

# 사전학습 모델 기반 앨범 아트와 가사 감정 구조의 이미지- 텍스트 멀티모달 분석

콘셉트 앨범을 중심으로

## Multimodal Image-Text Analysis of Album Art and Lyric Emotion Structures Using Pre-trained Models

focusing on the concept album

주 저 자 : 최서영 (Choi, Seo Young)      부산대학교 디자인학과 석사과정

교 신 저 자 : 김철기 (Kim, Cheol Ki)      부산대학교 디자인학과 교수  
ckkim@pusan.ac.kr

<https://doi.org/10.46248/kidrs.2025.4.476>

접수일 2025. 11. 19. / 심사완료일 2025. 12. 01. / 게재확정일 2025. 12. 08. / 게재일 2025. 12. 30.

이 논문은 부산대학교 우수대학원생 논문연구 지원 사업(2025)에 의하여 연구되었음

## Abstract

As the music industry has shifted to digital streaming, the influence of physical albums has declined, while album art remains a key visual and emotional interface on music platforms. This study applies the GoEmotion model to song lyrics to extract emotional data and formulates a similarity hypothesis between emotions and album art, which is tested using a pre-trained model and statistical analysis. The results show a strong alignment between the emotional characteristics of specific tracks and their album visuals, suggesting that album art can be designed to convey emotional narratives. By introducing quantitative analysis, this study complements predominantly qualitative album art research and supports future emotion-based visual design.

## Keyword

Emotion analysis(감정분석), Multimodal system(멀티모달 시스템), Album Art(앨범아트)

## 요약

최근 수년간 디지털 스트리밍 시대로 발전으로 음반 산업은 아날로그 CD 중심 물리적 매체 시대에서 스트리밍 기반 디지털 환경으로 전환되면서 실물 음반 영향력은 급격히 축소되었다. 그럼에도 앨범 아트는 여전히 시각적 매개체로서 중요한 역할을 담당하고 있으며, 다양한 음악 플랫폼은 디지털 앨범을 소개하는 핵심 인터페이스로 앨범 아트를 적극 활용하는 추세로 예술적·홍보적 장치로 진화하고 있음을 보여준다. 이에 본 연구는 앨범 아트의 정체성을 가사 텍스트에 GoEmotion 사전 학습 모델을 적용하여 감정 데이터로 재조명하였다. 그 후, 시각·감정 텍스트 매체의 연결성을 기반으로 유사도에 대한 가설을 설정하여 사전 학습 모델을 통해 유사성을 추출하고 통계 분석을 진행하였다. 그 결과, 특정 트랙의 감정과 앨범 아트의 유사도가 높음을 확인하였고, 이는 아트가 감정적 내러티브를 반영하는 방향으로 설계될 수 있음을 시사한다. 본 연구는 정성 중심으로 분석된 앨범 아트 연구에 정량적인 근거를 보완함으로써 향후 감정 기반 아트 수립의 기초 연구로 의의가 있다.

## 목차

### 1. 서론

- 1-1. 연구 배경
- 1-2. 연구 목적 및 방법

### 2. 이론적 배경

- 2-1. 앨범 커버 아트 분석
- 2-2. 텍스트-이미지 멀티모달 모델
- 2-3. 콘셉트 앨범 정의

### 3. 텍스트-이미지 멀티모달 분석

3-1. 데이터 수집 및 전처리

- 3-2. 감정 점수 및 감정 레이블 선정
- 3-3. Plutchik 기반 감정 매팅
- 3-4. CLIP 모델 기반 유사도 분석
- 3-5. BLIP 모델 기반 프롬프트 검증

### 4. 분석 결과

- 4-1. 통계 분석 결과
- 4-2. 시사점 및 논의점

### 5. 결론

#### 참고문헌

# 1. 서론

## 1-1. 연구 배경

최근 수년간, 기술의 발전으로 음반 산업은 아날로그 CD 시대에서 디지털, 스트리밍 중심의 디지털 시대로 부상하면서 물리 매체가 차지하는 영향력이 미미해졌다. 다양한 음악 플랫폼의 발전으로 가수의 팬이 아닌 대중 또한 대부분은 스트리밍 서비스를 이용하여 노래를 듣고 아티스트는 실물 CD 대신 디지털 음반을 발매하는 추세이다.<sup>1)</sup>

음악을 접하는 방식이 비닐 레코드 블록에서 디지털 장치의 작은 셀프일로의 극적인 변화에도 불구하고 앨범 아트는 여전히 큰 영향력을 가진다. 앨범 아트는 단순히 형식과 매체의 역할을 넘어 디지털 시대와 함께 진화하면서 홍보성에 더불어 예술성을 갖춘 수단으로 여겨진다. 디지털 시대에서 스포타이나 애플뮤직, 멜론, 벅스 등 여러 음악 플랫폼에서는 디지털 앨범을 전시하는데 앨범 아트를 적극적으로 사용하고 있다.<sup>2)</sup>

오늘날 음반 시장은 포화 상태로, 아티스트들은 기존 팬들뿐만 아니라 잠재적인 팬들의 시선을 사로잡아야 하고 이 몇 초의 기회는 앨범 아트가 제공한다. 앨범 아트는 청취자와 아티스트 간의 매개체로써 시각적인 분위기, 장르 그리고 아티스트의 정체성까지 포괄적으로 담아 전달해야 함을 의미한다. 즉 앨범 아트는 음반 전체 그리고 아티스트의 정체성으로 청취자들의 첫 시선을 끄는데 필수적인 관문 중 하나로 여겨진다.<sup>3)</sup> 대부분 앨범 아트는 디자이너, 일러스트레이터를 통해 제작되지만 아티스트 역시 본인의 추구성을 반영하여 소비자의 흥미를 유발해 유지하기 위해 시각적 표현

및 전시에 직접 참여하는 경우도 다수 존재한다.<sup>4)</sup> 음악을 접할 때 가장 먼저 인지하는 시각적인 인터페이스로써 앨범 아트는 음악계의 표지 역할로 시청각적 해석을 불러일으키고 새로운 음악을 소개하는 데 최적화된 방법으로 사용된다.<sup>5)</sup> 이런 앨범 아트는 청취자가 노래를 선택하는 데 있어 가장 중요한 요소로 음반 판매 마케팅과도 직결 적이기에 노래 가사들의 분위기나 장르, 특성을 포괄한 감정을 반영하여 앨범 아트로 녹여내는 것이 필수적이다.

이렇듯 앨범 아트의 영향력은 음반 시장에서 여전히 상장성을 가지며 본 연구에서는 국내 앨범을 기준으로 앨범 아트와 노래의 가사 속에 담긴 감정의 연관성을 정량적으로 분석하고자 한다. 특히 일반 앨범이 아닌 콘셉트 앨범을 선정해 앨범 내 트랙들의 서사적인 감정 구조를 도출하고 이를 기반으로 앨범 아트와의 연관 정도를 파악하면서 현재의 앨범 아트의 현황을 살펴보고, 향후 개선 방향에 관한 사전 연구로 목표로 한다.

## 1-2. 연구 목적 및 방법

본 연구는 앨범 내 트랙들의 가사 텍스트에 대한 감정 점수 도출 및 레이블 생성, 가설 설정 및 가설 하에서 생성된 레이블을 기반으로 이미지-텍스트 멀티모달 분석 시스템 적용, 가설의 통계적 분석 단계로 진행한다.

먼저 이미지-텍스트 멀티모달 모델에서 앨범 아트와 매칭시킬 가사의 세밀한 감정 텍스트를 뽑아내기 위해 GoEmotion 모델을 활용한다. 이 모델은 Google Research에서 2020년도에 공개한 대규모 감정 분류 데이터셋 GoEmotion으로 Reddit 댓글 58,000개에 대해 28개의 감정 레이블로 구성된 데이터셋을 사전 훈련 시킨 모델이다. 기존 Ekman 6가지 감정이나 기존 이분법적 감정 분류(긍정, 부정)와 달리 해당 모델은 보다 세분화된 감정 포착이 가능하며 이는 노래의 내포된 복합적인 감정을 도출하기에 적합하다. 해당 연구는 국내 앨범을 대상으로 하기에 GoEmotion 데이터셋을 한국어로 번역하여 훈련한 GoEmotion-Korean 모델을 선정하여 진행한다.<sup>6)</sup> 이 모델을 통해 본 연구

- 1) Jim Carroll, 'The art of the sleeve: every album cover tells a story', The Irish Times, 2016.10.13. (2025.10.26.) [www.irishtimes.com/culture/music/the-art-of-the-sleeve-every-album-cover-tells-a-story-1.2821084](http://www.irishtimes.com/culture/music/the-art-of-the-sleeve-every-album-cover-tells-a-story-1.2821084)
- 2) Tara Venkatesan, Qian Janice Wang, & Charles Spence, 'Does the typeface on album cover influence expectations and perception of music?', Psychology of Aesthetics, Creativity and the Arts, 2022, Vol.16, No.3, pp.487–503.
- 3) Impress Vinyl, 'Why Music Album Covers Are Still important In The Digital Age', Vinyl Record Covers Blog, 2025.04.15. (2025.10.26.) [vinylpressing.com.au/blog/vinyl-record-covers/why-music-album-covers-are-still-important-in-the-digital-age/](http://vinylpressing.com.au/blog/vinyl-record-covers/why-music-album-covers-are-still-important-in-the-digital-age/)

- 4) 박민희, 이은주, '인디 앨범 섬네일 디자인의 시각적 특성에 관한 연구', 일러스트레이션 포럼, 2022, Vol.23, No.73, p.93
- 5) 박주향, '파라텍스트로써의 앨범커버 연구 –비틀즈의 앨범커버를 중심으로–', 인문사회예술융합, 2023, Vol.45, No.10, p.399

는 앨범 내 트랙별 가사에 대한 대표 상위 감정 3개를 뽑아 레이블로 생성한다. 더 나아가 이 3개의 감정을 Plutchik 좌표계에 투영하여 최종적으로 1개의 대표 레이블을 도출하고 이를 활용해 감정 흐름과 최종 레이블을 생성한다. 앨범 커버 아트와 감정 간의 유사도 분석을 위해 앞서 생성한 레이블 기반으로 감정 텍스트 프롬프트를 생성하고 이미자-텍스트 멀티모달 모델을 통해 시각적 요소로써 감정을 대표하는지에 대해 다방면적인 관점에서 의미적 상호작용을 텁색하고 통계적 분석을 진행한다.

## 2. 이론적 배경

### 2-1. 앨범 커버 아트 분석

앨범 커버 아트에 관한 연구는 최근까지도 커버 아트의 중요성과 영향력이 언급되며 정성적, 정량적인 관점에서 국내외적으로 활발하게 이루어지고 있다.

정성적인 관점에서는 주로 색상, 질감, 그래픽 요소 등 시각적인 요소를 기준으로 앨범 아트의 시각적 효과를 측정하였다. 록 음악의 앨범 아트를 대상으로 커버 디자인을 단순한 홍보 수단을 넘어 예술 작품으로 바라보는 관점에서 미학적 매력을 가지는 매체로써 정성적인 분석을 진행하였다. 전달력, 심미성, 독창성, 주목성, 색상 표현을 기준으로 한 5점 척도 설문 조사를 통해 시각적 효과에 따른 호감도를 측정하는 사례분석을 바탕으로 앨범 아트가 구매 의향에 미치는 영향과 음악적 메시지 전달의 중요성을 강조하였다.<sup>7)</sup>

최근에는 딥러닝 기술의 발전으로 정성적인 기준을 잡고 정량적으로 접근을 통해 아트와 마케팅, 소비적 연관도를 살펴보았다. 아트 요소 중 선의 방향성과 아트의 선명도를 기준으로 한 정량적 분석에서는 심리학에서 수직 및 수평선의 영향력을 언급하며 3가지 가설 설정을 통해 진행하였다. 아트의 선명도와 음반 시장의 성과의 영향력, 아트 속 선의 방향과 음악 소비 인과관계, 여러 조건에서 선의 방향과 소비자 청취 경험의 영향력을 검증하여 선명도가 높을수록, 선/가장자리 방향이 있는 앨범이 더 청취를 유도할 것을 예상하였다.<sup>8)</sup>

6) monologg, “GoEmotions-Korean”, GitHub repository, (2025.10.25.)  
github.com/monologg/GoEmotions-Korean

7) 주효선, 송민정, ‘록 음악 앨범 커버의 시각적 디자인 요소 특징 연구’, 브랜드디자인학, 2021.12, Vol.19, No.4, pp.264–269

또한 국내의 경우, K-Pop 아이돌 앨범 아트를 대상으로 앨범 커버의 색채 분석을 통해 세대, 디자인 유형별 색채의 트렌드를 확인하였다. KSCA로 커버 아트의 주조 색상을 얻어 명도, 채도 기반으로 색상 분포를 확인하고 2세대는 중·저명도, 3세대는 중·고명도, 3.5세대는 주로 중·고채도를 사용함을 결론 내리며 세대별 사례 분석을 선, 선명도 요소를 통해 진행하였다.<sup>9)</sup> 이렇듯 특정 시각적 요소나 특정 장르의 앨범 아트를 기반으로 경향성을 측정하거나 세대별, 디자인 유형별로 사례 분석 위주로 분석이 이루어지고 있다. 모델을 활용하여 분석한 아트 연구의 경우는 주로 장르와 연관지어 정량적인 관점에서 진행되었다. 18개의 장르를 기준으로 앨범 데이터를 수집하고 아트에 따라 모델이 해당 장르를 분류하도록 훈련하며 아트가 음악 추천 알고리즘에서 활용할 수 있음을 실용적인 관점에서 제시하였다.<sup>10)</sup>

이처럼 정성적 연구들은 주로 색상, 질감, 선 등 시각적 요소의 미학성과 아트의 메시지 전달력을 평가하는 데 집중하였고 정량적 연구 역시 선명도나 선의 방향 등 특정 요소를 수치화해 소비 성과와의 관계를 분석하였다. 또한 모델 기반 아트 분석은 앨범 아트를 장르 분류나 추천 알고리즘과 같은 실용적, 기능적 용도에 활용하려는 경향이 강해 이미지가 내포한 서사적 의미를 해석하려는 시도는 미비하다. 이러한 흐름 속에서 본 연구는 앨범 아트와 가사 텍스트의 감정 간의 직접적인 연관성을 유사도를 기반으로 정량적으로 분석함으로써, 음악의 정서적 정보가 시각적 아트에 어떻게 반영되는지를 정량적으로 규명하고자 한다. 즉, 앨범 아트의 시각적 특징을 소비, 장르 중심의 해석을 넘어 가사의 감정과 시각적 연결 구조를 분석하는 새로운 관점을 제시하고자 한다.

### 2-2. 이미자-텍스트 멀티모달 시스템

최근 인공지능 기반의 시각 및 언어의 융합 연구가

8) Joye, Y., & Fennis, B. M, 'Set that record straight! Cardinal line orientations in music album artwork boost market performance and music consumption', Psychology&Marketing, 2024, pp.918–920

9) 황지선, 이윤진, ‘파이썬을 활용한 데이터 크롤링과 유형화에 따른 국내 아이돌 앨범 커버 색채 분석’, 한국색채학회, 2022.05, pp.31–33

10) Natalie Greenfield & Ngorli Paintsil, Album Covers Deserve Some Attention, Stanford University, 2024, pp.2–6.

활발하게 이루어지면서 이미지와 텍스트를 동시에 이해하는 멀티모달(Multimodal) 시스템이 다양한 분야에서 중요 도구로 부상하고 있다. 이런 시스템은 이미지에서 시각적인 피처 값을 추출하고 텍스트를 활용해 의미적 맥락을 연결함으로써 시각 자료가 담고 있는 메시지를 보다 정교하게 해석을 가능하게 한다. 대표적으로 CLIP(Contrastive Language-Image Pre-training)은 이미지와 텍스트를 동일 의미 공간에 매핑하여 공간을 학습함으로써 두 모달 간의 연관성 분석을 가능하게 하였고 여러 모달이나 분야(Downstream Task)에서 추가 학습 없이 적용하는 제로 학습이 가능하다.<sup>11)</sup> CLIP과 여러 멀티모달 모델의 발전으로 분야에 맞게 도메인 적응(Fine-tuning)하여 다양한 연구가 진행되고 있다.

특히 기존 감정 분석에서는 이미지나 텍스트 단일 모달리티에 집중되어 감정 정보의 좁고 불완전한 스펙트럼만 포착하였으나 다중 모달 시스템을 활용해 여러 모달리티에서 추출된 피처 값들로 연관도 분석과 특징 수준에서의 상호작용이 가능하였다. 특히 FAFMEAC(Feature -Attentive Multimodal Emotion Analyzer with CLIP)은 이미지 캡션 감정 레이블로(긍정·부정·중립) 분류된 이미지 데이터셋을 활용하여 CLIP 모델의 임베딩 공간을 시각적, 언어적 표현을 효율적으로 정렬하고 의미적 피처를 추출한다. 추가로 각 모달리티 인코더에서 개별 추출된 특징과 결합한 뒤, 어텐션 메커니즘과 커스텀 레이어를 적용하여 모달 간의 상호작용을 통해 더 정교화된 감정 다중 분류 학습이 가능하였다. 즉 CLIP 모델의 공간 특징을 활용하되, 추가적 파인튜닝 과정을 통해 정교한 다중분류의 가능성을 탐구하였다.<sup>12)</sup> 제로샷 학습의 일반화 검증 연구에서는 이미지의 품질뿐만 아니라 주관적인 느낌을 갖는 이미지를 평가하기 위해 CLIP을 활용한 결과, 기존 이미지 평가 체계를 넘어 정서적 특성을 해석할 수 있는지를 분석하였다.<sup>13)</sup> 또한 추상 미술 작품을 대상으

로도 CLIP을 활용하여 작품에서 유발되는 감정을 분류하였고, baseline보다는 유의미한 높은 정확도를 보였으며 이는 CLIP이 미술 작품에서도 감정 상태의 해석 가능성을 염불 수 있었다.<sup>14)</sup> 이를 통해 CLIP이 추가적인 학습 없이도 어느 정도 감정적인 피처를 인지함을 확인할 수 있었다.

멀티모달 모델, 특히 CLIP은 이미지-텍스트 임베딩 공간 구축을 통해 두 모달의 의미적 연관성 분석을 가능하게 하였다. 기존 연구는 일상 이미지 데이터셋을 기반으로 모델을 활용하거나 앨범 아트의 경우는 생성 중심의 활용으로 한정되어 있으며 앨범 아트와 감정의 연관도 기반 정량적 해석은 미비하다. 이에 맞춰 본 연구는 CLIP 멀티모달 공간을 활용해 앨범 아트와 앨범 내 가사에 내포된 감정의 관련도를 정량적으로 보완하여 기존 디자인 전략 체계에 기초적 근거를 제공하고자 한다.

### 2-3. 콘셉트 앨범 정의

콘셉트 앨범은 일반 음반과는 달리 각 트랙이 하나의 서사 또는 음악적인 구조로 연결되어 전체적으로 통일된 주제를 다루는 앨범이다. 콘셉트 앨범은 과거 레코드판이나 테이프를 통해 음반을 접하던 시대에서 끊기지 않고 마치 중간 광고가 존재하는 드라마처럼 내용을 담기 위해 탄생한 앨범이다. 이러한 구조는 멜로디, 리듬, 악기 구성, 가사 등 다양한 음악 요소를 통해 포괄적·서사적 구조로 연결될 수 있다. 대중음악 역사상 최초의 콘셉트 앨범은 프랭크 시나트라의 “In the Wee Small Hours(1995)”가 자주 언급되며 이후 비틀즈의 “Sgt. Pepper’s Lonely Hearts Club Band”를 통해 콘셉트 앨범은 대중적인 개념으로 자리 잡았다. 다만 이는 학자들의 다양한 관점에 따라 콘셉트 앨범의 기준이 방대하고 이에 따른 정의들 또한 매우 다양하기 때문에 현대 시대까지 여전히 모호하게 다루어지는 개념이다.<sup>15)</sup>

11) Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever, ‘Learning Transferable Visual Models From Natural Language Supervision’, Proceedings of ICML 2021, pp.8748–8751.  
arxiv.org/abs/2103.00020

12) Zichao Nie, ‘Feature-Attentive Multimodal Emotion Analyzer with CLIP’, ICICML, 2023, pp.318–320

13) Jianyi Wang, Kelvin C.K. Chan, Chen Change

Loy, ‘Exploring CLIP for Assessing the Look and Feel of Images’, arXiv preprint, 2022, pp.2557–2558  
arxiv.org/abs/2207.12396

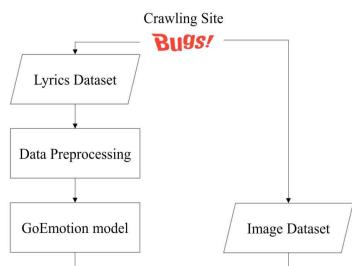
14) Hanna-Sophia Widholz, Ece Takmaz, ‘Decoding Emotions in Abstract Art: Cognitive Plausibility of CLIP in Recognizing Color–Emotion Associations’, arXiv Preprint, 2024, pp.3–4  
arxiv.org/abs/2405.06319

15) Sorensen, Paige C., ‘The Concept Album

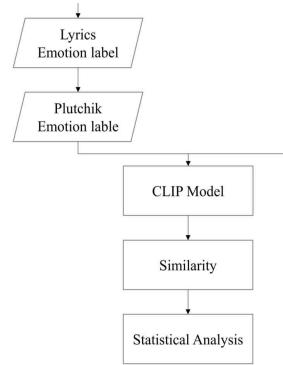
그럼에도 꾸준하게 언급될 만큼 음악의 역사에서 중요한 개념이고 특히 힙합이나 인디 장르에서 아티스트들은 여전히 서사적 구조를 반영해 트랙 간의 순서 배치를 신중히 고려하며 음반을 제작한다. 앨범 전체가 하나의 예술 작품으로써 의미를 가지고 서사적 구조를 통해 심층적인 경험을 제공하면서도 청취자의 공감을 불러일으킨다. 이러한 콘셉트 앨범의 정체성은 곧 가사 텍스트가 구성하는 스토리 전개와 감정 흐름에 의해 전개되며, 앞서 언급했듯 앨범 아트가 앨범의 정체성을 시각적으로 포괄해 보여주는 역할을 한다는 점은 가사 텍스트가 지닌 서사의 감정은 아트가 반영해야 할 핵심적인 내용이 될 수 있다. 이에 따라 본 연구는 콘셉트 앨범 내 트랙들의 가사 텍스트들에서 도출된 감정 궤적이 아트와 연관 있는 패턴을 보일 것이고, 해당 패턴이 앨범 아트에 얼마나 반영되어 있는지를 확인하는데 분석의 초점을 둔다. 나무위키, 위키피디아에 명시된 앨범과 아티스트의 인터뷰에서 앨범을 ‘서사적 구조’라고 언급한 국내 콘셉트 앨범을 기준으로 총 48개를 선정하여 분석을 진행한다. 콘셉트 앨범의 기준은 멜로디, 리듬 등 다양한 음악적 구성요소가 있으나 본 연구는 감정과 가장 직결된 가사를 기반으로 한다. 선정된 앨범의 장르는 주로 본인의 삶에 관해 이야기하는 힙합에 많이 분포되어 있으며 밸러드는 주로 옛날 가수의 앨범으로 분포되어 있다. 앨범 아트와 앨범의 감정의 관계 분석은 정성적 연구에서도 정량적인 뒷받침이 될 수 있는 시사점을 제공할 것으로 판단하여 콘셉트 앨범을 기준으로 분석을 진행한다.

### 3. 이미지-텍스트 멀티모달 분석

본 절에서는 이미지-텍스트 멀티모달 분석을 위한 파이프라인을 단계적으로 다음과 같이 소개한다.



Continuum', Honors Senior Capstone Projects. 43, 2019, pp.7-15



[그림 1] 분석 플로우차트

#### 3-1. 데이터 수집 및 전처리

콘셉트 앨범은 벅스 뮤직 사이트에서 크롤링을 통해 힙합, 인디/포크, 밸러드, 락/메탈, 댄스, 알앤비 장르에서 앨범 아트 48개의 이미지와 48개 내 앨범 곡들 총 528곡을 수집하였다.

[표 1] 장르별 콘셉트 앨범

장르	앨범명	곡 수
힙합	1Q87, LUCKYMAN, FOUNDER, TO DIE FOR, WONDERLAND, The movie star, 99, 500000, 선인장화, 송정맨션, The Aneodote, MODM, MODM2, MAINSTREAM, 2MANYHOMES, SOUR, CC, EAT, Moonshine, 'Not' Based On The True Story	221
인디 / 포크	daydream, Garion2, TEAMBABY, TEENTROUBLES, HardboiledCafe	73
밸러드	A Night in Seoul, ENDTHEORY, Cycle, The Dreams	49
댄스	Present, Hawwah, Moments in between, 서울여자, GROWTH THEORY, The Golden Age, Ilove, ERROR, MONKEYHOTEL	80
락 / 메탈	Issue, THIRSTY, Escaping Gravity, Holding Onto Gravity, Newton a Apple, Atomos	62
알앤비	130 mood, WATERFALL, Q, The Sandwich Artist	43

전처리 과정은 크게 3단계로, 가사 데이터를 대상으로 다음과 같이 진행하였다.

먼저 가사 텍스트의 정규화를 위해 \*, #, &, \$ 등 특수문자 처리 및 연속된 공백을 단일 공백으로 통합하는 과정을 진행하였다.

그 후, 한국어와 영문이 섞인 가사에 대해 영어로 이루어진 가사만 Google Translate API를 활용하여 한국어로 번역 과정을 거쳤다.

마지막으로 다음 단계에서 사전 모델을 통해 감정 추출 시, 가사 텍스트에서 감정 도출의 세밀도를 높이기 위해 전체 가사를 의미 단위의 구절로 분리하는 과정을 거쳤다. 가사의 특성을 고려하여 통계 기반 문장 분리기 KSS(Korean Sentence Splitter)와 형태소 분석기 Kiwi를 결합한 하이브리드 접근법을 이용하였다. KSS를 활용해 문장 경계를 1차적으로 식별한 후, 노래의 언어학적 특성을 고려하여 90자 이상 긴 문장만 Kiwi 분석기로 종결어미(EF, 예: “해요”, “네요”) 위치를 탐지한 결과를 기반으로 우선 분할하고, 추가로 연결어미(EC, 예: “고”, “며”) 중 문장 중반 이후에 있는 것을 기준으로 2차 분할을 수행함으로써 미세한 경계점을 보완하였다. 결과는 리스트 형태로 저장하여 다음 분석에서 활용하도록 열로 추가하였다.

분리된 구절	
0	「내 이름은 괴물 내 이름은 죄인 내 이름은 위선자 내 이름은 폐안」, ‘봄에 괴물..
1	「모 해병미 있어 나의 웃소매 끝에 두 헤드 흐르는 물줄기를 닦아낸 뒷에 모 해병..
2	『길고 긴 꿈을 꾸었나 모든 게 낯설기만 해 너무 많은 갑협을 네 옆에 두고 왔니..
3	「기분이 너무 무감각해 기분이 너무 무감각해 난」, ‘사람이었지 와로운 가득한 얼..
4	「내 민생 내 민생 꿈결 얼렁뚱말이며 일렁며 물결 흔들리 흔들리」, ‘술걸 기분은 기분..

[그림 2] 구절 리스트

### 3-2. 감정 점수 및 감정 레이블 선정

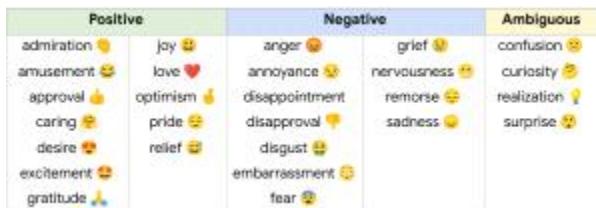
국내 노래 데이터셋은 가사에 대한 감정 레이블이나 점수 등 레이블링 된 데이터셋의 부재로 본 연구에서는 가사 형태에 적합한 구절 분리와 집계 방식을 거쳐 사전 학습된 모델로 감정과 점수를 계산하고 레이블링 하여 분석 데이터를 마련하였다. 사전 훈련된 모델 GoEmotion은 텍스트 임베딩 값을 학습하여 예측한 점수로 사전 정의된 28가지 감정들을 텍스트값이 얼마나 유사하나에 따라 0에서 1 사이 점수를 할당하고 1에 가까울수록 해당 감정과 유사함을 의미한다. 이를 활용해 앞 단계에서 저장한 구절 리스트마다 감정과 점수를 할당하며 한 노래의 가사는 보통 1~15개의 구절로 구성되어 구절마다 여러 감정과 점수를 가지게 된다. 추출된 감정과 점수를 곡 단위로 통합하기 위해 Top-3 상위 평균 집계법을 제안한다. 이 방법에서는 구절마다 할당된 28개의 감정에 대한 확률값 중 가장 높은 상위 3개의 감정만을 선택한다. 전체 구절을 순회하면서 선정된 감정이 몇 개의 구절에서 Top-3에 포함되었는지를 카운트하고 이를 활용해 출현 빈도 기반

평균으로 감정마다 확률값을 계산한다. 예시로 ‘슬픔’ 감정이 9개의 구절 중 4개의 구절 상위 감정으로 포함되어 있다면 해당 4개의 슬픔 확률 점수를 활용해 등장 횟수를 활용한 슬픔 감정의 최종 평균 점수를 도출한다. 이로써 확률값을 기준으로 노래마다 총 3개의 실제 가사의 감정과 가까운 의미 있는 감정만을 추출하여 낮은 확률의 노이즈 감정을 배제하고 실제 두드러지게 나타난 감정만을 고려하며 노래 전체의 정서적 특성을 적합하게 반영할 수 있다.

TOP3 감정_1013	
0	[(‘gratitude’, 0.9902), (‘approval’, 0.854), (...]
1	[(‘embarrassment’, 0.8747), (‘sadness’, 0.5904), (...]
2	[(‘sadness’, 0.7165), (‘neutral’, 0.7031), (...]
3	[(‘amusement’, 0.7712), (‘approval’, 0.7366), (...]
4	[(‘neutral’, 0.9314), (‘surprise’, 0.77145), (...]

[그림 3] 감정 및 점수 추출

### 3-3. Plutchik 감정 좌표 매팅



[그림 4] GoEmotions taxonomy: Includes 28 emotion categories, including “neutral”.<sup>16)</sup>

본 단계에서는 앞 단계에서 추출된 가사별 최상위 3개의 감정을 하나의 감정으로 선정하기 위한 기준을 정립한다. 단일 최상위 감정만을 활용하면 다음과 같은 한계를 가진다.

먼저 실제 노래의 가사가 내포하는 감정은 복합적인 상태로 caring, sadness, disappointment처럼 여러 감정이 혼재된 형태이며 이를 최상위 단일 감정만으로 축약하면 이중적인 상태가 소실되어 감정 간의 상호작용이나 인과관계를 무시할 수 있다. 예를 들어 joy와 trust의 감정에 의해 최상위 감정 love가 유발되면 joy

16) research.google, ‘GoEmotions: A Dataset for Fine-Grained Emotion Classification’, 2021.(2025. 11. 11.)  
research.google/blog/goemotions-a-dataset-for-fine-grained-emotion-classification/

와 trust의 감정을 무시한 채 love만으로 가사에 대한 감정을 정의 내리게 된다. 최상위 감정만으로는 이런 혼합성을 포착하지 못하여 편향된 결과를 보일 수 있다. 이에 따라 본 연구는 추출된 Top-3 감정을 Plutchik 감정 바퀴 모델의 좌표계를 정의하고 투영하여 복합적인 감정을 포괄하는 하나의 감정으로 레이블링하고자 한다.

[표 2] Plutchik 1차 감정 매핑

Plutchik	각도	GoEmotion 감정
joy	0°	joy, amusement, excitement, gratitude, pride, relief
trust	45°	approval, caring, admiration
fear	90°	fear, nervousness
surprise	135°	surprise, realization, confusion
sadness	180°	sadness, grief
disgust	225°	disgust, disapproval, embarrassment
anger	270°	anger, annoyance
anticipation	315°	anticipation, curiosity, desire

[표 3] Plutchik 2차 감정

Plutchik	각도	생성
love	22.5°	joy+trust 영역
submission	67.5°	trust+fear 영역
alarm	112.5°	fear+surprise 영역
disappointment	157.5°	surprise+sadness 영역
remorse	202.5°	sadness+disgust 영역
contempt	247.5°	disgust+anger 영역
aggressiveness	292.5°	anger+anticipation 영역
optimism	337.5°	anticipation+joy 영역

Plutchik 감정 바퀴<sup>17)</sup>는 8개의 기본 감정(분노, 공포, 슬픔, 혐오, 놀람, 기대, 신뢰, 기쁨)이 적대 관계, 인접 관계, 강도 단계, 혼합 단계의 계층적인 구조가 반영되어 동일 간격으로 각 감정을 배치해 혼합된 인접 감정이 2차 감정을 형성함을 제시한다. 동일 간격의 배치로 감정 간의 상대적 거리가 정의될 수 있고 각 감정은 고유한 방향성을 가지는 벡터로 표현되어 개별 감정을 극좌표계에서 벡터화하여 조합이 가능하다.<sup>18)</sup>

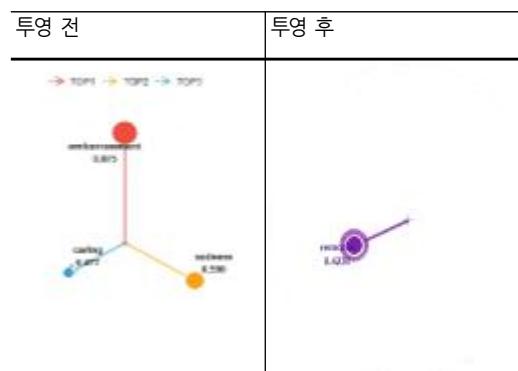
17) Robert Plutchik, 「The Nature of Emotion」, 2001, pp.347~350.

18) Alfonso Semeraro, Salvatore Vilella, Giancarlo Ruffo, 'PyPlutchik: Visualising and comparing

본 연구는 이런 구조를 활용해 8개의 기본 감정에 360도를 동일한 분할 각도로 부여하고 Top-3 세부 감정 레이블을 Plutchik 기본 감정과 의미상으로 대응하는 축의 각도로 매핑하였다. 세부 감정의 매핑은 가중 무게 중심 계산 시 3개의 감정 벡터가 합성되면서 기본 감정 사이의 Plutchik 2차 감정 영역으로 수렴할 수 있으며, 이를 통해 단일 감정만으로 표현할 수 없는 복합 감정 상태 포착이 가능하다. 감정에 대한 확률값은 벡터 크기(magnitude)로 변환되며 상위 3개의 감정은 극좌표계에서 벡터 합을 통해 단일 대표 감정으로 산출되어 최종 트랙의 감정으로 저장된다.

예를 들어 embarrassment(225°), sadness(180°), caring(45°)일 때, 각 감정은 서로 다른 축에 정의되어 다른 방향을 가지지만 감정 점수 레이블 확률값에 따라 확률 점수가 클수록 해당 감정에 매핑된 축으로 끌어 당겨져 최종 벡터는 각 감정 벡터의 크기(감정 점수)와 방향이 반영되어 점수가 가장 큰 disgust 영역(embarassment: 0.875)와 다음으로 큰 점수인 sadness 영역(sadness: 0.59)이 혼합된 remorse 감정 위치에 일치하게 수렴한다.

[표 4] Plutchik 투영 전후



이런 방식은 Plutchik의 감정 바퀴의 감정 간 구조적 관계를 유지하면서도 노래의 다중 감정을 복합적으로 감정 확률값과 축을 고려해서 하나의 감정으로 해석하는 데 적합하다.

### 3-4. CLIP 모델 기반 유사도 분석

본 절은 앨범 아트와 가사의 감정과 유사도 분석을 위해 이미지-텍스트 멀티모달 모델인 CLIP을 활용한다.

CLIP으로 이미지와 텍스트 간 관계 분석을 위해서는 텍스트 프롬프트 설계가 필수적이며 이에 따라 본 연구는 다음과 같이 3개의 가설 프레임워크를 설계하였다.

가설 1 (H1): 앨범 내 트랙의 감정 흐름이 앨범 아트와 연관성 있을 것이다.

가설 2 (H2): 앨범 내 타이틀 감정이 앨범 아트와 연관성 있을 것이다.

가설 3 (H3): 앨범 트랙 중 특정 감정이 앨범 아트와 연관성 있을 것이다.

CLIP 텍스트 프롬프트는 앞 단계에서 Plutchik 감정 투영을 통해 생성된 감정을 활용하여 생성한다. CLIP의 유사도 계산은 앨범 가사에 대해서 생성된 감정 텍스트 프롬프트와 앨범 아트가 얼마나 유사한가에 대해 Softmax 함수를 통해 상대적인 확률 분포로 변환된다. 유사도 점수는 코사인 유사도를 기반으로 스케일링 된 값으로 0에서 1 범위를 가지며, 1에 가까울수록 유사도가 높음을 의미한다. 단일 프롬프트 설계 시 해당 감정에 대한 절대적인 유사도만을 제공하여 명확한 비교를 할 수 없으므로 각 가설을 기반으로 다중 감정 프롬프트를 설계하여 가설을 정량적으로 설명하였다. 이에 따라 감정 레이블 열을 기준으로 “[Emotion] album” 형식으로 프롬프트를 설계하였다.

[표 5] 가설 별 프롬프트 형식

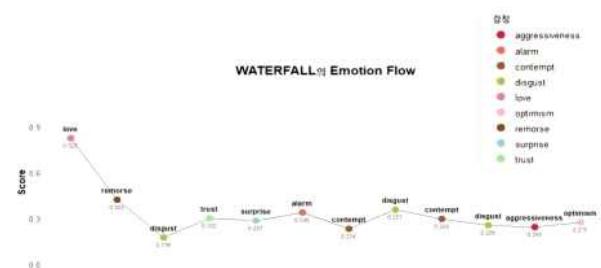
가설	트랙 감정 형식	프롬프트 형식
H1	Start Track Emotion	“{Emotion} album”
	Median Track Emotion	“{Emotion} album”
	End Track Emotion	“{Emotion} album”
	Album Flow Emotion	“The album flow is {Start Emotion}, {Median Emotion} and {End Emotion}.”
H2	타이틀 Track Emotion	“{Emotion} album”
	... 수록 N Track Emotion	...
	수록 1 Track Emotion	“{Emotion} album”
H3	... 수록 N Track Emotion	...
	“{Emotion} album”	

감정 텍스트 프롬프트는 해당 앨범 아트 이미지와 CLIP 공동 임베딩 공간에 정렬되어 유사도 벡터로 계산되며 각 유사도 점수는 가설마다 최상위 점수를 기준으로 [‘h1\_score’, ‘h2\_score’, ‘h3\_score’] 열에 저장하였다. 유사도 점수는 상대적인 확률값으로 비교를 진행하였으며, 1에 가까울수록 높은 유사도를 의미한다.

	album_name	h1_score	h2_score	h3_score
0	ERROR	0.638354	0.098816	0.388534
1	TEEN TROUBLES	0.601846	0.007842	0.657226
2	END THEORY	0.862011	0.272293	0.272293
3	THIRSTY	0.803303	0.094736	0.362680
4	daydream	0.556251	0.007543	0.425367

[그림 5] 가설별 유사도 점수

### 3-5. BLIP 모델 기반 프롬프트 검증



[그림 6] 앨범 내 트랙별 감정 흐름 및 점수

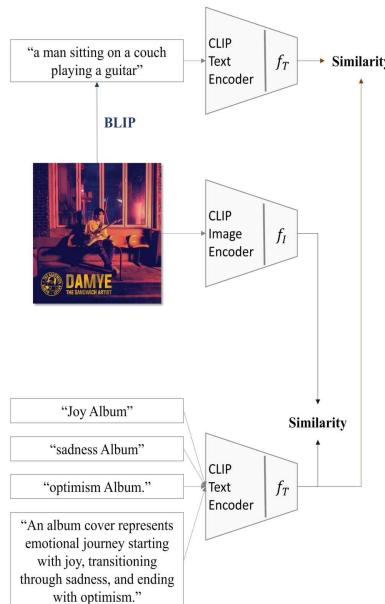
감정 텍스트 프롬프트에 입력된 감정은 GoEmotion 모델과 Plutchik 감정 바퀴를 통해 정량적으로 얻은 값으로 레이블의 불확실성이 존재한다. 이에 따라 본 단계에서는 BLIP(Bootstrapping Language-Image Pretraining) 모델을 보조적 도구로 감정 텍스트 프롬프트에 대한 신뢰성을 검증한다.

BLIP은 CLIP처럼 이미지와 텍스트를 공동 의미 공간으로 정렬하는 모델로, 공간 학습뿐만 아니라 모델이 직접 생성한 캡션을 학습에 재활용하는 텍스트 생성 성능을 갖춘 부트스트래핑 방식을 활용하여 더 정교한 이미지 설명이 가능하다.<sup>19)</sup> 즉 시각 정보를 언어 정보로 확장하는 기술로 본 단계에서는 앨범 아트의 설명 캡션을 생성하여 감정 텍스트 프롬프트와 캡션의 의미적 유사도를 측정하고 감정 레이블의 신뢰도를 확인하는 보조적 검증 장치로 활용하였다.

기존 CLIP 유사도 분석에 사용한 텍스트 프롬프트와 BLIP을 통해 생성된 아트의 캡션 유사도를 비교한다.

19) Junnan Li, Dongxu Li, Caiming Xiong, Steven Hoi, ‘BLIP: Bootstrapping Language–Image Pre-training for Unified Vision–Language Understanding and Generation’, ICML, 2022, pp.12888–12890

가설 각각에 대해 앨범 단위로 감정 프롬프트를 비교하고 분석한 결과, 대부분 앨범 가설에 대한 유사도가 0.6 이상의 결과로 모델로 생성된 감정 프롬프트가 타당함을 입증하였다.



[그림 7] BLIP-텍스트 프롬프트 유사도 예시

## 4. 분석 결과

### 4-1. 통계 분석 결과

본 단계는 지정한 가설들의 타당성을 증명하기 위해 통계 분석을 진행하였다. 본 연구의 데이터는 일반 앨범이 아닌 국내 콘셉트 앨범을 중심으로 수집하였기에 대부분 장르가 힙합(64%)으로 수집되었고 이런 불균형은 콘셉트 앨범의 성격 자체가 특정 장르에 편향되었기 때문이다. 콘셉트 앨범은 본인의 이야기를 음악으로 풀어내는 힙합 문화에서의 서사와 감정 흐름을 중시하는 전통과 밀접하게 관련되었기 때문에 힙합 장르에 대부분 분포되었다. 이에 따라 본 연구는 장르 구분 없이 전체 데이터(N=48)를 대상으로 정의된 가설(H1, H2, H3)의 유사도 점수를 기반하여 통계 분석을 진행하였다.

[표 6] 기술통계

Hypothesis	H1	H2	H3
------------	----	----	----

N	48	47	48
M	0.641	0.075	0.485
SD	0.207	0.096	0.190
Median	0.648	0.029	0.471
Range	[0.291, 0.946]	[0.00, 0.363]	[0.151, 0.949]
Shapiro-W	0.917	0.775	0.917
P	0.02**	<0.001***	.274
정규성	위배	위배	만족

전반적으로 유사도는 H1 > H3 > H2 일관된 패턴으로 명확한 차이가 관찰되었으며 이런 패턴이 통계적으로 유의미한지 검증 과정을 거친다. 세 가설은 동일 데 이터셋을 앨범 단위로 세 가지 조건으로 진행된 연구로 구체적인 가설 각각의 쌍별 비교를 위해 대응 표본 t-검정(paired text)를 선택하였다. H1의 높은 유사도 분포, H2의 낮은 유사도 분포로 인해 H1, H2은 정규 분포를 위배하였으나 표본 크기 N > 30을 만족하므로 Wilcoxon 검정을 통해 교차 검증을 진행하였다.

[표 7] 통계 분석 결과

분석	통계량	결과	P	효과
ANOVA	F	128.85	<0.001***	$n^2 = 0.665$ -
	Mauchly's W	0.982	.665	
t-검정	H1- H2	t= 15.94	<0.001***	d= 3.49 d= 0.79 d= -2.72
	H1- H3	t= 4.04	<0.001***	
	H2- H3	t= -11.62	<0.001***	
Wilcoxon	H1- H2	W=1128	<0.001***	-
	H1- H3	W= 922	<0.001***	
	H2- H3	W= 6	<0.001***	

먼저 ANOVA 결과, 세 가설의 유의미한 차이가 나타났으며 구형성 가정도 만족하였다. 이는 세 가설이 우연이 아닌 확실한 차이가 있음을 의미한다. 각 쌍별로 차이가 유의미한지 비교를 위해 t-검정을 실시한 결과, H1 > H2 > H3으로 모든 비교 쌍에서 p < 0.001로 유의미한 차이를 나타냈다. Wilcoxon 검정을 통해서도 모든 비교에서 H3 > H2 패턴을 확인할 수 있었다.

### 4-2. 시사점 및 논의

본 연구는 총 48개의 앨범 단위로 528개 곡을 대상으로 GoEmotion 모델을 통해 상위 3개의 감정을 얻고, Plutchik 좌표계로 투영하여 최종적으로 1개의

감정을 얻었다. 이를 기반으로 감정 텍스트 프롬프트를 설계하여 아트와 이미지-텍스트 멀티모달 분석을 진행하였다. 결과는 통계적 검증을 통해 앨범 아트가 H1> H3> H2 순서로 유사도 패턴을 확인하였고 각 가설의 차이가 유의미함을 보였다.

H1 가설에서 가장 높은 분포를 보면 대부분이 앨범 전체 흐름에 대한 감정 프롬프트가 아닌 시작 또는 마지막 트랙 감정에서 높은 유사도를 보였다. 이는 콘셉트 앨범이 서사를 여는 첫 곡의 감정적 고조나 감정 여정이 마무리되는 마지막 곡의 감정이 아트와 가장 연관도 있게 정렬됨을 의미한다. 즉, 앨범의 감정은 동적인 변화를 보이나 앨범 아트는 가설 H1, H3에서 큰 값을 보이며 여정의 시작점, 도착점 그리고 특정 트랙의 감정이 아트로 수렴되는 경향을 보였다. 반면 앨범 흐름 자체를 정의한 프롬프트("예: Sadness-> angry-> optimism")와는 유사도가 낮아 이는 정적인 앨범 아트가 동적인 형태의 감정 여정을 담기에는 한계가 존재함을 의미한다.

또한 H2에서는 타이틀곡의 감정과 앨범 아트는 완전한 무관성을 보였는데 앨범 발매 시 타이틀곡 선정을 가장 중요시하는 상업적인 기준과 다르게 감정 자체를 디자인 요소로 고려하지 못함을 정량적으로 확인하였다.

방법론에서 일반 앨범이 아닌 기준 및 정의가 방대하고 모호하지만, 서사적인 성격을 띠는 콘셉트 앨범을 선정하여 감정 여정을 정량적으로 도출해 흐름을 시각화하였고, 음악의 감정과 아트를 공동 공간에 정렬하는 멀티모달 모델로 정량화한 시도로, 아트 디자인 분야에서 데이터 기반 분석 방법론을 제시하였다. 다만 정량적인 접근만으로는 콘셉트 앨범 아트의 디자인 요소와 복합적 감정을 충분히 포착하기에 한계가 존재한다. 이에 따라 정성적 연구와 통합적 접근을 통해 디자이너의 해석, 청취자의 경험, 디자인 요소(색상, 타이포그래피, 레이아웃 등)를 종합적으로 포괄하는 이해 기반 연구가 필요하다. 또한, 콘셉트 앨범의 성격으로 장르가 특정 장르로 편향된 점을 고려하여 좀 더 느슨한 콘셉트 앨범 기준을 잡고 다양한 장르에서 확장 수집이 필요하다. 후속 연구에서는 장르 확장 및 장르별 분석, 디자인 기반 질적 연구와 결합, 동적 시각화 실험을 통해 동적 아트 분석을 위한 방안을 모색해야 한다.

## 5. 결론

앨범 아트는 음반의 시각적 정체성을 담는 핵심 요소로, LP 시대에서 시작하여 현재의 디지털 스트리밍 환경까지도 중요한 역할로 언급된다. 이에 따라 앨범 아트 디자인에 관한 분석은 꾸준히 이루어져 왔으나, 대부분 시각적 특성이나 개별 사례분석, 혹은 추천 알고리즘을 위한 장르별 아트 생성 연구로 국한되어 왔다. 본 연구는 선행 연구를 다른 관점에서 앨범 아트가 음악 감정 구조와 어떻게 정렬되는지를 정량적으로 분석한 점에서 의의를 가진다.

본 연구는 감정을 사전 학습 모델을 통해 정량화하여 노래 데이터셋의 감정 레이블 부재를 해결하고 정의한 가설을 기반으로 설계한 감정 프롬프트를 CLIP 모델을 활용해 앨범 아트와 의미적 유사도를 비교하였다. 세 가설 중 어느 것이 아트와 가장 강하게 정렬되는지를 검증하였고, 48개 앨범을 대상으로 분석한 결과, 앨범 아트는 전체 감정의 흐름보다는 시작점, 마지막 트랙의 감정(H1) 또는 특정 트랙의 감정(H3)과 관련도가 높음을 보였다. 이는 정적인 아트가 시간적 감정 구조를 담는 데 한계가 있으며 서사의 극적인 감정 만을 포착함을 의미한다.

특히 본 연구는 콘셉트 앨범이라는 서사적 성격을 감정을 통해 정량화한 최초 시도로서 자연어 처리와 멀티모달 모델을 결합한 방법론을 제시하였다. 동시에 정적인 아트 이미지와 감정 반영의 한계를 언급하며 음악 요소 중 가사의 가장 중요한 트리거인 감정을 동적 흐름으로 담아낼 새로운 시각 아트 기법(인터렉티브, 애니메이션 아트 등)의 필요성을 제시한다. 후속 연구에서는 정량, 정성적 통합 접근을 통해 콘셉트 앨범의 감정 흐름을 구현하는 방법을 모색할 예정이다. 이러한 접근은 디지털 스트리밍 시대에 새 방향성을 제시할 것이며, 현재 세계관을 구축해 부흥 중인 K-POP 장르에서도 유의미한 시사점을 제공할 것이다. 본 연구의 확장을 통해 복합적 구조를 분석하고 동적 시각화 기술과 결합하여 음악의 감정 여정까지 구현하는 기반이 될 것이다.

## 참고문헌

1. Robert Plutchik,『The Nature of Emotion』, 2001
2. 박민희, 이은주, ‘인디 앨범 섬네일 디자인의 시각적 특성에 관한 연구’, 일러스트레이션 포럼, 2022
3. 박주향, ‘파라텍스트로써의 앨범커버 연구 -비틀즈의 앨범커버를 중심으로’, 인문사회예술융합, 2023
4. 주효선, 송민정, ‘록 음악 앨범 커버의 시각적 디자인 요소 특징 연구’, 브랜드디자인학, 2021
5. 황지선, 이윤진, ‘파이썬을 활용한 데이터 크롤링과 유형화에 따른 국내 아이돌 앨범 커버 색채 분석’, 한국색채학회, 2022
6. Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever, ‘Learning Transferable Visual Models From Natural Language Supervision’, Proceedings of ICML 2021
7. Alfonso Semeraro, Salvatore Vilella, Giancarlo Ruffo, ‘PyPlutchik: Visualising and comparing emotion-annotated corpora’, PLOS ONE, 2021
8. Hanna-Sophia Widhoelzl, Ece Takmaz, ‘Decoding Emotions in Abstract Art: Cognitive Plausibility of CLIP in Recognizing Color-Emotion Associations’, arXiv Preprint, 2024
9. Jianyi Wang, Kelvin C.K. Chan, Chen Change Loy, ‘Exploring CLIP for Assessing the Look and Feel of Images’, arXiv preprint, 2022
10. Joye, Y., & Fennis, B. M, ‘Set that record straight! Cardinal line orientations in music album artwork boost market performance and music consumption’, Psychology&Marketing, 2024
11. Junnan Li, Dongxu Li, Caiming Xiong, Steven Hoi, ‘BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation’, ICML, 2022
12. Natalie Greenfield & Ngorli Paintsil, ‘Album Covers Deserve Some Attention’, Stanford University, 2024
13. Sorensen, Paige C., ‘The Concept Album Continuum’, Honors Senior Capstone Projects. 43, 2019
14. Tara Venkatesan, Qian Janice Wang, &Charles Spence, ‘Does the typeface on album cover influence expectations and perception of music?’, Psychology of Aesthetics, Creativity and the Arts, 2022
15. Zichao Nie, ‘Feature-Attentive Multimodal Emotion Analyzer with CLIP’, ICICML, 2023
16. Impress Vinyl, ‘Why Music Album Covers Are Still important In The Digital Age’, Vinyl Record Covers Blog, 2025.04.15.(2025.10.26.)
17. Jim Carroll, ‘The art of the sleeve: every album cover tells a story’, The Irish Times, 2016.10.13. (2025.10.26.)
18. research.google, ‘GoEmotions: A Dataset for Fine-Grained Emotion Classification’, 2021.(2025.11.11.)
19. github.com