 요소가 시간의 흐름에 따라 어떻게 변화하는지를 분석할 수 있는 유용한 자료이다.

2024년 8월 20일 출시된 《검은신화: 오공》은 중국 고전 소설 『서유기』와 중국 신화를 주요 소재로 활용한 AAA급 액션 RPG로, 출시 직후 전 플랫폼 누적 판매량 1,000만 장을 돌파하며 국제 게임시장에서 높은 관심을 받았다[1]. 이 작품은 중국 신화와 전통적 세계관을 서사와 캐릭터에 구현하는 동시에 보스 전투, 난이도, 그래픽, 성능 및 소울라이크 장르의 특성을 포함하고 있다. 따라서 이용자 리뷰에는 게임플레이와 품질에 대한 평가뿐 아니라 중국 신화와 문화적 소재에 관한 담론도 함께 나타날 수 있다.

그러나 《검은신화: 오공》을 직접 다룬 학술 연구는 아직 제한적이다. Tang 등[2]은 관련 YouTube 댓글 7,604건을 대상으로 감성분석과 LDA 토픽모델링을 수행하였으나, 실제 플레이 경험이 반영된 Steam 리뷰와 장기간의 담론 변화는 충분히 다루지 못하였다. 기존 게임 리뷰 연구에서는 Steam 리뷰의 도움성[3], 이용자의 주요 관심 요소[4], 토픽과 감성 정보를 활용한 리뷰 분류[5], 플레이타임에 따른 이용자 경험의 차이[6] 등이 분석되었다. 최근 Singh와 O'Hagan[7]은 소셜 VR 게임인 Rec Room의 Steam 리뷰 약 40,000건에 VADER와 BERTopic을 적용하여 감성과 주요 담론의 장기적 변화를 분석하였다. 그러나 이 연구는 소셜 VR 커뮤니티에서 나타나는 신흥 문제와 쟁점의 모니터링에 초점을 두었으며, 문화적 소재를 핵심으로 하는 게임에서 해당 소재가 이용자 담론 속에서 장기간 지속되는 양상이나 LDA와 BERTopic이 제시하는 분석 수준의 차이는 다루지 않았다.

따라서 《검은신화: 오공》 출시 이후 장기간 축적된 Steam 리뷰를 대상으로 키워드, 감성 및 토픽 구조의 시기별 변화를 종합적으로 분석할 필요가 있다. 특히 중국 신화와 문화적 소재가 출시 초기의 일시적 관심에 그치는지, 이후의 이용자 담론에서도 지속되는지를 살펴볼 필요가 있다. 아울러 확률 기반 LDA와 문맥 임베딩 기반 BERTopic을 병행하여, 하나의 게임 리뷰 코퍼스에서 포괄적인 담론 구조와 세분화된 문맥 기반 주제가 어떠한 공통점과 차이를 보이는지 비교할 필요가 있다.

이에 본 연구는 2024년 8월 20일부터 2026년 6월 17일 까지 Steam에 게시된 영어 리뷰 중 플레이타임이 1시간 이상인 66,981건을 대상으로 출시 직후, 중기 및 후기의 키워드, 감성 및 토픽 변화를 분석하였다. TF-IDF 기반 키워드 분석과 VADER 감성분석을 수행하고, LDA와 BERTopic을 병행 적용하여 두 모형에서 나타나는 주요 담론과 시기별 변

화 방향의 공통점과 차이를 살펴보았다.

본 연구의 연구문제는 다음과 같다.

첫째, 《검은신화: 오공》 영어 Steam 리뷰의 주요 키워드와 감성 분포는 출시 이후 시기별로 어떻게 변화하는가?

둘째, 영어 Steam 리뷰에서 나타나는 주요 토픽은 무엇이며, 게임플레이와 중국 신화·문화 관련 토픽의 비중은 시기별로 어떻게 변화하는가?

셋째, LDA와 BERTopic에서 나타나는 주요 담론의 구조와 변화 방향에는 어떠한 공통점과 차이가 있는가?

본 연구는 약 22개월간 축적된 대규모 영어 Steam 리뷰를 분석하여, 이용자 담론이 출시 초기의 캐릭터·미프와 시장 관련 화제에서 후기의 보스 전투, 소울라이크 장르 및 게임 품질 평가 중심으로 변화하였음을 확인하였다. 또한 중국 신화와 문화적 소재가 장기간 독립적인 담론 주제로 유지되었음을 제시하였다. 방법론적으로는 LDA가 포괄적인 담론 구조를 제시하고 BERTopic이 이를 세분화된 문맥 기반 토픽으로 구체화함을 보여줌으로써, 두 모형의 상호보완적 활용 가능성을 확인하였다. 이러한 결과는 문화 기반 게임 연구에서 문화적 소재의 담론적 지속성을 분석하는 근거를 제공하며, 게임 개발·운영 측면에서는 출시 이후 변화하는 이용자 관심과 품질 평가를 장기적으로 관찰할 필요성을 시사한다.

II. 이론적 배경 및 선행연구

2-1 게임 콘텐츠의 문화적 재현과 이용자 담론

게임은 서사, 캐릭터, 공간 및 시각적 표현을 통해 역사, 신화, 가치관 및 문화적 이미지를 재현하는 디지털 문화콘텐츠이다. Hall[8]의 인코딩·디코딩 관점에 따르면, 미디어 콘텐츠의 의미는 수용자에게 일방적으로 전달되는 것이 아니라 개인의 사회·문화적 배경과 경험에 따라 다르게 해석된다. 따라서 게임에 구현된 신화와 문화적 소재 역시 제작자의 의도와 동일한 방식으로만 수용되는 것이 아니라, 이용자의 플레이 경험과 기존 지식에 따라 다양하게 이해되고 평가될 수 있다.

Schröder[9]는 미디어 수용을 단순한 긍정과 부정의 반응이 아니라 의미의 이해와 평가가 결합된 다차원적 과정으로 보았다. 이러한 관점에서 온라인 리뷰는 이용자의 수용 경험 전체를 직접 보여주는 자료라기보다, 콘텐츠 경험 이후 형성된 해석과 평가가 선택적으로 표출된 담론 자료로 볼 수 있다. 따라서 게임 콘텐츠를 분석할 때에는 감성의 방향뿐 아니라 이용자가 어떠한 요소를 중심으로 작품을 해석하고 평가하는지를 함께 살펴볼 필요가 있다.

Shaw[10] 역시 비디오게임 문화를 고정된 실체가 아니라 사회적·문화적 담론을 통해 구성되는 대상으로 보았다. 이는 게임의 문화적 의미를 이해하기 위해 작품 자체의 표현뿐 아니라 이용자가 이를 설명하고 평가하는 담론도 함께 분석해야 함을 시사한다. 종합하면, 게임에 구현된 문화적 소재의 의

미는 작품 내부에 고정되어 있기보다 이용자의 경험과 해석이 표출되는 담론을 통해 구성된다.

이에 본 연구는 영어 Steam 리뷰를 이용자의 중국 문화에 대한 태도나 국가 이미지의 변화를 직접 측정하는 자료로 간주하지 않는다. 대신 중국 신화와 문화적 소재가 어떠한 키워드와 토픽으로 나타나며, 그 비중이 시간의 흐름에 따라 어떻게 변화하는지를 분석하여 해당 소재의 담론적 지속성을 살펴보고자 한다.

2-2 Steam 리뷰 데이터와 장기적 게임 리뷰 연구

게임 이용자 리뷰는 플레이 경험에 대한 평가와 감정이 텍스트 형태로 축적된 비정형 데이터로, 이용자 반응과 주요 담론을 분석하는 데 활용된다. 특히 Steam 리뷰는 리뷰 원문과 함께 작성 시점, 플레이타임 및 추천 여부 등의 정보를 제공하므로, 이용자의 플레이 경험과 평가를 연계하여 분석하는 데 유용하다[3],[6]. 또한 작성 시점을 기준으로 리뷰를 배열할 수 있어, 게임 출시 이후 관심과 평가요소가 시간의 흐름에 따라 어떻게 변화하는지를 장기적으로 관찰할 수 있다.

다만 Steam 리뷰는 자발적으로 리뷰를 작성한 이용자의 의견으로 구성되므로 전체 게임 이용자를 대표하는 확률표본으로 보기 어렵다. 또한 짧은 감상, 반복적인 밈, 반역적 표현이 포함될 수 있으며, 업데이트, 시상식 또는 논란을 계기로 리뷰 작성이 특정 시기에 집중될 가능성도 있다. 따라서 분석 결과는 전체 이용자의 보편적 인식으로 일반화하기보다, 특정 기간 Steam 플랫폼의 리뷰 작성자들 사이에서 형성된 담론으로 한정하여 해석할 필요가 있다.

Steam 리뷰를 활용한 선행연구에서는 리뷰의 정보적 가치와 이용자 경험이 다양한 관점에서 분석되어 왔다. 강하나 등[3]은 Steam 리뷰의 도움성에 영향을 미치는 요인을 분석하였고, 위민영 등[4]은 VR 게임 리뷰에 텍스트마이닝을 적용하여 이용자의 주요 관심 요소를 도출하였다. 김태국과 김길환[5]은 토픽모델링과 감성정보를 결합한 게임 리뷰 분류 방법을 제시하였으며, 최영우와 유승호[6]는 플레이타임에 따라 이용자 경험과 주요 키워드가 달라질 수 있음을 확인하였다.

Singh와 O'Hagan[7]은 소셜 VR 게임인 Rec Room의 Steam 리뷰 약 40,000건에 감성분석과 BERTopic을 적용하여 이용자 커뮤니티에서 나타나는 문제와 담론의 장기적 변화를 분석하였다. 이는 Steam 리뷰가 특정 시점의 반응을 파악하는 자료를 넘어, 시간에 따라 변화하는 평가와 쟁점을 추적하는 장기적 자료로 활용될 수 있음을 보여준다.

그러나 기존 연구는 주로 리뷰의 도움성, 감성분류, 이용자 경험 또는 특정 게임요소의 분석에 초점을 두었다. 장기간 축적된 단일 게임의 리뷰를 대상으로 키워드, 감성 및 토픽 구조의 변화를 함께 분석하고, 서로 다른 토픽모형을 병행하여 결과의 공통점과 차이를 비교한 연구는 상대적으로 제한적이다.

2-3 텍스트마이닝 기반 게임 리뷰 분석: 키워드·토픽·감성 분석

텍스트마이닝은 대규모 비정형 텍스트에서 주요 어휘, 잠재 주제 및 감성적 특성을 정량적으로 파악하는 방법이다. 게임 리뷰에는 이용자의 플레이 경험과 평가가 비형식적인 언어로 축적되므로, 키워드 분석, 토픽모델링 및 감성 분석을 함께 적용하면 이용자 담론의 내용과 정서적 방향을 종합적으로 살펴볼 수 있다.

TF-IDF(term frequency-inverse document frequency)는 특정 문서 또는 문서 집합에서 자주 출현하면서 전체 코퍼스에서는 상대적으로 드물게 나타나는 단어에 높은 가중치를 부여하는 방법이다. 이를 통해 각 시기나 문서 집합을 특징짓는 주요 어휘를 식별할 수 있다.

토픽모델링은 대규모 문서 집합에 내재된 잠재 주제 구조를 비지도 방식으로 도출하는 방법이다. Blei 등[11]이 제안한 LDA(Latent Dirichlet Allocation)는 각 문서를 여러 토픽의 혼합으로, 각 토픽을 단어의 확률분포로 가정하는 생성 확률 모형이다. LDA는 연구자가 토픽 수를 사전에 설정해야 하며, 일반적으로 Coherence, Perplexity 및 토픽의 해석 가능성을 종합적으로 고려하여 적정 토픽 수를 결정한다[12]. 다만 단어의 출현 빈도에 기반한 단어주머니 표현을 사용하므로, 단어의 순서와 문맥적 의미를 직접 반영하기 어렵다는 한계가 있다.

Grootendorst[13]가 제안한 BERTopic은 문장 임베딩, 차원 축소 및 밀도 기반 군집화를 결합한 문맥 기반 토픽모델링 기법이다. Sentence-BERT 계열의 임베딩 모형[14]을 활용하여 문서를 의미 벡터로 변환하고, 문맥적으로 유사한 문서를 군집화한 후 클래스 기반 TF-IDF를 통해 대표 키워드를 도출한다. BERTopic은 문맥적 유사성을 반영하여 세분화된 주제를 도출하는 데 유용하지만, 분석 결과가 임베딩 모형과 차원 축소 및 군집화 매개변수에 따라 달라질 수 있으며, 군집 밀도가 낮은 문서는 이상치로 분류될 수 있다.

LDA와 BERTopic은 문서 표현과 토픽 도출 방식에서 차이를 보인다. 김선옥과 양기덕[15]은 두 모형의 병행 활용이 기존 토픽을 구체화하고 새로운 주제를 발견하는 데 기여할 수 있다고 보았으며, Dehghani와 Zaman[16]도 게임 리뷰에 복수의 토픽모형을 적용하여 이용자 피드백을 다각적으로 해석할 수 있음을 제시하였다. 따라서 두 모형의 결과를 함께 검토하면 LDA가 제시하는 포괄적인 담론 구조와 BERTopic이 도출하는 세분화된 문맥 기반 주제를 상호보완적으로 해석할 수 있다. 다만 두 모형에서 유사한 결과가 나타났다는 사실을 통계적 교차 검증이나 분석 결과의 강건성으로 직접 해석해서는 안 된다.

감성 분석은 텍스트에 나타난 긍정과 부정 등의 정서적 방향과 강도를 정량화하는 방법이다. Hutto와 Gilbert[17]가 제안한 VADER(Valence Aware Dictionary and sEntiment Reasoner)는 소셜미디어 텍스트에 특화된 사전·규칙 기반 도구로, 감성 어휘뿐 아니라 대문자, 문장부호, 정도부사 및 부

정어를 함께 고려하여 compound 점수를 산출한다. 별도의 학습 데이터 없이 비형식적인 영어 텍스트에 적용할 수 있다는 장점이 있으나, 반어, 밈, 은어 및 문맥 의존적 표현을 정확하게 판별하는 데에는 한계가 있다.

Singh와 O'Hagan[7]은 Steam 리뷰에 VADER와 BERTopic을 적용하여 감성과 주요 담론의 장기적 변화를 분석함으로써 두 분석 방법의 결합 가능성을 보여주었다. 그러나 이 연구는 소셜 VR 커뮤니티에서 나타나는 신흥 문제와 쟁점의 모니터링에 초점을 두었으며, 문화적 소재가 이용자 담론에서 장기간 지속되는 양상이나 LDA와 BERTopic이 제시하는 토픽 구조의 공통점과 차이는 다루지 않았다.

이에 본 연구는 [7]의 분석 범위를 확장하여 TF-IDF 기반 시기별 주요 어휘 분석과 VADER 기반 감성 분석을 수행하였다. 또한 확률 기반 LDA와 문맥 임베딩 기반 BERTopic을 병행 적용하여 이용자 담론의 포괄적인 구조와 세분화된 문맥 기반 토픽을 도출하고, 두 모형에서 나타나는 주요 주제와 시기별 변화 방향의 공통점과 차이를 상호보완적으로 비교하였다.

III. 연구 방법

3-1 연구 설계 및 절차

본 연구는 《검은신화: 오공》 출시 이후 약 22개월간 Steam에 게시된 영어 리뷰를 대상으로 이용자 담론과 감성의 시기별 변화 양상을 탐색적으로 분석한 실증 연구이다. 변수 간 인과관계나 특정 이론을 검증하기보다, 리뷰에 나타난 주요 키워드, 감성 및 토픽의 분포와 변화 방향을 기술하고 해석하는 데 목적이 있다. 연구는 데이터 수집, 품질 필터링, 텍스트 전처리, 분석 기간 구분 및 세부 분석의 순서로 수행하였으며, TF-IDF 기반 키워드 분석, VADER 감성분석, LDA 및 BERTopic 토픽모델링을 적용하였다.

이용자 담론의 시기별 변화를 파악하기 위해 전체 분석 기간을 출시 직후, 중기 및 후기로 구분하였다. 출시 직후는 게임 출시와 함께 관심과 리뷰 작성이 집중된 2024년 8월 20일부터 9월 30일까지로 설정하였다. 중기는 초기의 집중적인 리뷰 활동이 감소하고 The Game Awards를 비롯한 주요 시상 시즌이 포함된 2024년 10월 1일부터 2025년 2월 28일까지로 구분하였다. 후기는 시상 시즌 이후 장기적인 플레이 경험과 평가가 축적된 2025년 3월 1일부터 2026년 6월 17일까지로 설정하였다.

이러한 시기 구분은 동일한 기간 단위나 통계적 변화점 분석에 따른 것이 아니라, 출시 초기 반응, 시상 시즌 및 장기 이용 단계라는 게임 수용 과정의 특성을 반영한 조작적 구분이다. 각 시기의 기간과 리뷰 수가 서로 다르므로 원시 빈도의 단순 비교는 지양하고, 감성 및 토픽 구성비, 1,000토큰당

상대 빈도와 월별 변화 추이를 중심으로 해석하였다. 분석 기간별 리뷰 분포는 표 1과 같다.

토픽 분석은 최종 기본 코퍼스 중 각 분석 방법의 문서 선정 기준을 충족한 리뷰를 대상으로 LDA와 BERTopic을 각각 적용하였다. 이후 두 모형에서 도출된 토픽의 구성과 시기별 분포를 비교하여 공통적으로 나타나는 주요 담론과 변화 방향을 살펴보았다. 또한 투입 문서 집합의 차이가 비교 결과에 미치는 영향을 점검하기 위해 두 모형의 선정 기준을 모두 충족한 리뷰를 공통표본으로 구성하고 민감도 분석을 수행하였다. 두 모형의 비교는 성능의 우열이나 통계적 교차검증이 아니라, 서로 다른 분석 원리에서 도출된 결과를 상호보완적으로 해석하기 위한 절차로 한정하였다.

표 1. 분석기간 구분 및 영어 Steam 리뷰 분포

Table 1. Period classification and distribution of English-language steam reviews

Period	Date Range	Period Description	No. of Reviews	%
Launch Period	2024.08.20. -2024.09.30	Concentrated review activity immediately after release	42,860	64.0%
Middle Period	2024.10.01. -2025.02.28	Decline in initial review activity and inclusion of the major award season	13,702	20.5%
Late Period	2025.03.01. -2026.06.17	Accumulation of long-term gameplay experiences and evaluations	10,419	15.6%
Total	2024.08.20. -2026.06.17		66,981	100.0%

Note: June 2026 data include reviews collected through June 17. Percentages may not total 100.0% due to rounding.

3-2 데이터 수집

분석자료는 Steam 플랫폼에 게시된 《검은신화: 오공》의 영어 이용자 리뷰이다. Steam 리뷰 API를 활용하여 게임 출시일인 2024년 8월 20일부터 2026년 6월 17일까지 작성된 리뷰를 수집하였다. 수집 항목은 리뷰 원문, 작성 시점, 추천 여부, 플레이타임 및 Steam 구매 여부 등의 메타데이터로 구성하였다.

수집 결과 총 84,951건의 영어 리뷰를 확보하였다. 이후 동일한 리뷰 ID로 중복 수집된 문서와 빈 텍스트를 제거하고, 플레이타임 기준의 품질 필터링을 적용하여 최종 기본 코퍼스 66,981건을 확정하였다. 수집자료와 최종 코퍼스의 현황은 표 2와 같다.

분석방법별로 요구되는 최소 토큰 수와 전처리 기준이 서로 다르므로, LDA와 BERTopic에는 최종 기본 코퍼스 가운데 각 분석기준을 충족한 리뷰만 투입하였다. 분석방법별 텍스트 정규화, 문서 선정 기준 및 최종 투입 문서 수는 3-3절에 제시하였다.

본 연구는 영어로 작성된 Steam 리뷰만을 대상으로 하므로, 분석결과를 전체 게임 이용자나 특정 국가 이용자의 인식으로 일반화하기 어렵다. 따라서 연구결과는 해당 기간 Steam 플랫폼의 영어 리뷰 작성자들 사이에서 형성된 담론으로 한정하여 해석하였다.

표 2. 영어 Steam 리뷰의 수집 및 최종 코퍼스 현황

Table 2. Collection and final corpus of English-language steam reviews

Data Source	Language	Collection Period	Collected	Final Base Corpus
Steam User Reviews	English	2024.08.20.-2026.06.17.	84,951	66,981

Note: The final base corpus was obtained after removing duplicate and empty reviews and applying playtime-based quality filtering.

3-3 데이터 전처리

데이터 전처리는 품질 필터링과 분석방법별 텍스트 정규화의 두 단계로 수행하였다. 먼저 품질 필터링 단계에서는 동일한 리뷰 ID로 중복 수집된 문서와 빈 텍스트를 제거하였다. 또한 플레이 경험이 초기 단계에 한정된 리뷰가 분석결과에 미치는 영향을 줄이기 위해 플레이타임이 1시간 미만인 리뷰를 제외하였다. 플레이타임에 따라 이용자 경험과 리뷰의 주요 내용이 달라질 수 있다는 선행연구[6]를 참고하였으나, 1시간 기준은 본 연구의 목적에 따라 설정한 조작적 기준이다. 다만 이 기준으로 인해 초기 실행 오류, 성능 문제 및 환불 경험이 반영된 단시간 플레이 리뷰가 제외될 수 있으므로 결과 해석 시 이를 고려하였다. 이러한 기준을 적용한 결과, 최초 수집된 영어 리뷰 84,951건 중 17,970건이 제외되었으며, 최종 기본 코퍼스는 66,981건으로 확정되었다.

텍스트 정규화는 각 분석방법의 특성에 따라 차별적으로 적용하였다. VADER 감성분석에는 대소문자, 문장부호, 부정어 및 강조 표현과 같은 감성 단서를 보존하기 위해 HTML 태그, URL 및 Steam 마크업 등 분석에 불필요한 요소만 제거한 텍스트를 사용하였다. 반면 TF-IDF 키워드 분석과 LDA 토픽모델링에는 영문 소문자화와 특수문자 제거를 수행한 후, spaCy의 en_core_web_sm 모델을 활용하여 표제어를 추출하였다.

이후 NLTK 영어 불용어에 game, play, good 등 전체 코퍼스에서 빈번하게 출현하지만 주제 간 구분력과 해석 기여도가 낮은 일반어를 추가하여 확장 불용어 사전을 구성하였다. 또한 monke를 monkey로, goty를 game_of_the_year로 변환하는 등 주요 오타, 축약어 및 유사 표현을 통합하였다. 다만 china와 chinese는 문법적 기능과 사용 맥락이 다르므로 병합하지 않고 독립된 토큰으로 유지하였다. 분석의 재현성을 위해 확장 불용어와 변환어의 전체 목록은 별도로 기록하였다.

BERTopic 분석에는 문서의 문맥적 의미와 원문의 표현을

최대한 보존하기 위해 표제어 추출과 확장 불용어 제거를 적용하지 않은 노이즈 제거 텍스트를 사용하였다. 이에 따라 monke와 같은 원문 표현도 BERTopic 입력자료에서는 그대로 유지하였다. TF-IDF와 VADER 감성분석에는 최종 기본 코퍼스 66,981건을 사용하였다. LDA에는 정규화 후 3개 이상의 유효 토큰이 남은 리뷰 40,086건을, BERTopic에는 노이즈 제거 후 5개 이상의 토큰으로 구성된 리뷰 45,427건을 각각 투입하였다. 분석방법별 입력자료, 문서 선정 기준 및 최종 투입 문서 수는 표 3과 같다.

표 3. 분석방법별 입력자료 및 문서 선정 기준

Table 3. Input data and document selection criteria by analysis method

Analysis Method	Input Text	Selection Criterion	No. of Documents
TF-IDF	Lemmatized and stopword-filtered text	Final base corpus	66,981
VADER	Noise-removed original text	Final base corpus	66,981
LDA	Lemmatized and normalized text	At least 3 valid tokens	40,086
BERTopic	Noise-removed text without lemmatization	At least 5 tokens	45,427

Note: Document counts differ because preprocessing and minimum-token criteria varied by analysis method. The common sample included reviews that met the selection criteria for both LDA and BERTopic.

두 토픽모형의 투입 문서 집합 차이를 점검하기 위해 리뷰 ID를 기준으로 LDA와 BERTopic의 문서 선정 기준을 모두 충족한 리뷰의 교집합을 구성하였다. 그 결과 공통표본은 총 38,000건으로, LDA 투입 문서 40,086건의 94.8%, BERTopic 투입 문서 45,427건의 83.7%에 해당하였다. 시기별로는 출시 직후 24,320건(64.0%), 중기 7,790건(20.5%), 후기 5,890건(15.5%)으로 나타났다. 이 공통표본은 두 모형의 전체표본 분석결과와 비교하기 위한 민감도 분석에 활용하였다.

3-4 분석 방법

영어 Steam 리뷰에 나타난 이용자 담론과 감성의 시간적 변화를 분석하기 위해 TF-IDF 기반 키워드 분석, VADER 감성분석, LDA 및 BERTopic 토픽모델링을 수행하였다.

먼저 시기별 핵심 어휘와 변화 양상을 파악하기 위해 scikit-learn의 TfidfVectorizer를 활용하여 TF-IDF(term frequency-inverse document frequency)를 산출하였다. 분석 매개변수는 max_features=500, min_df=5, max_df=0.8로 설정하였다. 전체 기본 코퍼스에서 동일한 어휘사전과 IDF 값을 산출한 후 이를 출시 직후, 중기 및 후기 자료에 공통으로 적용하여 시기별 평균 TF-IDF와 상위 키워드 순위를 비교하

였다. 또한 주요 키워드의 월별 출현빈도를 해당 월의 전체 토큰 수로 표준화하여 1,000토큰당 상대빈도를 산출하고 변화 추이를 분석하였다.

영어 리뷰의 감성은 Hutto와 Gilbert[17]가 제안한 VADER를 활용하여 분석하였다. 대소문자, 문장부호, 부정어 및 강조 표현이 감성 판단에 활용된다는 점을 고려하여 표제어 추출과 불용어 제거 이전의 노이즈 제거 텍스트를 입력자료로 사용하였다. 각 리뷰에 대해 -1에서 1 사이의 compound 점수를 산출하였으며, VADER의 표준 분류기준에 따라 0.05 이상은 긍정, -0.05 이하는 부정, 그 사이는 중립으로 분류하였다. 이후 월별 평균 compound 점수와 긍정·중립·부정 리뷰의 비율을 산출하여 시기별 감성분포를 기술적으로 비교하였다.

잠재 주제 구조를 파악하기 위해 Blei 등[11]이 제안한 LDA를 적용하였다. 정규화 후 3개 이상의 유효 토큰이 남은 리뷰 40,086건을 대상으로 gensim 사전을 구축하였으며, 출현 문서 수가 10건 미만인 단어와 전체 문서의 70%를 초과하여 출현한 단어는 제외하였다. 적정 토픽 수를 결정하기 위해 토픽 수 K를 3에서 12까지 변화시키며 C_v Coherence[12]와 Log Perplexity를 산출하고, 두 지표와 토픽의 해석 가능성을 종합적으로 검토하였다. 최종 모형은 K=7, passes=20, alpha=auto, eta=auto, random_state=42로 설정하였다. 시기별 토픽 구성비는 각 리뷰를 최대 확률의 지배적 토픽에 배정하여 산출하였다.

LDA의 단어빈도 기반 한계를 보완하고 리뷰의 문맥적 의미를 반영하기 위해 Grootendorst[13]가 제안한 BERTopic을 병행 적용하였다. 분석에는 표제어 추출 이전의 노이즈 제거 텍스트를 사용하였으며, 5개 이상의 토큰으로 구성된 리뷰 45,427건을 투입하였다. 문서 임베딩에는 Sentence-BERT 계열의 all-MiniLM-L6-v2 모형[14]을 사용하였고, UMAP을 통한 차원 축소와 HDBSCAN 군집화를 적용하였다. 최소 토픽 크기는 50으로 설정하였으며, 토픽 표현에는 영어 불용어를 제거한 1~2그램 CountVectorizer(min_df=10)와 클래스 기반 TF-IDF(c-TF-IDF)를 적용하였다. 별도의 토픽 축소와 이상치 재배정은 수행하지 않았으며, UMAP과 HDBSCAN의 세부 매개변수는 부록에 제시하였다.

BERTopic에서 특정 군집에 포함되지 않은 문서는 이상치로 처리하였다. LDA와 BERTopic의 토픽명은 상위 키워드와 대표 리뷰를 바탕으로 부여하였다. 제1저자가 토픽명의 초안을 작성하고, 공동연구자의 검토와 협의를 거쳐 최종 확정하였다. 시기별 BERTopic 토픽 비율은 각 시기의 비이상치 문서를 기준으로 산출하였으며, 해당 비율은 전체 투입 문서가 아니라 토픽에 배정된 문서 내 구성비를 의미한다.

BERTopic에서 도출된 75개 토픽 중 본문에는 문서 수를 기준으로 상위 8개 토픽을 제시하였다. 상위 8개 토픽에는 총 19,793건의 문서가 포함되었으며, 이는 전체 비이상치 문서 33,522건의 약 59.0%에 해당한다. 나머지 67개 토픽도 분석에서 제외하지 않고 전체 토픽 분포와 시기별 구성비 산출에

포함하였으며, 본문에서는 가독성과 해석의 집중도를 고려하여 개별 설명만 생략하였다.

마지막으로 LDA와 BERTopic에서 도출된 토픽의 구성과 시기별 변화 방향을 비교하여 두 모형에서 공통적으로 나타난 주요 담론과 차이를 살펴보았다. 분석방법별 투입 문서 집합의 차이가 비교 결과에 미치는 영향을 점검하기 위해 리뷰 ID를 기준으로 두 모형의 문서 선정 기준을 모두 충족한 공통 표본을 구성하였다. 공통표본에서는 각 모형의 기존 전처리 방식과 최종 토픽 배정 결과를 유지한 상태에서 시기별 토픽 구성비를 재산출하였다. 이후 전체표본과 공통표본에서 나타난 주요 담론의 구성비와 변화 방향을 비교하여 결과의 일관성을 검토하였다. 이 민감도 분석은 문서 집합의 차이가 분석 결과에 미치는 영향을 점검하기 위한 것으로, 두 모형의 성능 우열이나 통계적 교차검증을 의미하지 않는다.

모든 분석은 Python 3 기반의 Google Colab 환경에서 수행하였다. 분석의 재현성을 높이기 위해 LDA와 UMAP의 난수 시드를 42로 고정하였으며, BERTopic의 세부 분석 설정은 부록에 제시하였다.

IV. 분석 결과

4-1 분석자료의 월별 분포

최종 기본 코퍼스인 영어 Steam 리뷰 66,981건의 월별 분포는 그림 1과 같다. 리뷰는 게임 출시 직후인 2024년 8월과 9월에 집중되었으며, 이 기간에 작성된 리뷰는 총 42,860건으로 전체의 64.0%를 차지하였다. 특히 2024년 8월은 8월 20일부터 31일까지의 부분일 자료임에도 가장 많은 리뷰가 작성되었고, 9월에도 높은 수준이 이어졌다.

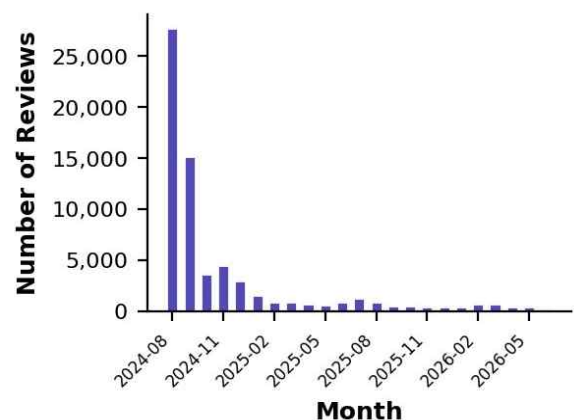


그림 1. 영어 Steam 리뷰의 월별 분포

Fig. 1. Monthly distribution of English-language steam reviews

2024년 10월 이후 월별 리뷰 수는 크게 감소하였으며, 2025년부터는 비교적 낮은 수준에서 증감을 반복하였다. 다

만 수집 종료 시점까지 리뷰가 지속적으로 작성되어, 출시 초기의 집중적인 반응 이후에도 이용자 평가가 장기간 축적되었음을 확인할 수 있다.

시기별 리뷰 수는 출시 직후 42,860건(64.0%), 중기 13,702건(20.5%), 후기 10,419건(15.6%)이었다. 세 시기는 기간과 리뷰 수가 서로 다르므로, 이후 분석에서는 원시빈도의 단순 비교보다 감성 및 토픽 구성비, 1,000토큰당 상대빈도와 월별 변화 추이를 중심으로 해석하였다. 또한 2026년 6월 자료는 6월 17일까지 수집된 부분월 자료이므로 해당 월의 결과는 제한적으로 해석하였다.

4-2 시기별 키워드 및 상대빈도 분석

시기별 평균 TF-IDF 상위 10개 키워드는 그림 2와 같다. monkey는 출시 직후 0.145, 중기 0.138, 후기 0.092로 세 시기 모두 가장 높은 값을 나타냈다. 이는 원숭이 캐릭터 관련 표현이 분석기간 전반의 주요 어휘였으나, 후기에는 상대적 중요성이 감소했음을 보여준다. game_of_the_year는 출시 직후와 중기에 두 번째로 높은 값을 보였으나 후기에는 상위 10개에서 제외되었다. 이는 작품을 올해의 게임 수준 또는 후보로 평가하는 표현과 시상 관련 화제가 두 시기에 상대적으로 두드러졌음을 보여준다.

반면 boss는 출시 직후 3위에서 후기 2위로 상승하였으며, combat, hard, soul, fight, gameplay 등 전투, 난이도 및 장르적 특성과 관련된 어휘는 후기에도 상위권에 포함되었다. 이는 이용자 담론의 중심이 초기의 캐릭터·밈과 올해의 게임 평가 및 시상 관련 화제에서 보스전, 전투 체계, 난이도 및 실제 플레이 경험에 관한 평가로 점차 이동했음을 시사한다. 시기별 TF-IDF는 전체 코퍼스에서 산출한 동일한 어휘사전과 IDF 값을 공통 적용하여 비교하였다.

핵심 키워드의 월별 출현빈도를 1,000토큰당 상대빈도로 환산한 결과는 그림 3과 같다. 문화·정체성 관련 키워드 중 wukong은 분석기간 전반에 걸쳐 5.2~10.6회로 가장 높은 수준을 유지하였다. chinese는 2.9~5.9회로 china의 0.3~1.3회보다 높게 나타났다. 이는 영어 리뷰에서 국가명을 직접 지칭하는 명사형 표현보다 chinese mythology, chinese culture와 같이 문화적·신화적 대상을 수식하는 표현이 상대적으로 자주 사용되었을 가능성을 보여준다. 다만 두 단어는 문법적 기능과 결합 방식이 다르므로, 빈도 차이만으로 중국 문화에 대한 이용자의 태도나 평가 방향을 판단하기는 어렵다.

mythology는 분석기간 전반에 걸쳐 1.6~3.0회 범위에서 반복적으로 나타났다. 이는 신화적 소재가 특정 시기에만 일시적으로 언급된 것이 아니라 출시 이후에도 지속적인 담론 요소로 나타났음을 보여준다. 다만 뚜렷한 증가 추세보다는 일정한 범위 내에서 월별 변동을 보인 것으로 해석하는 것이 적절하다.

게임플레이 및 품질 관련 키워드에서는 boss가 16.8~26.9회로 분석기간 전반에 걸쳐 가장 높은 상대빈도를 보였

다. story, combat, graphic, hard도 월별 변동은 있었으나 전 기간에 걸쳐 비교적 높은 수준을 유지하였다. 이는 출시 초기의 화제성이 감소한 이후에도 이용자의 관심이 보스전, 전투 시스템, 서사, 그래픽 및 난이도와 같은 실제 플레이 경험의 주요 요소에 지속적으로 집중되었음을 보여준다. 다만 후기 일부 월은 리뷰 수와 전체 토큰 수가 적어 상대빈도의 변동성이 커질 수 있으므로, 개별 월의 수치보다 전반적인 변화 방향을 중심으로 해석하였다.

종합하면, 출시 초기에는 캐릭터·밈과 올해의 게임 평가 및 시상 관련 어휘가 두드러졌으나, 후기에는 보스전, 전투, 난이도 및 장르적 특성과 관련된 어휘의 상대적 중요성이 높아졌다. 중국 신화와 문화적 소재는 뚜렷한 증가세를 보이지는 않았지만 분석기간 전반에 걸쳐 반복적으로 나타나, 작품의 신화적 정체성이 장기적인 이용자 담론에서도 지속되었음을 보여준다.

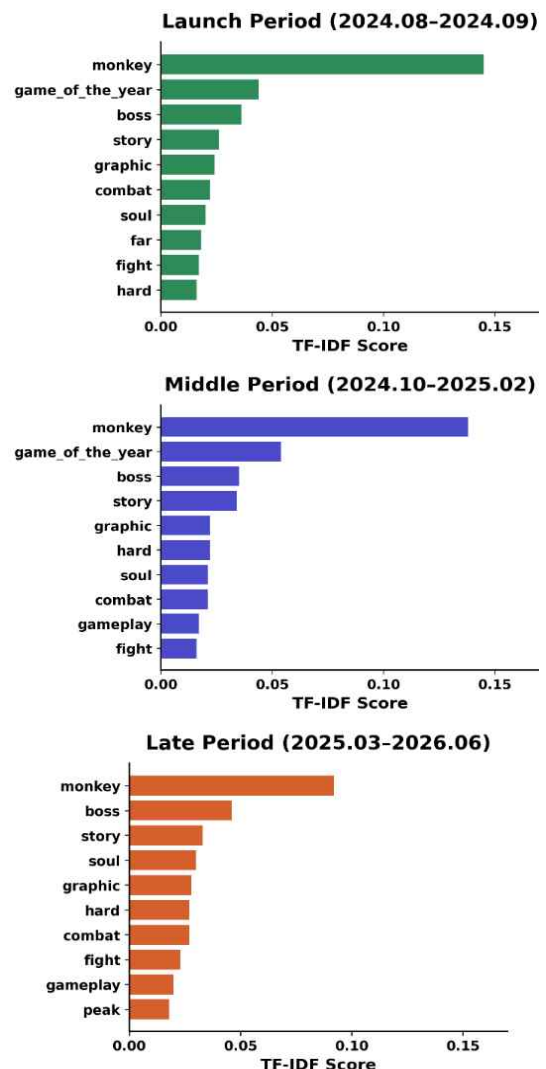


그림 2. 시기별 TF-IDF 상위 10개 키워드

Fig. 2. Top 10 TF-IDF keywords by period

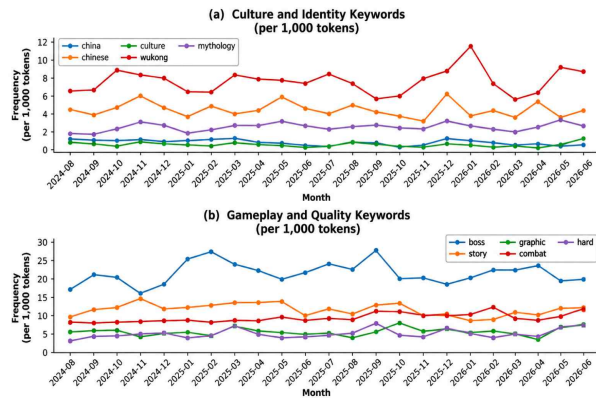


그림 3. 핵심 키워드의 월별 상대빈도: (a) 문화·정체성, (b) 게임플레이·품질

Fig. 3. Monthly relative frequencies of selected keywords per 1,000 tokens: (a) Culture and identity and (b) Gameplay and quality

4-3 감성 분석 결과

VADER 감성분석 결과, 월평균 compound 점수는 분석기간 전반에 걸쳐 0.32~0.42 범위의 양의 값을 유지하였다. 월별 긍정 리뷰 비율은 58.5~68.0%, 부정 리뷰 비율은 8.4~16.5% 범위에서 변동하였다. 2026년 6월의 긍정 리뷰 비율은 68.0%로 나타났으나, 해당 자료는 6월 17일까지 수집된 부분월 자료이므로 제한적으로 해석할 필요가 있다. 월별 평균 compound 점수와 긍정·부정 리뷰 비율의 변화는 그림 4와 같다.

시기별 감성분포는 그림 5와 같다. 긍정 리뷰 비율은 출시 직후 60.3%, 중기 60.4%, 후기 63.2%로 나타나 후기에는 출시 직후보다 2.9%p 높았다. 중립 리뷰 비율은 출시 직후 27.6%에서 중기 28.8%로 소폭 증가한 뒤 후기에는 22.7%로 감소하였다. 부정 리뷰 비율은 출시 직후 12.1%, 중기 10.8%, 후기 14.1%로 나타났다.

후기에는 중립 리뷰 비율이 감소한 반면 긍정과 부정 리뷰 비율이 모두 증가하였다. 이는 감성분류 결과가 중립보다 긍정 또는 부정 범주에 더 많이 분포한 양상으로 볼 수 있다. 다만 이러한 변화만으로 이용자 평가의 양극화나 감성표현의 강화를 단정하기는 어렵다. 전반적으로는 모든 시기에서 긍정 리뷰 비율이 60% 이상으로 나타나, VADER 분류상 긍정 범주가 우세한 경향이 지속되었다.

4-4 LDA 토픽 모델링 결과

적정 토픽 수를 결정하기 위해 토픽 수 K를 3개부터 12개까지 변화시키며 C_v Coherence와 Log Perplexity를 산출하였다. 그 결과 K=7에서 Coherence가 0.5785로 가장 높았으며, Log Perplexity도 비교 모형 가운데 상대적으로 양호한 수준을 보였다. K=11에서도 비교적 높은 Coherence가

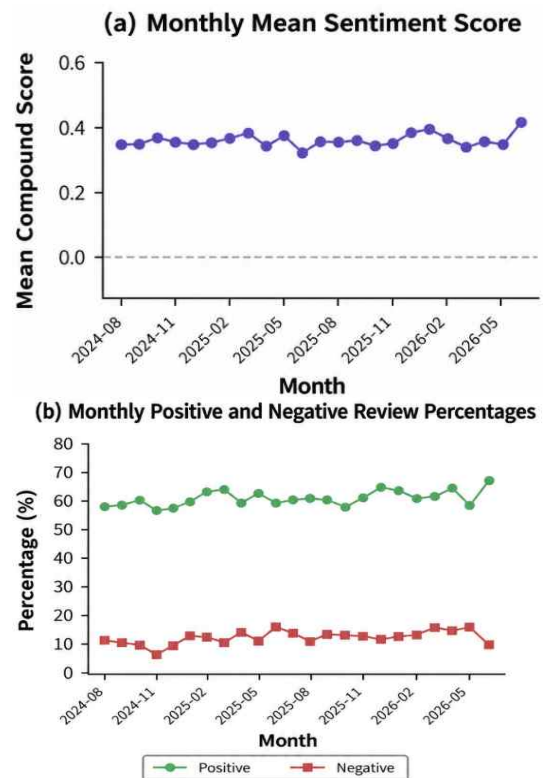


그림 4. 월별 감성 변화: (a) 평균 compound 점수, (b) 긍정·부정 리뷰 비율

Fig. 4. Monthly sentiment trends: (a) Mean compound score and (b) Positive and negative review ratios

Sentiment Distribution by Period

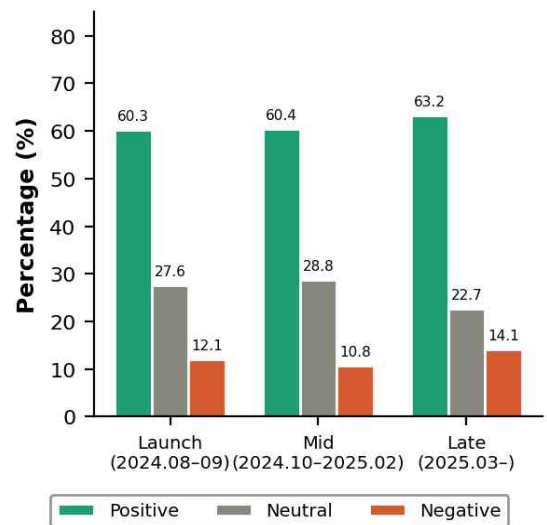


그림 5. 시기별 감성분포

Fig. 5. Sentiment distribution by period

나타났으나, 토픽의 해석 가능성과 모형의 간결성을 종합적으로 고려하여 K=7을 최종 선택하였다. 토픽 수에 따른 모형 평가 결과는 그림 6과 같다.

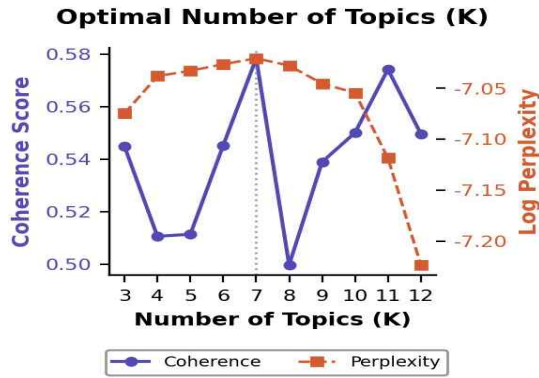


그림 6. 토픽 수에 따른 Coherence 및 Log Perplexity 변화
Fig. 6. Coherence and log perplexity by number of topics

최종 LDA 모형의 토픽은 상위 키워드와 대표 리뷰를 함께 검토하여 명명하였다. T1은 see, wait, say, people 등 일반적인 서술 표현이 중심을 이루어 특정 게임요소와 직접 연결되지 않는 비구체적 일반 반응으로 해석하였다. T2는 boss, chapter, fight, enemy, map, level 등을 포함한 보스 전투·레벨 디자인, T3은 wukong, black, myth, journey, chinese, west로 구성된 중국 신화·여정, T4는 crash, setting, performance, fps 등을 중심으로 한 기술·성능, T5는 story, soul, combat, graphic, beautiful 등을 포함한 전반적 게임 품질·호평, T6은 grind, bad, bug, average 등을 중심으로 한 명시적 부정 평가, T7은 monkey, stick, return, king 등으로 구성된 원숭이 밈으로 해석하였다.

각 리뷰를 최대 확률의 지배적 토픽에 배정하여 시기별 구성비를 산출한 결과는 그림 7과 같다. 가장 높은 비중을 차지한 토픽은 전반적 게임 품질·호평(T5)과 보스 전투·레벨 디자인(T2)이었다. T5는 출시 직후 32.3%, 중기 35.0%, 후기 35.8%였으며, T2는 각각 29.7%, 29.6%, 36.4%를 차지하였다. 후기에는 두 토픽의 합계가 72.2%로 나타나, 이용자 담론이 전반적 게임 품질과 보스 전투 및 레벨 설계에 관한 평가를 중심으로 형성되었음을 보여준다.

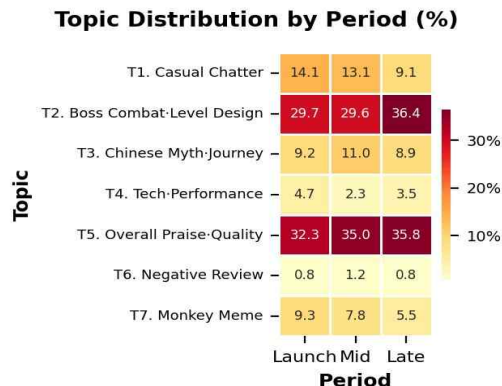


그림 7. LDA 토픽의 시기별 분포
Fig. 7. LDA topic distribution by period

중국 신화·여정(T3)은 출시 직후 9.2%, 중기 11.0%, 후기 8.9%로 나타나 세 시기 모두 약 9~11%의 비중을 유지하였다. 이는 중국 신화와 『서유기』 관련 담론이 출시 초기의 일시적 화제에 그치지 않고 후기에도 독립적인 토픽으로 유지되었음을 보여준다.

원숭이 밈(T7)은 출시 직후 9.3%에서 중기 7.8%, 후기 5.5%로 지속적으로 감소하였다. 비구체적 일반 반응(T1)도 14.1%에서 13.1%, 9.1%로 감소하였다. 반면 보스 전투와 게임 품질 관련 토픽의 비중은 증가하여, 후기 리뷰에서는 구체적인 플레이 경험과 평가요소가 상대적으로 두드러졌다. 기술·성능(T4)은 출시 직후 4.7%에서 중기 2.3%로 감소한 뒤 후기에는 3.5%로 소폭 증가하였다.

명시적 부정 평가(T6)는 전 기간 0.8~1.2%의 낮은 비중을 보였다. 다만 이는 부정적 감성을 지닌 전체 리뷰의 비율이 아니라 bad, bug, grind 등의 표현이 하나의 토픽으로 응집된 문서의 비율이다. 부정적 평가는 보스 전투, 난이도 및 기술·성능 등 다른 토픽에도 포함될 수 있으므로 VADER 감성분석의 부정 리뷰 비율과 직접 비교하기는 어렵다.

종합하면, LDA 분석에서는 출시 초기의 원숭이 밈과 비구체적 일반 반응 관련 토픽이 감소한 반면, 후기에는 보스 전투·레벨 디자인과 전반적 게임 품질 평가의 비중이 확대되었다. 중국 신화·여정 토픽은 전 기간 일정한 비중을 유지하여, 작품의 문화적 소재가 장기적인 이용자 담론에서도 지속적으로 나타났음을 보여준다.

4-5 BERTopic 토픽모델링 결과

문맥 임베딩 기반 BERTopic을 적용한 결과 총 75개의 토픽이 도출되었다. 이 가운데 특정 토픽에 배정되지 않은 이상치 문서는 11,905건으로, 전체 입력 문서 45,427건의 26.2%를 차지하였다. 이에 따라 특정 토픽에 배정된 비이상치 문서는 총 33,522건이었다. 본문에는 문서 수를 기준으로 선정한 상위 8개 토픽을 제시하였으며, 이들 토픽에는 총 19,793건의 문서가 포함되어 전체 비이상치 문서의 약 59.0%를 차지하였다. 나머지 67개 토픽도 분석에서 제외하지 않고 전체 토픽 구성과 시기별 비율 산출에 포함하였다.

가장 규모가 큰 토픽은 wukong, myth, black myth를 주요 키워드로 하는 오공·신화(BT0)로, 4,087건의 문서가 포함되었다. 이어 기술·성능(BT1, 2,698건), 원숭이 캐릭터(BT2, 2,641건), 보스 전투(BT3, 2,453건), 중국 신화·문화(BT4, 2,342건) 순으로 나타났다. 특히 오공·신화(BT0)와 중국 신화·문화(BT4)가 별개의 토픽으로 도출된 것은 작품명과 주인공을 중심으로 한 신화적 정체성 담론과 중국의 신화·문화적 배경을 직접 언급한 담론이 서로 다른 문맥적 군집으로 구분되었음을 보여준다.

시기별 상위 토픽의 분포는 그림 8과 같다. 각 비율은 시기별 비이상치 문서를 기준으로 산출한 것으로, 전체 투입 리뷰 대비 비율이 아니라 토픽에 배정된 문서 내 구성비를 의미한다.

다. 오공·신화(BT0)는 출시 직후 11.4%, 중기 13.1%, 후기 14.4%로 증가하였으며, 상위 8개 토픽 중 모든 시기에서 가장 높은 비중을 차지하였다. 이는 비이상치 문서 내에서 작품의 신화적 정체성과 관련된 담론의 상대적 비중이 후기까지 증가했음을 보여준다.

보스 전투(BT3)는 출시 직후 6.7%에서 중기 7.1%, 후기 10.3%로 증가하였으며, 소울라이크 장르(BT6)도 5.4%에서 6.7%, 9.1%로 상승하였다. 반면 원숭이 캐릭터(BT2)는 8.2%에서 7.8%, 6.7%로 감소하였고, monke, return monke 등의 표현으로 구성된 원숭이 밈(BT5)도 7.6%에서 4.5%, 2.9%로 줄어들었다. 이는 출시 초기의 캐릭터·밈 중심 담론이 점차 감소하고, 후기에는 보스 전투와 장르적 특성에 관한 구체적인 평가가 상대적으로 두드러졌음을 보여준다.

중국 신화·문화(BT4)는 출시 직후 7.4%, 중기 6.4%, 후기 6.0%로 다소 감소하였으나, 세 시기 모두 독립된 토픽으로 유지되었다. 따라서 중국 문화 관련 담론이 확대되었다고 보기는 어렵지만, 출시 이후에도 일정한 비중으로 지속되었다고 해석할 수 있다. 기술·성능(BT1)은 출시 직후 9.3%에서 중기

4.4%로 감소한 뒤 후기에는 7.5%로 다시 증가하여, 성능과 최적화 관련 평가가 시기별로 변동하였음을 보여준다.

종합하면, BERTopic 분석에서는 작품의 신화적 정체성과 관련된 담론이 주요 토픽으로 지속되는 가운데, 초기의 캐릭터·밈 중심 담론은 감소하고 보스 전투와 소울라이크 장르 관련 담론의 비중은 증가하였다. 이는 이용자 담론의 중심이 출시 초기의 화제성에서 실제 게임플레이 경험과 장르적 평가로 점차 이동하였음을 보여준다.

4-6 LDA와 BERTopic 결과 비교 및 공통표본 민감도 분석

LDA와 BERTopic의 분석 특성은 표 5와 같다. LDA는 단어의 확률분포를 기반으로 7개의 비교적 포괄적인 토픽을 도출한 반면, BERTopic은 문맥 임베딩과 밀도 기반 군집화를 통해 75개의 세분화된 토픽을 도출하였다. 또한 LDA는 각 문서를 여러 토픽의 확률분포로 표현하며, 본 연구에서는 최대 확률의 지배적 토픽을 기준으로 시기별 구성비를 산출하였다. 반면 BERTopic은 문맥적으로 유사한 문서를 군집에 배정하고, 군집의 밀도가 충분하지 않은 문서를 이상치로 분류한다는 차이가 있다.

두 모형은 토픽의 수와 세분화 수준에서는 차이를 보였으나, 주요 담론과 시기별 변화 방향에서는 유사한 경향이 나타났다. 먼저 LDA의 보스 전투·레벨 디자인(T2)과 전반적 게임 품질·호평(T5), BERTopic의 보스 전투(BT3), 소울라이크 장르(BT6), 전투·그래픽(BT7)은 모두 게임플레이와 품질 평가가 이용자 담론의 주요 주제로 나타났음을 보여준다.

신화·문화 관련 담론도 두 모형에서 공통적으로 확인되었다. LDA의 중국 신화·여정(T3)은 세 시기 동안 약 9~11%의 비중을 유지하였다. BERTopic에서는 오공·신화(BT0)의 비중이 후기까지 증가한 반면, 중국 신화·문화(BT4)는 소폭 감소하면서도 독립된 토픽으로 유지되었다. 이는 중국 신화 관련 담론이 두 모형에서 지속적으로 나타난 가운데, BERTopic이 이를 작품의 신화적 정체성과 중국의 문화적 배경에 관한 담론으로 보다 세분화하여 제시했음을 보여준다.

시기별 변화 방향에서도 공통점이 나타났다. LDA의 원숭이 밈(T7)과 BERTopic의 원숭이 캐릭터(BT2) 및 원숭이 밈(BT5)은 후기까지 비중이 감소하였다. 반면 보스 전투와 소울라이크 장르 관련 토픽의 비중은 증가하여, 이용자 담론이 출시 초기의 캐릭터·밈 중심 반응에서 실제 게임플레이와 장르적 평가 중심으로 변화하는 경향이 두 모형에서 공통적으로 확인되었다.

종합하면, LDA는 이용자 담론의 포괄적 구조와 전반적인 변화 방향을 제시한 반면, BERTopic은 이를 세분화된 문맥 기반 토픽으로 구체화하였다. 두 모형에서는 게임플레이와 품질 평가의 중심성, 중국 신화 관련 담론의 지속성, 그리고 캐릭터·밈 중심 반응에서 게임플레이·장르 평가 중심 담론으로의 변화라는 공통된 흐름이 나타났다.

표 4. BERTopic 상위 8개 토픽

Table 4. Top eight BERTopic topics

Topic	Topic Label	Count	Representative Keywords
BT0	Wukong and Mythology	4,087	wukong, myth, black myth
BT1	Technology and Performance	2,698	settings, fps, pc, performance
BT2	Monkey Character	2,641	monkey, stick, monkey game
BT3	Boss Combat	2,453	boss, bosses, boss fights
BT4	Chinese Mythology and Culture	2,342	chinese, china, mythology
BT5	Monkey Meme	2,099	monke, return monke, stick
BT6	Soulslike Genre	2,083	souls, souls like, soulslike
BT7	Combat and Graphics	1,390	combat, great, graphics

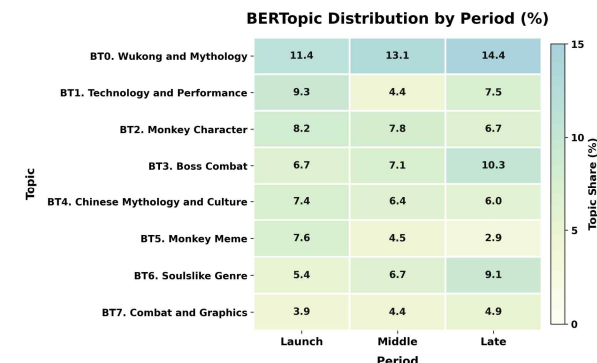


그림 8. BERTopic 상위 8개 토픽의 시기별 분포

Fig. 8. Distribution of the top eight BERTopic topics by period based on non-outlier documents

표 5. LDA와 BERTopic의 분석 특성 비교

Table 5. Comparison of LDA and BERTopic

Category	LDA	BERTopic
Analysis Basis	Bag-of-words and probabilistic modeling	Contextual embeddings and clustering
No. of Topics	7, researcher-selected	75, parameter-derived
Topic Granularity	Broad topics	Fine-grained topics
Document Assignment	Topic probabilities	Cluster or outlier
Contextual Information	Not directly reflected	Reflected
Outlier Ratio	Not applicable	26.2%
Documents Analyzed	40,086	45,427

두 모형의 투입 문서 집합이 서로 다르다는 점을 고려하여, 두 모형의 문서 선정 기준을 모두 충족한 공통표본 38,000건을 대상으로 민감도 분석을 수행하였다. 공통표본에서도 각 모형의 기존 전처리 방식과 토픽 배정 결과를 유지한 상태에서 시기별 토픽 구성비를 재산출하고, 이를 전체표본의 결과와 비교하였다.

분석 결과, 공통표본에서도 전체표본과 유사하게 캐릭터·ميم 관련 토픽의 비중은 후기까지 감소한 반면, 보스 전투와 게임플레이 관련 토픽의 비중은 증가하였다. 중국 신화와 문화 관련 토픽도 분석기간 전반에 걸쳐 독립적인 주제로 유지되었다. 일부 토픽의 세부 구성비에는 차이가 있었으나, 주요 담론의 구성과 시기별 변화 방향은 전체표본과 공통표본에서 대체로 일관되게 나타났다. 이는 투입 문서 집합의 차이가 본 연구의 주요 해석 방향을 크게 변화시키지 않았음을 보여준다. 다만 이러한 결과는 두 모형의 통계적 일치나 성능 우열

을 검증한 것이 아니라, 문서 집합 차이에 따른 결과의 민감도를 점검한 것으로 해석해야 한다.

V. 논 의

본 연구의 감성분석 결과, VADER 분류상 긍정 범주가 전 기간에 걸쳐 우세하게 나타나 출시 초기의 관심이 감소한 이후에도 작품에 대한 긍정적 평가가 일정 수준 유지되었음을 확인하였다. 이는 플랫폼과 텍스트 특성의 차이에도 불구하고, 관련 YouTube 댓글에서 중립 및 긍정 감성이 우세하게 나타났다고 보고한 Tang 등[2]의 결과와 유사한 방향성을 보인다. 다만 YouTube 댓글과 Steam 리뷰는 작성 목적, 이용자 구성 및 실제 플레이 경험의 반영 정도가 다르므로, 두 결과는 직접적인 일치나 재현이 아닌 참고적 비교로 해석할 필요가 있다. 또한 후기에는 중립 범주가 감소하고 긍정과 부정 범주가 함께 증가하였으나, 이를 이용자 평가의 양극화로 단정하기보다 감성분류 결과가 중립보다 긍정 또는 부정 범주에 상대적으로 더 많이 분포한 것으로 해석하는 것이 적절하다.

이용자 담론은 출시 초기의 캐릭터·ميم과 올해의 게임 평가 및 시상 관련 화제에서, 시간이 지나면서 보스 전투, 장르적 특성 및 게임 품질 평가 중심으로 변화하였다. 이는 출시 당시의 화제성에 기반한 반응이 장기적인 플레이 경험의 축적과 함께 구체적인 게임요소에 대한 평가로 전환되었음을 의미한다. 이러한 결과는 Steam 리뷰가 이용자의 플레이 경험과 주요 관심 요소를 반영한다는 선행연구[4],[5]와 맥을 같이하며, 플레이타임에 따라 리뷰의 주요 내용이 달라질 수 있다고 본 최영우와 유승호[6]의 결과와도 연결된다. 본 연구는

표 6. 전체표본과 공통표본의 주요 토픽 구성비 비교

Table 6. Comparison of major topic proportions between the full and common samples

Model	Topic	Sample	Launch Period	Middle Period	Late Period	Temporal Pattern
LDA	Boss Combat and Level Design (T2)	Full Sample	29.7%	29.6%	36.4%	Increase
		Common Sample	29.9%	29.8%	36.1%	Increase
	Chinese Mythology and Journey (T3)	Full Sample	9.2%	11.0%	8.9%	Sustained
		Common Sample	9.4%	10.8%	9.1%	Sustained
	Monkey Meme (T7)	Full Sample	9.3%	7.8%	5.5%	Decrease
		Common Sample	9.1%	7.6%	5.7%	Decrease
BERTopic	Boss Combat (BT3)	Full Sample	6.7%	7.1%	10.3%	Increase
		Common Sample	6.8%	7.3%	10.1%	Increase
	Chinese Mythology and Culture (BT4)	Full Sample	7.4%	6.4%	6.0%	Sustained
		Common Sample	7.2%	6.5%	6.1%	Sustained
	Monkey Meme (BT5)	Full Sample	7.6%	4.5%	2.9%	Decrease
		Common Sample	7.5%	4.6%	3.0%	Decrease

Note: LDA proportions were calculated based on all documents included in each period, whereas BERTopic proportions were calculated based on non-outlier documents assigned to topics in each period.

이러한 논의를 시기별 변화의 관점으로 확장하여, 동일한 게임에서도 출시 이후 시간의 경과에 따라 이용자 담론의 중심이 달라질 수 있음을 보여주었다.

Singh와 O'Hagan[7]의 연구와 본 연구는 Steam 리뷰를 활용하여 감성과 담론의 장기적 변화를 분석하였다는 점에서 공통적이다. 그러나 [7]은 소셜 VR 커뮤니티에서 나타나는 신흥 문제와 쟁점의 모니터링을 목적으로 VADER와 BERTopic을 적용한 반면, 본 연구는 문화적 소재를 핵심으로 하는 단일 게임을 대상으로 TF-IDF를 추가하고 LDA와 BERTopic을 병행하였다. 이를 통해 이용자 관심의 장기적 변화뿐 아니라 중국 신화와 문화적 소재의 담론적 지속성, 그리고 두 토픽모형이 제시하는 분석 수준의 차이를 함께 확인하였다는 점에서 기존 연구를 확장하였다.

중국 신화와 문화적 소재는 두 토픽모형에서 후기까지 독립적인 주제로 유지되었다. 이는 중국 신화와 『서유기』 관련 소재가 출시 초기의 홍보나 일시적인 화제에 그치지 않고, 작품의 정체성을 구성하고 이용자가 게임을 해석하는 지속적인 담론 요소로 기능했음을 보여준다. 이러한 결과는 문화적 소재의 의미가 이용자의 경험과 기존 지식을 통해 해석된다는 Hall[8]의 관점, 미디어 수용을 의미의 이해와 평가가 결합된 과정으로 본 Schröder[9]의 논의, 게임문화를 이용자 담론을 통해 구성되는 대상으로 본 Shaw[10]의 관점과 연결된다. 다만 본 연구는 이들 이론이나 이용자의 중국 문화에 대한 태도 변화를 직접 검증한 것이 아니므로, 결과는 중국 신화와 문화적 소재가 장기간 독립적인 리뷰 담론으로 지속되었다는 범위로 제한하여 해석할 필요가 있다.

방법론적으로 LDA는 이용자 담론의 포괄적인 구조를 제시한 반면, BERTopic은 신화, 캐릭터, 밈, 성능, 전투 및 장르 관련 담론을 세분화하여 제시하였다. 이는 두 모형의 병행 활용이 포괄적인 토픽을 구체화하고 하위 담론을 탐색하는데 유용하다고 본 김선옥과 양기덕[15], Dehghani와 Zaman[16]의 논의와 부합한다. 또한 공통표본 민감도 분석에서도 캐릭터·밈 중심 담론의 감소, 게임플레이 관련 담론의 증가 및 중국 신화·문화 담론의 지속이라는 주요 변화 방향이 전체표본과 대체로 유사하게 나타났다. 이는 분석방법별 투입 문서 집합의 차이가 본 연구의 핵심적인 해석 방향을 크게 변화시키지 않았음을 보여준다. 다만 LDA와 BERTopic은 전처리 방식과 문서 배정 원리가 다르므로, 이러한 결과를 두 모형의 통계적 일치나 성능 우열을 검증한 것으로 확대 해석해서는 안 된다.

실무적으로는 게임 출시 초기의 판매량이나 화제성뿐 아니라, 출시 이후 장기간 축적되는 전투, 난이도, 서사, 그래픽, 성능 및 장르 관련 평가를 지속적으로 관찰할 필요가 있다. 특히 문화적 소재를 활용한 게임에서는 해당 소재가 단기적인 마케팅 요소에 머무는지, 장기적인 작품 정체성과 이용자 담론으로 이어지는지를 함께 점검해야 한다. 이러한 장기적 리뷰 분석은 업데이트 우선순위 설정, 품질 개선 및 문화 기반 게임 콘텐츠의 운영 전략을 수립하는 데 활용될 수 있다.

VI. 결론 및 제언

본 연구는 2024년 8월 20일부터 2026년 6월 17일까지 Steam에 게시된 영어 리뷰 중 플레이타임이 1시간 이상인 66,981건을 대상으로 TF-IDF 키워드 분석, VADER 감성분석, LDA 및 BERTopic 토픽모델링을 수행하였다.

분석 결과, VADER 분류상 긍정 범주는 전 기간에 걸쳐 우세하게 나타났다. 이용자 담론은 출시 초기의 캐릭터·밈과 올해의 게임 평가 및 시장 관련 화제에서 후기의 보스 전투, 소울라이크 장르 및 게임 품질 평가 중심으로 변화하였다. 중국 신화와 문화적 소재는 분석기간 전반에 걸쳐 독립적인 담론 주제로 지속되었다. 또한 LDA와 BERTopic은 토픽의 세분화 수준에서는 차이를 보였으나, 게임플레이와 품질 평가의 중심성, 중국 신화 관련 담론의 지속성 및 초기 화제성에서 실제 플레이 경험 중심 담론으로의 변화라는 공통된 흐름을 제시하였다.

본 연구는 약 22개월간 축적된 대규모 Steam 리뷰를 활용하여 이용자 감성과 담론의 장기적 변화를 분석하였다는 데 의의가 있다. 또한 LDA와 BERTopic을 병행하여 포괄적인 담론구조와 세분화된 문맥 기반 토픽을 함께 제시하고, 공통표본 민감도 분석을 통해 주요 변화 방향의 일관성을 점검하였다. 이를 통해 장기적 게임 리뷰 분석에서 복수의 토픽모형을 상호보완적으로 활용할 수 있는 가능성을 제시하였다. 산업적 측면에서는 출시 이후 변화하는 이용자 관심과 품질 평가를 지속적으로 관찰하고, 이를 게임 업데이트와 운영 전략에 반영할 필요가 있음을 시사한다.

본 연구의 한계는 다음과 같다. 첫째, 분석대상이 영어 Steam 리뷰에 한정되어 있고 자발적으로 리뷰를 작성한 이용자의 의견만 포함하므로, 연구결과를 전체 게임 이용자의 인식으로 일반화하기 어렵다. 향후에는 다른 언어권의 Steam 리뷰와 Metacritic, Reddit 및 YouTube 등 다양한 플랫폼의 자료를 비교할 필요가 있다.

둘째, 플레이타임 1시간 미만의 리뷰를 제외함으로써 초기 실행 오류, 성능 문제 및 환불 경험이 충분히 반영되지 않았을 가능성이 있다. 후속 연구에서는 전체 리뷰와 플레이타임 집단별 결과를 비교하여 필터링 기준에 따른 민감도를 검토할 필요가 있다.

셋째, VADER는 반어, 밈, 은어 및 문맥 의존적 표현을 완전히 반영하기 어렵다. 향후에는 Steam의 추천·비추천 정보, 수동 라벨링 자료 및 문맥 기반 감성분류모형을 활용하여 감성분류의 타당성을 검증할 필요가 있다.

넷째, 공통표본 민감도 분석을 통해 주요 변화 방향의 일관성을 점검하였으나, LDA와 BERTopic은 전처리 방식과 문서 배정 원리가 서로 다르며 BERTopic에서는 일부 문서가 이상치로 분류되었다. 향후에는 시기별 이상치의 특성과 매개변수 변화에 따른 토픽 안정성을 추가로 검토할 필요가 있다.

마지막으로 본 연구는 중국 신화와 문화적 소재의 담론적

지속성을 확인하였으나, 이것이 이용자의 중국 문화 인식이나 국가 이미지 변화로 이어졌는지는 검증하지 못하였다. 향후에는 문화 관련 리뷰의 문맥 분석과 수동 코딩, 설문 및 인터뷰를 병행하여 문화적 소재가 이용자에게 해석되고 평가되는 과정을 보다 정교하게 분석할 필요가 있다.

부 록

표 7. BERTopic 분석의 주요 설정

Table 7. Main settings for BERTopic analysis

Analysis Component	Main Settings
Embedding Model	all-MiniLM-L6-v2
UMAP	n_neighbors=15, n_components=5, min_dist=0.0, metric='cosine', random_state=42
HDBSCAN	min_cluster_size=50, min_samples=10, metric='euclidean', cluster_selection_method='eom'
Topic Representation	CountVectorizer(ngram_range=(1,2), min_df=10, stop_words='english'), c-TF-IDF
Topic Post-processing	Topic reduction and outlier reassignment were not applied
Analysis Environment	Python 3, Google Colab

참고문헌

- [1] Xinhua. "Black Myth: Wukong" Becomes Smash Hit with over 10 Million Copies Sold [Internet]. Available: <https://english.news.cn/20240824/a351c8bb12074ea2b86c0862b53593a2/c.html>.
- [2] Q. Tang, X. Bai, and F. Gan, "On Sentiment Recognition Mechanism in Black Myth: Wukong Player Communication on YouTube," *Frontiers in Psychology*, Vol. 16, 1625671, October 2025. <https://doi.org/10.3389/fpsyg.2025.1625671>
- [3] H.-N. Kang, H.-R. Yong, and H.-S. Hwang, "A Study of Factors Influencing Helpfulness of Game Reviews: Analyzing STEAM Game Review Data," *Journal of Korea Game Society*, Vol. 17, No. 3, pp. 33-44, June 2017. <https://doi.org/10.7583/JKGS.2017.17.3.33>
- [4] M. Wui, J.-Y. Na, and Y. I. Park, "A Study on the Elements of Interest for VR Game Users Using Text Mining and Text Network Analysis: Focused on STEAM User Review Data," *Journal of Korea Game Society*, Vol. 18, No. 6, pp. 69-82, December 2018. <https://doi.org/10.7583/JKGS.2018.18.6.69>
- [5] T.-K. Kim and K. Kim, "Sentiment Classification of Game Review Data Using Topic Modeling," *Journal of Digital Contents Society*, Vol. 22, No. 9, pp. 1477-1486, September 2021. <https://doi.org/10.9728/dcs.2021.22.9.1477>
- [6] Y.-W. Choi and S. Ryu, "Key Words Research of Players' Experience and Presence in FPS Genre: Focusing on Game Play Time and Steam Reviews," *Journal of Korea Game Society*, Vol. 21, No. 6, pp. 13-30, December 2021.
- [7] A. Singh and J. O'Hagan, "Exploring Topic Modelling of User Reviews as a Monitoring Mechanism for Emergent Issues Within Social VR Communities," arXiv:2406.03994, June 2024. <https://doi.org/10.48550/arXiv.2406.03994>
- [8] S. Hall, Encoding/Decoding, in *Culture, Media, Language: Working Papers in Cultural Studies*, 1972-79, London, UK: Hutchinson, pp. 128-138, 1980.
- [9] K. C. Schröder, "Making Sense of Audience Discourses: Towards a Multidimensional Model of Mass Media Reception," *European Journal of Cultural Studies*, Vol. 3, No. 2, pp. 233-258, 2000. <https://doi.org/10.1177/136754940000300205>
- [10] A. Shaw, "What Is Video Game Culture? Cultural Studies and Game Studies," *Games and Culture*, Vol. 5, No. 4, pp. 403-424, 2010. <https://doi.org/10.1177/1555412009360414>
- [11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, 2003.
- [12] M. Röder, A. Both, and 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)*, Shanghai, China, pp. 399-408, February 2015. <https://doi.org/10.1145/2684822.2685324>
- [13] M. Grootendorst, "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure," arXiv:2203.05794, March 2022. <https://doi.org/10.48550/arXiv.2203.05794>
- [14] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, China, pp. 3982-3992, November 2019. <https://doi.org/10.18653/v1/D19-1410>
- [15] S. Kim and K. Yang, "Topic Model Augmentation and Extension Method Using LDA and BERTopic," *Journal of the Korean Society for Information Management*, Vol. 39, No. 3, pp. 99-132, September 2022. <https://doi.org/10.3743/KOSIM.2022.39.3.099>
- [16] F. Dehghani and L. Zaman, "Exploring Players' Perspectives: A Comprehensive Topic Modeling Case Study on Elden Ring," *Information*, Vol. 15, No. 9, 573,

September 2024. <https://doi.org/10.3390/info15090573>

- [17] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8, No. 1, pp. 216-225, May 2014. <https://doi.org/10.1609/icwsm.v8i1.14550>



임준형 (June-Hyeong Lim)

2026년 : 한국폴리텍II대학 인천캠퍼스
반도체공정과 2년제
학위과정 재학

2016년~2017년: 그랜드하얏트호텔 인천공항지점 근무

2018년~2020년: ㈜거산투어 근무

2020년: ㈜아이에스테크놀로지 근무

2024년: ㈜정임종합건설 근무

※관심분야: 데이터 분석, 디지털 인문학, 반도체 공정 등



배현영 (Hyun-Young Bae)

2024년 : 한국기술교육대학교

IT융합과학경영산업대학원

(공학석사 - AI융합교육학과)

2026년 : 서강대학교

가상융합전문대학원 박사과정

수료

2025년~현 재: 이엑스솔루션 수석연구원

2025년~현 재: 한국기술교육대학교

IT융합과학경영산업대학원 AI융합교육학과

겸임교수

※관심분야: 인공지능, 퍼지컬AI, 가상융합, 디지털트윈 등