ISSN 2508-2817 KiDRS Korea Institute of Design Research Society 통권 38호, 2026. Vol.11, No.1

AIGC 기술 기반 고전시가 멀티모달 시청각 커뮤니케이션 디자인 연구

청옥안 원석(青玉案·元夕)과 동짓달 기나긴 밤을"을 중심으로

A research on Multimodal Audiovisual Communication Design of Classical Poetry Based on AIGC Technology

Focusing on QingYuAn: Lantern Festival Night and On a Long Winter Solstice Night

주 저 자 : 종 림 (Zhong, Lin) 한양대학교 디자인학부 시각디자인전공 박사과정

교 신 저 자 : 송민정 (Song, Min Jung) 한양대학교 디자인대학 커뮤니케이션디자인학과 교수
mjung111@hanmail.net

https://doi.org/10.46248/kidrs.2026.1.35

접수일 2026. 02. 15. / 심사완료일 2026. 02. 26. / 게재확정일 2026. 03. 03. / 게재일 2026. 03. 30.

Abstract

The advancement of digital transformation and AIGC (Artificial Intelligence Generated Content) technologies is reshaping cultural engagement, yet classical poetry remains largely confined to text-based interpretation. This study reinterprets Stuart Hall's encoding-decoding theory to develop a multimodal communication design strategy for recontextualizing the aesthetic value of classical poetry in contemporary media environments. Using Qing Yu An · Yuan Xi by Xin Qiji and Dongjitdal Ginagin Bameul by Hwang Jini as case studies, a multi-layered encoding system was constructed to translate text, imagery, and rhythm into audiovisual forms through a five-stage process: text analysis, storyboard design, audiovisual generation, dynamic implementation, and interaction. A/B testing and expert evaluation demonstrated improved content fidelity and artistic representation compared to general AI models, while CLIP similarity scores indicated higher semantic alignment. The study suggests that AIGC can function as an adaptive design system mediating audiovisual reinterpretation and human-AI collaboration in classical poetry.

Keyword

Artificial Intelligence (인공지능), Classical Poetry (고전시가) , Multimodal Communication (멀티모달 커뮤니케이션) , Digital Diffusion (디지털 확산) , Encoding-Decoding (인코딩-디코딩)

요약

디지털 전환과 AIGC(Artificial Intelligence Generated Content) 기술의 발전은 전통문화 향유 방식을 재구성하고 있으나, 고전시가 전승은 여전히 텍스트 중심 해석에 머물러 있다. 본 연구는 스튜어트 홀(Stuart Hall)의 인코딩-디코딩 이론을 재 해석하여 고전시가의 미학적 가치를 현대 미디어 환경에서 재현하기 위한 멀티모달 커뮤니케이션 디자인 전략을 탐색하였다. 사례로 신기질의 〈청옥안·원석〉과 황진이의 〈동짓달 기나긴 밤을〉을 선정하여, 텍스트·심상·운율을 시청각 기호로 변환하는 다층적 인코딩 체계를 설계하고, '텍스트 분석-스토리보드-시청각 생성-동적 구현-상호작용'의 5단계 프로세스를 적용한 인터랙티브 플랫폼 프로토타입을 구현하였다. A/B 테스트와 전문가 평가 결과, 제안 모델은 범용 AI 대비 콘텐츠 충실도와 예술적 재현성에서 향상된 성과를 보였으며, CLIP Score 분석에서도 의미 정합성이 높게 나타났다. 본 연구는 AIGC를 고전시가의 시청각적 재현과 인간-AI 협업을 매개하는 지능형 설계 체계로 확장할 수 있는 가능성을 제시한다.

목차

1. 서론

1-1. 연구 배경 및 목적

고전시가는 동양 문화의 귀중한 자산으로서, 다원적이고 혁신적인 방식으로 세계로 확산하며 국제 문화 교류에서 생동적인 역량을 보여주고 있다. 최근 아시아에서 유럽과 북미 시장에 이르기까지, 동양 전통문화, 특히 고전시가 문화에 대한 시장 수요가 지속해서 증가하고 있으며, 기존의 출판물 중심의 향유 방식에서 공연, 디지털 미디어, 교육 통합 등 다양한 경로로 확장되고 있다. 글로벌 최대 스트리밍 플랫폼 중 하나인 TikTok에서 발표한 "2022 고전시가 데이터 보고서"에 따르면, TikTok에서 고전시가 관련 동영상 누적 조회수는 178억 회에 달했으며, 전년 대비 168% 증가하였다. 동시에 TikTok 전자상거래 플랫폼에서의 고전시가 관련 서적 판매량은 전년 대비 588% 증가하였다. [1)]

한국은 동아시아 한자 문화권의 중요한 구성원으로서, 중국 고전시가에 대한 깊은 역사적 연계성과 독자적인 문화적 정체성을 갖고 있다. 한국 전통 한시 창작은 천 년 이상의 역사를 이어왔으며, 9세기 말부터 20세기 초까지 풍부한 시가 전통을 형성하였다. 이러한 역사적, 문화적 연계는 현대 한국에서 고전시가 문화의 확산에 견고한 토대를 제공한다.

한국 사회에서 고전시가에 관한 관심은 일상생활의 여러 측면에서 나타난다. 2020년 코로나19 팬데믹 동안, 한국 정부와 언론은 중국 고전시가를 활용하여 연대와 희망을 표현하였으며, 이는 일상생활에서 정서적 위안을 제공하는 역할을 했다. 2024년 1월, 서울시는 외국 시를 선정하여 지하철 안전문에 게시함으로써 국내외 방문객을 환영하였다. 이때 선정된 시에는 이백, 두목, 신기질 등 중국의 저명한 문인들의 고전시가가 포함되었다. 명동역의 〈산행〉 (두목), 홍익대학교 역의 〈추노아·박산도중벽〉 (신기질), 대림역의 〈산중 문답〉 (이백)이 포함되었다. 이 조치는 각국 국민으로부터 높은 평가를 받았다.[2)]

특히 10~20대 젊은 세대의 손에서 고전시가는 혁신적 부활을 경험하고 있다. TikTok에서 가장 인기 있는 챌린지는 "랩으로 당나라 시 외우기"로, 젊은이들은 "황하 원상 백운 간"을 3연속 운율(韻律, rhythm)로 나누고 배경 비트는 K-pop 리듬을 사용한다. 이러한 혁신적 향유 방식은 젊은 층에서 고전시가의 새로운 생명력을 부여한다.

비록 고전시가가 현대 확산 과정에서 눈에 띄는 성과를 거두었으나, 여전히 여러 도전과제가 존재한다. 첫째, 멀티모달 커뮤니케이션은 언어적 복잡성과 문화적 오해라는 이중 장벽에 직면한다. 고전시가의 풍부한 시적 이미지와 정교한 언어적 표현은 소통 과정에서 그 고유한 미적 정취와 함축적 의미를 상실할 위험이 있어, 목표 문화권의 온전한 이해를 어렵게 한다. 둘째, 콘텐츠 내용과 형식은 여전히 신기술과의 결합이 필요하며, 단순한 내용 이전만으로는 심층적 감정 공감을 유발하기 어렵다.

본 연구는 멀티모달 생성형 AIGC 기술이 전통 고전시가의 디지털 확산과 혁신적 표현에 미치는 잠재적 활용 가능성을 탐구하는 데 목적을 두고 있으며, "텍스트-이미지-영상-사운드"의 멀티모달 인코딩 프레임을 구축하여 고전시가 이미지의 시각화, 영상화 및 청각적 구현을 실현한다. 이론적 측면에서, 본 연구는 인코딩-디코딩 이론을 기반으로 고전시가의 매체 간 변환 과정에서의 인코딩 메커니즘과 수용자의 디코딩 행동을 분석하고, 다양한 기호 (sign) 양식 간 구조적 동형성과 미학적 공명 원리를 규명한다. 실천적 측면에서, 중국 송대 신기질의〈청옥안·원석(靑玉案·元夕)〉과 조선 중기 황진이의〈동짓달 기나긴 밤을〉을 사례로 삼아 고전시가 멀티모달 창작의 기술적 경로와 운영 방법을 탐색한다. 종합적으로, 본 연구는 디지털 시대 한중 고전

1) TikTok, 2022 고전시가 데이터 보고서, (2022.06.21.) https://mp.weixin.qq.com/s/k4Jary0ezL-P5xKC5YDbzA

2) 중국뉴스넷, 중국 고전시시, 한국 서울 지하철에 입성하다, (2024.09.20.) https://www.chinaqw.com/sp/2024/01-22/371446.shtml

시가의 문화적 확산과 교육적 보급, 기술적 적합 경로 탐색에 대한 이론적 근거와 기술적 참고를 제공함으로써, 전통문화의 신매체 환경에서의 활성화와 혁신적 발전을 촉진하는 데 기여하고자 한다.

1-2. 연구 방법 및 범위

본 연구는 문헌 고찰, 사례 분석, 프로토타입 개발 및 실증 검증의 통합적 방법론을 통해 진행되었다. 첫째, 문헌 고찰을 통해 인코딩-디코딩 이론과 AIGC 기술의 문화콘텐츠 적용 현황을 분석하여 연구의 이론적 토대를 마련하였다. 둘째, 신기질의〈청옥안·원석〉과 황진이의〈동짓달 기나긴 밤을〉을 핵심 사례로 선정하여, 텍스트 분석, 스토리보드 설계, 시청각 생성에 이르는 AIGC 기반 실증 창작 연구를 수행하였다. 셋째, 실증 과정을 기반으로 '입력-자동 인코딩-멀티모달 생성-상호작용'이 가능한 고전시가 AIGC 인터랙티브 플랫폼 프로토타입을 설계·구현하였다. 넷째, 구축된 시스템의 효용성을 검증하기 위해 일반 사용자와 전문가 그룹을 대상으로 A/B 테스트와 정량적·정성적 평가를 수행하여 제안 모델의 타당성을 입증하였다.본 연구의 핵심 검증 대상은 사용자 생성 행위 자체가 아니라, 제안된 멀티모달 인코딩 구조가 고전시가의 의미를 효과적으로 전달하는가에 있다. 사용자 평가는 해당 구조의 전달 효율성과 의미 재현 가능성을 확인하기 위한 보조적 검증 절차로 설정되었다. 본 연구의 범위는 이러한 이론적, 실천적, 검증적 과정을 통해 디지털 시대 고전시가의 창의적 확산과 인간-AI 협업을 위한 구체적인 방법론과 기술적 경로를 제시하는 데 있다.

2. 이론적 배경

2-1. 고전시가 멀티모달 확산

최근 연구에 따르면 고전시가의 멀티모달 확산은 점차 발전하고 있으며, 특히 중국 고전시가를 중심으로 한 디지털 시청각 실천이 활발히 이루어지고 있다. 텍스트에서 이미지로의 전환 측면에서, Cai 등(2023)은 '황학고시(黃鶴故詩)' 디자인 프로젝트를 통해 황학루 관련 고전시가를 디지털 몰입 환경으로 재해석하였다. 그는 AI 회화와 VR 체험을 결합한 고전시가 교육 및 커뮤니케이션 디자인 방식을 모색함으로써, 인공지능과 가상현실 환경에서의 고전시가 문화의 디지털 보존과 확산에 실질적 참고 경로를 제시하였다.[3)]

고전시가 상호작용 창작 측면에서 Guo 등(2019)은 인간-기계 협업형 중국 고전시가 생성 시스템인 Jiuge를 제안하여, 사용자가 다회 수정과 입력(키워드, 텍스트, 이미지 등)을 통해 창작에 동적으로 참여할 수 있도록 하였으며, 참여성과 맞춤성을 높였다. [4)]

강병규(2019)는 인공신경망 기반 AI의 문학 창작 적용을 논의하며, AI가 대규모 고전시가 데이터를 심층 학습함으로써 자연스러운 고전시가를 생성할 수 있음을 밝히고, 이를 인간 창작의 보조 도구이자 시 창작과 이해의 인지 확장 수단으로 보았다.[5)]

Zheng(2025)은 중국 최초의 텍스트-영상 AI 애니메이션 "천추시송(千秋詩頌)"을 사례로, AI 기술이 고전시가 단편 영상 창작에 적용되는 시각화 전략과 문화 확산 효과를 분석하였다. 연구는 교육 지향적 소재 선정, 수묵 미학 기반 시각 구성, 그리고 멀티모달 협업 전략의 세 측면을 중심으로 고전시가의 매체 전환 과정을 고찰하였다. 분석 결과, AI 기술은 창작 효율과 확산 효과를 높였으며, 청년층의 전통문화 인식을 강화하였으나, 동시에 콘텐츠 동질화와 저작권 문제 등 기술 윤리의 새로운 과제를 초래하였다. 이에 따라 연구는 인문적 판단과 기술적 창의의 균형을 유지하며, 효율성과 깊이를 겸비한 새로운 문화 확산 모델 구축의 필요성을 제시하였다.[6)]

종합적으로 현재 연구는 고전시가 단일 모달리티에서 완전한 시청각 및 몰입형 창작 환경으로 확장되는 경향을 보이고 있으나, 한국 고전시가 분야에 AIGC 기

3) Cai Xinyuan, Chen Qiuchan, Zhang Jian, 'AI gongchuang de Zhongguo gushi wenhua VR tiyan xitong sheji yu shixian', PUBLISHING JOURNAL, 2023. 08, Vol.31, No.4, p.80.

4) Guo Zhipeng, Xiaoyuan Yi, Maosong Sun, Wenhao Li, Cheng Yang, Jiannan Liang, Huimin Chen, Yuhui Zhang, Ruoyu Li, 'Jiuge: A Human-Machine Collaborative Chinese Classical Poetry Generation System', 『Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations』, Association for Computational Linguistics, 2019, pp.25-30.

5) 강병규, 'AI의 중국 古典詩歌 창작 —詩語의 학습과 생성', 중국문학, 2019. 01, Vol.100, pp.183-208.

6) Zheng, Junlin, 'Research on Visualization Strategies for Classical Poetry Short Videos Based on AI Technology: A Case Study of 'Ode to the Ages'', New Media Research, 2025. Vol.11, No.12, pp.109-113.

술을 접목한 표현에 관한 연구는 여전히 부족하며, 음악 및 해설 생성을 결합한 연구도 거의 이루어지지 않고 있다.

2-2. 멀티모달 및 AIGC 기술

멀티모달 이론은 AI 디자인 연구의 이론적 토대를 제공하며, 체계기능언어학(SFL)의 관점을 도입하여 이미지와 음성 등 비언어적 기호의 의미 생성 과정을 규명하는 데 그 주안점을 둔다.

최근 AIGC 기술을 활용한 멀티모달 연구는 창작, 분석, 교육, 상호작용 등 다각적인 측면에서 진행되고 있다. 창작 측면에서 Yang 등(2024)은 대규모 언어 모델 기반의 크로스모달 중국 시 창작 시스템을 제안하였으며, 텍스트와 이미지 입력을 모두 지원하고 반복적 최적화와 설명 가능한 주석 기능을 갖추어 시 생성에서 크로스모달 입력의 우수성을 검증하였다.[7] 분석 및 보존 측면에서 박라미 등(2025)은 안토니오 가우디 건축을 사례로 멀티모달 AI 기반의 디자인 요소 텍스트화(Textualization) 방법을 제시하였다. 이는 이미지와 텍스트 분석을 결합하여 건축물의 주요 디자인 요소를 추출하고 설명을 자동 생성함으로써, AI 예술교육과 디지털 문화유산 보존에 실질적 적용 모델을 마련하였다.[8] 교육적 활용 측면에서 김선영(2024)은 S대학의 멀티모달 생성형 AI 콘텐츠 창작 교육 프로그램을 분석하였다. 해당 연구는 10개의 융합 전공 학생팀이 다양한 매체(텍스트, 이미지, 오디오, 영상)를 활용해 콘텐츠를 제작하는 과정을 교육 단계, 경험 공유, 성과 확산의 세 단계로 체계화하여 멀티모달 AI 교육의 개발 및 운영 전략을 제시하였다.[9] 상호작용 시스템 측면에서 송영훈 등(2025)은 이미지 객체 탐지와 텍스트 모델을 결합한 멀티모달 인터랙티브 시스템을 통해 시각 정보와 자연어 질의의 동시 처리를 구현하였다. 실험 결과, 본 시스템은 공항 내비게이션 및 학습 보조 등 다양한 분야에서 정교한 실시간 상호작용이 가능함을 확인하였으며, 데이터 불확실성 및 오류 관리 등 향후 발전 과제를 논의하였다.[10]

이처럼 AIGC 기술은 지속적으로 발전하고 있으나, 현재의 범용 모델은 심층적인 기호 추론과 고유한 문화적 맥락을 이해하는 데 있어 여전히 한계를 보인다. 특히 시각언어 모델은 특정 문화적 이미지를 처리할 때 표면적인 패턴 매칭에 의존하는 경향이 있어, 동양의 미학적 맥락에 대한 이해 부족이나 문화적 오독을 초래하기도 한다. 따라서 고전시가와같이 함축적 의미와 문화적 상징이 풍부한 콘텐츠를 온전히 구현하기 위해서는, 범용 모델의 한계를 보완할 수 있는 특화된 인코딩 메커니즘과 멀티모달 생성 전략이 요구된다.

2-3. 인코딩-디코딩 이론

스튜어트 홀(Stuart Hall)의 인코딩-디코딩 이론은(Encoding-Decoding Theory) 미디어 메시지의 생산과 수용 과정을 설명하는 핵심 프레임워크로, 1973년 유럽위원회 콜로키움에서 처음 제기된 후 1980년 수정·보완을 거쳐 정립되었다. 홀은 미디어 텍스트가 단순히 투명한 정보를 전달하는 것이 아니라, 송신자의 의미 구조가 기호화되는 '인코딩(Encoding)'과 수용자가 자신의 사회·문화적 배경을 바탕으로 의미를 재해석하는 '디코딩(Decoding)'의 결합으로 이루어진다고 보았다. 이는 커뮤니케이션이 선형적 전달이 아닌, 의미의 생산과 재구성, 그리고 잠재적 저항이 공존하는 역동적 순환 과정임을 시사한다.[11]

이러한 이론적 관점은 디자인 및 공학 분야로 확장되어 적용되고 있다. 디자인 커뮤니케이션 측면에서 손주현(2012)은 아이덴티티 디자인의 과정을 인코딩과 디코딩 메커니즘으로 분석하였다. 그는 기업의 정체성을 개념화하고 이를 시각적·언어적 기호로 변환하는 과정을 인코딩으로, 수용자가 디자인 결과물을 해석하여 목표 이미지를 인지하는 과정을 디코딩으로 정의하며, 이 과정에서 발생하는 '객체-정체성-이미지' 간의 간극을 줄이기 위한 기호학적 접근의 중요성을 강조하였다.[12]

---

7) Liwen Yang, Zhidong Zhang, Kaipeng Niu, Sitian Pan, Weiping Zhu, Chao Ma, 'Large Model Based Crossmodal Chinese Poetry Creation', 『2024 IEEE Smart World Congress (SWC)』, 2024, pp.27－34.

8) 박라미, 유진, 최유리, 오효정, 'AI를 활용한 멀티모달 분석 기반 디자인 언어 이해: 가우디 건축 사례를 중심으로', 디지털콘텐츠학회논문지, 2025. 05, Vol.26, No.5, pp.1161－1169.

9) 김선영, '대학교육에서 멀티모달 기반 생성형 AI를 활용한 콘텐츠 제작 프로그램 사례 연구', 한국콘텐츠학회논문지, 2024. 11, Vol.24, No.11, pp.609－617.

10) 송영훈, 김남기, 정경용, '객체 탐지와 텍스트 융합 기반 멀티모달 인터랙티브 시스템', 한국정보기술학회논문지, 2025. 07, Vol.23, No.7, pp.115－122.

11) Hall, S., "Encoding/Decoding", Culture, Media, Language, Hutchinson , 1980. 1972-79 , pp. 128-138.

기술적 구현 측면에서 Zhao 와 Lee(2022)는 Transformer-XL 기반의 고전시가 생성 시스템을 통해 인코딩-디코딩의 공학적 모델을 제시하였다. 해당 모델은 다중 헤드 어텐션(Multi-head Attention) 메커니즘을 통해 고전시가의 운율 규칙과 의미 정보를 인코딩하고, 디코딩 단계에서 해당 규칙에 부합하는 시구를 생성한다. 또한, BERT와 음운 점 검기를 활용한 2차 디코딩 평가를 통해 생성된 텍스트의 의미적 유창성과 운율 적합성을 검증하였다.[13)]

이상의 논의를 종합하면, 현대의 고전시가 확산은 단순한 텍스트 전달을 넘어 시각·청각·언어가 결합된 멀티모달 커뮤니케이션으로 확장되고 있다. AI 기반 생성 모델은 시적 이미지의 재구성과 정서적 몰입을 가능하게 하였으나, 여전히 깊이 있는 문화 맥락의 해석과 전달에는 한계를 보인다. 본 연구는 인코딩-디코딩 이론을 AIGC 환경에 재적용하여, AI가 생성하는 시각·청각 기호를 단순한 정보 처리가 아닌 문화적 의미의 '재부호화' 과정으로 해석한다. 이를 바탕으로 AIGC 기술을 매개로 한 고전시가의 멀티모달 확산 구조를 분석하고, 인공지능과 인간의 상호작용 속에서 발현되는 새로운 시적 경험과 문화 향유 양식을 탐색하고자 한다.

3. 고전시가 멀티모달 인코딩 및 디코딩 메커니즘

3-1. 고전시가 인코딩 메커니즘

고전시가는 고도로 응축된 언어와 복합적인 이미지 층위를 특징으로 하여, 이를 AIGC 모델이 즉각적으로 해석하기에는 의미적 모호성과 문화적 맥락의 장벽이 존재한다. 따라서 본 연구는 고전시가의 텍스트 정보를 '구조화된 데이터'로 변환하는 인코딩 메커니즘을 제안한다. 이는 시어에 내재된 문화적 이미지, 정서적 색채, 운율적 패턴을 AI가 연산 가능한 '의미 기호 데이터베이스'로 체계화하는 과정이다. 이 과정은 단순한 기술적 번역을 넘어, AIGC 환경에서 전통, 시정(詩情, poetic sentiment) 의 미학적 가치를 보존하고 동적 시청각 콘텐츠로 재활성화하는 전제 조건이 된다.이 과정은 기호학적 관점에서 볼 때, 언어적 '기표(Signifier)'로만 존재하던 고전시가의 텍스트를 해체하여, 그 이면에 담긴 문화적·정서적 '기의(Signified)'를 추출하고 이를 시청각적 생성을 위한 새로운 기호 체계로 재편하는 작업이다.

본 연구는 생성 모델(Midjourney, Stable Diffusion 등)의 문법 논리에 맞춰, 고전시가의 정보를 '텍스트(구조)-이미지(심상, 心象, poetic imagery)-음률(운율)'의 3단계 레이어로 변환하는 멀티모달 인코딩 시스템을 구축하였다.

텍스트 층: 구조적 벡터화 텍스트 층에서는 시가의 언어적 리듬과 수사적 관계를 구조 템플릿으로 인코딩한다. 각 시구는 글자 수(W), 시체(Form), 장법(Structure)에 따라 다음과 같이 벡터화된다.

행 및 시체 인코딩: 오언(S5) 및 칠언(S7) 절구와 율시, 혹은 한국 시조의 음보율(3-4-4조 등)을 기준으로 각 행을 구조화 단위(Ustruct)로 정의한다.

장법 구조(Zk): 시의 서사적 흐름인 기승전결(起承轉合)을 위치 변수 [Z1: Intro, Z2: Develop, Z3: Turn, Z4: Conclude] 로 매핑하여 시상 전개에 따른 연출의 근거를 마련한다.

기호학적 관점에서 이 과정은 언어적 기표(Signifier)를 구조 단위로 분절하여, 생성 모델이 연산 가능한 형식적 기의(Signified)로 치환하는 1차 기호화(primary signification)에 해당한다.

이미지 층: 시각적 파라미터 매핑 고전시가의 추상적 시어를 구체적인 시각 프롬프트로 변환하기 위해, 장면 정보를 다차원적 속성 기반의 생성 프레임으로 정의한다. 또한, '고월(孤月)', '잔월(殘月)' 등 다양한 유의어를 의미 군집화(Semantic Clustering)하여 통일된 시각 코어(예: 'YUE')로 치환함으로써 생성의 일관성을 확보한다:

(FRAMEn):{Scene, Time, Light, Camera, Character, Action, Mood}

이는 소쉬르(Saussure)의 기호 체계에서 자의적(arbitrary) 기호인 시어가, 도상성(iconicity)을 지닌 시각 기호로 전환되는 과정으로 볼 수 있다.

음률 층: 청각적 기호 변환 시의 운율과 감정의 강약을 청각적 신호로 변환한다. 한자의 평측(平仄)이나

12) 손주현, 박진숙, '아이덴티티디자인 커뮤니케이션 과정 중의 오류 발생원인에 관한 연구', Archives of Design Research, 2012. 08, Vol.25, No.3, pp.152-161.

13) Jianli Zhao, Hyo Jong Lee, 'Automatic Generation and Evaluation of Chinese Classical Poetry with Attention-Based Deep Neural Network', Applied Sciences, 2022. 01, Vol.12, No.13, p.6497.

우리말의 음보를 이진 코드(Binary Code)로 치환하여 낭송의 리듬과 배경음의 템포를 제어한다.

평측 이진화: 평성(0) / 측성(1), 예: "명월(평측)" = [0, 1], 운모 클러스터(韻母, rhyme cluster): 압운 (押韻, rhyme) 정보를 그룹화하여 음색 생성의 변수로 활용한다.

평측의 이진화는 시의 음성적 기표를 지표적 기호(Indexical Sign)로 변환하는 절차로, 청각적 기호 생성의 물질적 근거를 제공한다.

규칙 기반 매핑 함수 (Mapping Function): 앞서 정의한 각 레이어의 정보는 최종적으로 시청각 생성을 위한 파라미터 함수로 통합된다. 본 연구의 인코딩 메커니즘은 복잡한 딥러닝 연산이 아닌, 고전시가의 고유 규칙을 명확한 파라미터로 치환하는 규칙 기반 매핑을 따른다.

Vi = fvis(Wi, Zk)+{Scene, Light, Action, ...}

시각화 함수(Vi): 시구의 핵심어(Wi)와 장법(章法) 구조 내 위치(Zk)를 입력변수로 하여, 동일한 시어라도 기승전결(起承轉結) 의 위치에 따라 조명과 연출이 달라지도록 시각적 속성을 도출한다.

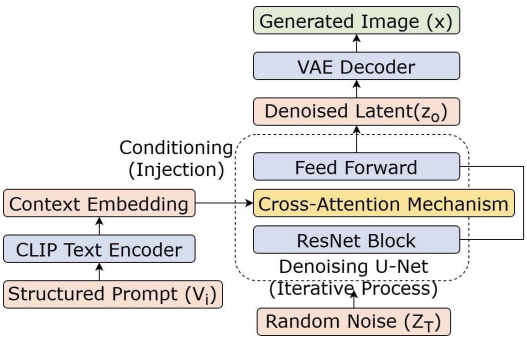
Ai =faud (Bi, Yi)+{Pitch, Speed, Timbre}

청각화 함수(Ai): 평측 이진 코드(Bi)와 운모 클러스터(Yi)를 입력받아 낭송의 음고, 속도, 음색을 제어하는 오디오 벡터를 생성함으로써, 시의 운율 정보가 물리적 음향 신호로 직역되도록 한다.

3-2. 고전시가 멀티모달 디코딩 메커니즘

본 연구에서 디코딩은 앞서 인코딩된 구조화된 데이터(Ptotal)를 수용자가 감각할 수 있는 물리적 미디어(영상, 소리)로 변환하는 과정이다. 이는 단순한 정보의 번역을 넘어, 생성 모델의 잠재 공간 내에서 추상적 기호가 구체적 실체로 물질화되는 근본적 작동 원리를 따른다.기호학적 차원에서 이 디코딩 과정은 추상적인 '상징적 기호(Symbolic Sign, 텍스트)'가 AIGC 모델의 잠재 공간(Latent Space)을 거쳐 구체적인 '도상적 기호(Iconic Sign, 시각 이미지)' 및 '지표적 기호(Indexical Sign, 청각 및 운율)'로 물화(Materialization)되는 의미 생성 과정이다.

시각적 디코딩은 텍스트 프롬프트(Vi)를 픽셀 단위의 이미지로 복원하는 과정으로, 잠재 확산 모델(Latent Diffusion Model, LDM) 아키텍처를 기반으로 구현된다. LDM은 픽셀 공간이 아닌 압축된 잠재 공간에서 연산을 수행하여 효율성을 극대화한다. 구체적으로, 인코딩된 텍스트는 CLIP(Contrastive Language-Image Pretraining Score) 인코더를 통해 임베드되어 U-Net 구조의 교차 주의(Cross-Attention) 레이어에 주입되며, 노이즈가 제거되는 역확산(Reverse Diffusion) 단계를 거쳐 구체적인 시각 정보로 복원된다. 이 과정의 핵심 논리는 '속성 매핑(Attribute Mapping)'이다. 시스템은 시어에 내재된 '정서적 키워드(예: 슬픔)'를 AI가 인지 가능한 '시각적 파라미터(예: 낮은 명도, 차가운 색조, 확산된 광원)'로 치환한다. 즉, 문학적 의미를 빛과 공간의 조형 언어로 재해석하는 논리적 과정이다.



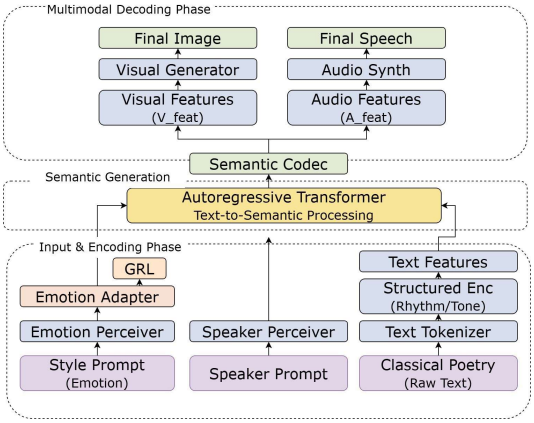
[그림 1] LDM 모델 아키텍처 원리

〈그림 1〉은 본 연구의 핵심 사례인 신기질의 〈청옥안·원석〉에 대한 시각적 디코딩 과정을 예시로 보여준다. '동풍(東風)', '화천수(花千樹)', '더불거(寶馬)' 등의 시어는 LDM 아키텍처를 통해 화려한 원소절(元宵節)의 야경과 축제 분위기로 전이된다. 반면, '등불이 희미한 곳(燈火闌珊處)'이라는 시구는 '어두운 명도', '고립된 피사체', '배경 흐림' 효과로 매핑되어, 군중 속 고독과 성찰이라는 시의 주제 의식을 입체적인 회화로 재현한다. 이를 통해 원시의 구조적 미학을 직관적이고 감각적으로 시각화한다.

청각적 디코딩은 인코딩된 운율 벡터(Ai)를 물리적인 음향 파형으로 합성하는 단계이다. 본 연구는 오픈소스 모델인 Index TTS2의 아키텍처를 차용하여 이 과정을 구현하였다. 이 단계의 중심 절차는 '운율적 추론(Prosodic Inference)'이다. 인코딩 단계에서 정의된 평측과 압운 정보는 모델 내부의 '길이 임베딩(Duration Embedding)' 레이어를 통과하며 발화 속도와 리듬으로 변환된다. 또한, 감정 벡터(Emotion Vector)는 '감정-음색 분리' 모듈을 통해 화자의 고유

음색을 유지하면서도 시의 정서(예: 애절함, 웅장함)에 맞는 억양(Pitch)과 떨림을 생성한다.[14]

GPT 기반 잠재 특성(Latent Feature) 융합은 발음의 명료도와 자연스러움을 극대화하며, 사용자의 자연어 명령(예: "잔잔하고 슬픈 어조로 낭독")을 정밀한 감정 벡터로 변환하여 정서적으로 풍부한 음성 출력을 가능하게 한다. 이를 통해 텍스트로만 존재하던 시의 운율은 시간성을 가진 동적인 청각 경험으로 재구성된다.



[그림 2] 전체 시스템 아키텍처

최종적으로 시각적 디코딩(영상)과 청각적 디코딩(오디오) 결과물은 타임라인 상에서 정교하게 동기화되어 통합된 멀티모달 콘텐츠를 형성하며, 이에 대한 전체 시스템 아키텍처는 <그림 2>과 같다. 이 과정은 일방적인 생성을 넘어, 사용자가 생성 결과에 개입하고 조율하는 'Human-in-the-Loop(인간 참여형)' 구조를 핵심으로 한다. 이러한 디자인은 고전시가의 멀티모달 확산을 시각화하고 몰입감을 제공하는 동시에, 수용자의 능동적 참여와 심층적 감상을 실현한다. 후술할〈청옥안·원석〉과〈동짓달 기나긴 밤을〉의 사례 분석 역시 이 시스템을 기반으로 수행된다. 궁극적으로 멀티모달 디코딩은 단순한 정보의 재현을 넘어, 상호작용을 통한 피드백 기반의 미학적 경험 최적화 메커니즘으로 기능하며, 디지털 환경 속에서 고전시가의 효과적인 확산을 가능하게 한다.

14) Siyi Zhou, Yiquan Zhou, Yi He, Xun Zhou, Jinchao Wang, Wei Deng와/과Jingchen Shu, 'IndexTTS2: A Breakthrough in Emotionally Expressive and Duration-Controlled Auto-Regressive Zero-Shot Text-to-Speech', arXiv, 2025, arXiv:2506.21619

4. AIGC 기반 고전시가 멀티모달 창작 실증 연구

4-1. 스크립트 구축과 스토리보드 디자인

본 절에서는 신기질의〈청옥안·원석〉과 황진이의 〈동짓달 기나긴 밤을〉을 대표 사례로 삼아, 본 연구에서 제시하는 인코딩-디코딩 메커니즘 이론을 기반으로 고전시가의 AIGC 멀티모달 생성 과정에서 나타나는 이미지 전이와 시청각적 상호작용 경로를 분석하였다. 스토리보드 구축, 영상 합성, 음악 구성은 텍스트의 의미 정보를 구체적인 형상과 청각적인지로 매핑하는 변환 레이어에 해당하며, 이는 멀티모달 고전시가 생성 체계의 필수적인 과정이다.

이미지 층위 인코딩에서는 시문의 주요 장면을 중심으로 의미에서 시각으로의 대응 관계를 구축한다. Midjourney V7 이미지 생성 모델은 빛과 그림자, 공간, 동작을 활용하여 시간의 흐름을 조형적 언어로 재구성한다.〈청옥안·원석〉의 이미지 생성 상세 파라미터는 <표 1>에, 〈동짓달 기나긴 밤을〉의 상세 파라미터는 <표 2>에 제시되어 있다. 예를 들어 "동풍(東風)"과 "화천수(花千樹)"와 같은 언어적 기표(Signifier)는 단순히 텍스트를 넘어, Midjourney V7의 파라미터 제어를 통해 '송대 원소절 밤의 화려한 풍경'이라는 시각적 도상(Icon)으로 변환된다. 이는 고전시가의 함축적 의미(기의)가 시각적 기호 체계로 성공적으로 치환되었음을 보여주는 기호학적 실천이다. 시각적으로 구현된 다중 카메라 시점과 조명은 원시의 정서를 담아내는 지표적(Indexical) 역할을 수행한다. "동짓달 기나긴 밤을"에서는 흔들리는 촛불과 느리게 이어지는 영상의 선율로 '시간의 정적 흐름'을 형상화한다.

[표 1] 청옥안·원석 이미지 생성 모델 파라미터

분류	항목	설정값/내용
플랫폼/모델	플랫폼 명	무계AI (Midjourney V7)
기본 설정	화폭 비율	16:9
	해상도	2912x1632 px
	랜덤 시드	382570
스타일 제어	예술 스타일	중국풍 회화
	예술 기법	수묵화, 수채 물감, 손 그림 예술품, 붓 그림
	예술 유파	민간 예술, 고전 건축주의
	작가 스타일	장택단(张择端)
파라미터	스타일 차이화	10%
	스타일 예술화	75%
	샘플링 모드	DPM++ SDE

		Karras
	CLIP Skip	9.5
	ENSD	9
융합 모델	만화 삽화	0.2
	중국풍 회화책	0.8
	현대 수묵	0.7
	NiJi 고풍	0.2
부정 프롬프트	부정 키워드	왜곡된 인체, 투시에 맞지 않는 구조, 많은 손가락, 무작위성

음률 층위 인코딩에서는 시의 평측 리듬과 운자 구조를 분석하여 IndexTTS 모델 및 Sound Style Transfer 모델에 입력하고, 피치, 속도, 볼륨 파라미터를 시의 운율과 정서에 맞춰 조정함으로써 배경음과 시적 리듬이 상호 호응하는 청각적 흐름을 형성한다. 특히 "동짓달 기나긴 밤을"의 사운드 생성에서는 거문고 음색과 여성의 저음 보컬을 기반으로 음색 스타일 전이 기법을 적용하여 시 속의 정서적 경험을 청각적으로 재현한다.

[표 2] 동짓달 기나긴 밤을 이미지 생성 모델 파라미터

분류	항목	설정값/내용
플랫폼/모델	플랫폼 명	무계AI (Midjourney V7)
기본 설정	화폭 비율	16:9
	해상도	2912x1632 px
	랜덤 시드	619245
스타일 제어	예술 스타일	한국풍 정서 동양적 미학
	예술 기법	평면 일러스트, 적은 질감
	예술 유파	낭만주의 상징주의
	작가 스타일	미야자키 하야오
파라미터	스타일 차이화	10%
	스타일 예술화	75%
	샘플링 모드	DPM++ SDE Karras
	CLIP Skip	9.5
	ENSD	9
융합 모델	만화 삽화	0.2
	중국풍 회화책	0.8
	현대 수묵	0.7
	NiJi 고풍	0.2
부정 프롬프트	부정 키워드	왜곡된 인체, 투시에 맞지 않는 구조

이러한 스크립트 기반으로 시스템은 자연언어 분석과 의미 분할을 통해 시문에 내재한 시공간, 인물, 정서 구조를 재 인코딩하고, 각 의미 단위를 이미지 생성 모델이 호출할 수 있는 장면 모듈로 변환한다. 스토리보드 생성 단계에서는 '의미 노드-카메라 단위-영상 서사'의 논리를 축으로 Midjourney V7 모델의 프롬프트 제어 및 이미지 임베딩(Image Embedding) 기법을 활용하여 의미에서 영상으로의 구조적 대응을 구축한다. "청옥안·원석"의 인코딩 키워드 및 스토리보드는 <표 3>에, "동짓달 기나긴 밤을"의 인코딩 키워드 및 스토리보드는 <표 4>에 제시되어 있다. --ar 파라미터를 통한 이미지 화면비(예: 16:9 또는 21:9) 제어, fps (frames per second) 조절을 통한 시간 시퀀스 제어, 그리고 조명 변수 조절을 통해 시의 리듬과 감정의 흐름에 맞는 스토리보드를 자동 생성하며, 이를 통해 언어적 논리와 조형적 논리의 동적 정렬을 달성한다. 또한 수동 검수 및 모델 보정 절차를 도입하여 화면 연출과 시적 분위기의 일체성을 확보함으로써, 고전시가가 입체적인 영상미로 생동감 있게 구현되도록 하였다.

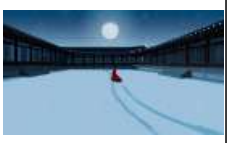
[표 3] 청옥안 원석(青玉案 元夕) 시나리오 및 스토리보드 인코딩

스토리보드	[FRAME$_n$]
	Scen: capital city night Time:night Light:lanterns fireworks Camera: aerial dolly Character: crowd Action: fireworks Emotion: festive Sound:crowd fireworks Narration: none
	Scene: street festival Time: lantern night Light: lantern glow smoke Camera: midclose-up Character: dragon dancers Action: celebration Emotion: joy Sound: drums Narration: ambient
	Scene: scholar walking Time:night Light: moonlight lanterns Camera: follow shot Character: scholar Action: walking Emotion: reflective Sound: crowd fades Narration: inner thought
	Scene:within the crowd Time: late night Light: flickering lanterns Camera: handheld tracking Character: scholar Action: searching Emotion: tense Sound: breathing Narration: none

	Scene: woman under the moon Time: midnight Light:cool moon warm lantern Camera: close-up Character: woman Action:gazing Emotion:stillness Sound: wind Narration: inner thought
	Scene: distant fade-out Time: night's end Light:ink wash Camera: pull-back Character: woman leaving Action: merging light Emotion: serenity Sound: silence Narration: poem appears

디코딩 단계에서는 시간적 시퀀스에 따른 장면 구성과 조명 변화 디자인을 통해 고전시가의 정취를 동적으로 시각화하며, 관객은 영상의 리듬과 청각적 운율, 그리고 실시간 상호작용을 통해 시의 의미를 입체적으로 재해석한다. 영상 생성 단계에서는 스토리보드를 기본 입력으로 삼아, 정지 이미지와 시공간 제어 파라미터를 Midjourney V7의 이미지-영상 변환 기능 및 특정 영상 생성 신경망에 적용함으로써 동적 영상 변환을 수행한다. 생성 과정에서는 의미 가중치 제어(Semantic Weight Control)와 감정-운율 매핑 로직을 활용하여 빛, 색조, 구도의 변화가 시의 정서적 흐름과 일치하도록 조정한다. 또한 카메라 이동 속도, 초점, 심도 등의 시네마틱 파라미터를 체계적으로 설정하여 시구의 운율이 영상의 동적 흐름으로 자연스럽게 전이되도록 구현한다. 최종적으로 시청각 동기화의 완성도를 높이기 위해 영상과 오디오 데이터 간의 정밀한 정렬을 수행함으로써, 화면의 리듬과 청각적 구조가 시간축 상에서 유기적으로 결합되도록 하였다.

[표 4] 동짓달 기나긴 밤을 시나리오 및 스토리보드 인코딩

스토리보드	[FRAME$_n$]
	Scene: winter night Time: moonrise Lighting: cold blue tone Camera: forward dolly Character: none Action: bamboo shadows swaying Emotion: tranquility Sound: wind Narration: none
	Scene: interior Time: deep night Lighting: candlelight + moonlight Camera: close-up Character: woman Action: sitting still Emotion: calm Sound: faint crackling fire Narration: none
	
	Scene: by the window Time: midnight Lighting: moon shadows Camera: close-up Character: woman Action: gazing outside Emotion: longing Sound: falling snow Narration: inner monologue
	Scene: courtyard Time: late night Lighting: lantern + snow glow Camera: long take Character: none Action: swaying lantern in the wind Emotion: solitude Sound: wind and snow Narration: none
	Scene: dreamscape Time: between reality Lighting: hazy glow Camera: slow tracking Character: ethereal woman Action: walking on snow Emotion: detachment Sound: dreamlike music Narration: none
	Scene: sky and earth Time: deep midnight Lighting: fusion of snowlight and moonlight Camera: pull-back Character: none Action: poem appears on screen Emotion: serenity Sound: fading silence Narration: none

4-2. 인터랙티브 플랫폼 프로토타입 디자인

본 연구는 고전시가 인코딩-디코딩 메커니즘을 시각화하고 비전공자도 쉽게 참여할 수 있는 멀티모달 인터랙티브 플랫폼을 구현하였다. 이 플랫폼은 '입력 및 인코딩', '파라미터 설정', '결과 생성 및 최적화'로 이어지는 3단계 순환 구조를 통해 복잡한 AIGC 기술 흐름을 직관적인 그래픽 사용자 인터페이스(GUI)로 시각화하였다 〈그림 3〉.

시스템 아키텍처는 프런트엔드, 미들웨어, 백엔드의 모듈형 구조로 설계되었으며, 그 핵심은 자동 생성 공정마다 사용자가 개입할 수 있는 인간 참여형 개념의 도입이다. 중앙 제어 허브는 음성 복제, 얼굴 변환, 스타일 전이 등 6대 기능 모듈로 구성되어 사용자가 의

[그림 3] 고전시가 입력 및 메인 레이아웃 디자인

미 벡터와 시각 매개변수를 직접 미세 조정할 수 있도록 지원한다.

[그림 4] 멀티모달 파라미터 제어 및 동기화 인터페이스

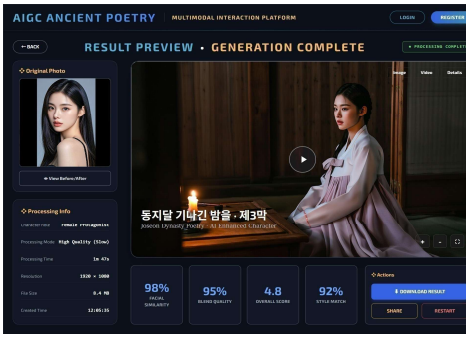
플랫폼 인터페이스는 정보 처리 효율을 위해 고정 비율 분할 레이아웃을 채택하였다 〈그림 4〉. 좌측 입력부는 점선과 아이콘으로 행동 유도성을, 우측 제어부는 슬라이더와 실시간 동기화로 조작 정밀도를 강화하였으며, 하단 액션 바는 시각적 위계를 통해 직관적 과업 수행을 유도한다.

인물 교체 및 스타일 변환 인터페이스 〈그림 5~6〉은 전후 비교 보기 기능과 상세 메타데이터를 제공하여 시스템 상태의 가시성을 확보하였으며, 사용자는 이를 통해 다양한 스타일을 선택하고 변환 강도를 조절하여 〈그림 7〉과 같이 최적화된 결과를 즉시 확인할 수 있다.

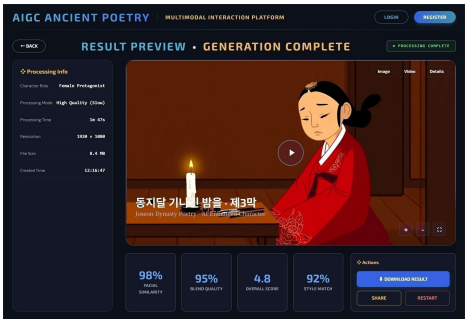
4-3. 실증 결과 평가 및 분석

본 연구는 고전시가 멀티모달 생성 시스템의 생성 결과물 적절성 및 사용자 경험의 합리성을 다각적으로

[그림 5] 스타일 변환 인터페이스



[그림 6] 인물 교체 인터페이스



[그림 7] 최종 멀티모달 결과 출력 및 피드백 인터페이스

검증하기 위해 정성적 및 정량적 평가를 병행하였다. 평가 방법론은 주관적 사용자 평가와 객관적 지표 분석이라는 두 가지 주요 차원에서 설계되었다.

생성 시스템의 실제 사용자 경험과 고전시가 원작과의 일치도를 심층적으로 파악하고, 제안하는 인코딩-디코딩 메커니즘의 유효성을 검증하기 위해 전문적인 다차원 평가 메커니즘을 설계하였다. 설문 문항은 5점 리커트 척도(1: 전혀 그렇지 않다 ~ 5: 매우 그렇다)를 사용하여 응답자의 만족도를 정량적으로 측정하였으며,

다음의 세 가지 전문적인 평가 차원을 중심으로 구성하였다:

원작 충실도 (Content Fidelity)는 원시의 주요 이미지, 주제, 중심 정서가 시각 및 청각 요소에 얼마나 정확하고 명확하게 반영되었는가를 평가한다. 이는 본 연구의 인코딩 메커니즘이 원작의 의미를 얼마나 효과적으로 보존하고 전달하는지를 가늠하는 주요 지표이다.

예술적 재현성 (Artistic Representativeness): 생성된 멀티모달 콘텐츠가 미적 품질, 시각적 구도, 청각적 음색 및 억양의 자연스러움, 그리고 전반적인 분위기의 통일성 측면에서 얼마나 높은 예술적 완성도를 가지는지를 평가한다. 이는 디코딩 메커니즘의 조형적, 음향적 구현 능력을 측정한다.

사용자 상호작용성 (User Interactivity): 플랫폼의 시청각 생성 효과 조정 기능, 실시간 피드백 및 보완 기능(내레이션, 자막 등)이 사용자 경험을 얼마나 효과적으로 향상시키며, 고전시가에 대한 깊이 있는 이해와 몰입을 돕는지를 평가한다.

본 연구에서 제안하는 맞춤형 AIGC 플랫폼의 성능 우위성을 객관적으로 검증하기 위해 A/B 테스트(A/B Testing) 방식을 도입하였다. 평가 참여자는 일반 사용자 60인과 고전시가 및 AI 예술 분야 전문가 5인, 총 65명으로 구성되었다. 일반 사용자 집단은 20대부터 40대까지 다양한 연령층으로 구성되었으며, AIGC 도구 사용 경험이 있는 그룹과 없는 그룹을 균형 있게 배정하여 표본의 편향성을 최소화하였다. 전문가 그룹은 고전문학 연구자 2명, 미디어 인터랙션 디자이너 2명, AI 생성형 모델 개발자 1명으로 구성하여 다각적인 시각을 확보하고자 하였다. 이들은 다음의 두 가지 조건 하에 멀티모달 콘텐츠를 생성하고 평가하였다:

A그룹 (대조군): 기존의 범용적인 순수 AI 플랫폼(DreaminaAI 기본 버전)을 활용하여 콘텐츠 생성.

B그룹 (실험군): 본 연구에서 구축한 고전시가 특화 맞춤형 플랫폼을 활용하여 콘텐츠 생성.

수집된 데이터의 통계적 유의성을 검증하기 위해 통계 분석 소프트웨어(SPSS 26.0)를 활용하였다. 본 실험은 총 65명의 표본(n=65)이 대조군(A그룹: 범용 AI 플랫폼)과 실험군(B그룹: 본 연구의 맞춤형 플랫폼)의 조건을 모두 수행하는 개체 내 설계(Within-subjects design)로 진행되었다. 따라서 두 조건 간의 평가 지표(원작 충실도, 예술적 재현성, 사용자 상호작용성) 평균 차이를 분석하기 위해 대응표본 t-검정(Paired t-test)을 실시하였으며, 통계적 유의수준은 $p < 0.05$로 설정하여 가설을 검증하였다.

평가 대상 시가(詩歌)는 중·한 고전시가 중에서 보편적인 인지도와 다양한 표현 가능성을 지닌 10편을 선정하였다. 각 참여자는 두 그룹의 플랫폼을 모두 사용하여 동일한 10편의 시가에 대한 멀티모달 콘텐츠를 직접 생성하고, 조정 기능을 사용한 후 앞서 정의된 세 가지 평가 차원(원작 충실도, 예술적 재현성, 사용자 상호작용성)에 따라 설문에 응답하였다.

데이터의 신뢰성과 타당성 확보를 위해, 설문 문항은 고전시가 AIGC 창작 관련 문헌 검토 및 전문가 자문을 통해 내용 타당성을 확보하였다. 또한, 평가 참여자들에 대한 사전 교육을 통해 평가 기준에 대한 이해도를 높여 평가자 간 신뢰도를 확보하고자 노력했다. A/B 테스트를 통한 대조군 및 실험군 비교는 측정 도구의 준거 타당성을 간접적으로 강화하며, 본 연구에서 제안하는 시스템의 유효성을 뒷받침한다. 아울러, 본 연구의 핵심 목적인 '의미 전달'과 '상호작용'의 중요도를 고려하여, 전문가 자문을 거쳐 도출된 가중치를 각 평가 항목에 차등 적용함으로써 종합 점수의 객관성을 높였다. 설문조사를 통해 얻은 일반 사용자 그룹과 전문가 그룹의 항목별 가중 평균 점수는 아래 <표 5>와 같다.

[표 5] A/B 테스트 기반 그룹별(A/B) 가중치 적용 평가 결과 비교

평가 항목	세부 평가 항목	일반 사용자 (N=60)		전문가 (N=5)	
		A	B	A	B
원작 충실도 가중치 0.4	시각적 묘사 정확도	3.1	4.3	2.9	4.5
	시각적 정서 전달력	3.3	4.4	2.6	4.1
	청각적 운율 및 정서 표현	2.9	4.1	3.1	4.3
예술적 재현성 가중치 0.3	시각적 미학적 품질	3.3	4.2	2.5	4.1
	청각적 음성 자연스러움	3.1	4.6	2.2	3.5
	멀티모달 통합의 조화로움	3.4	4.5	3.1	3.9
사용자 상호작용성 가중치 0.3	기능 조작의 용이성	3.5	4.6	2.5	4.1
	고전시가 이해 및 몰입 증진	3.6	4.5	3.2	4.2

통계 분석 결과, 일반 사용자 및 전문가 그룹 모두에서 실험군(B그룹)이 대조군(A그룹)에 비해 통계적으로 유의미하게 높은 점수를 기록하였다($p < 0.05$). 구

체적으로 살펴보면, 범용 AI를 활용한 A그룹은 시각적 묘사에서 피상적인 패턴 매칭에 머무른 반면, 제안된 시스템(B그룹)은 시의 장법 구조와 운율을 알고리즘의 파라미터로 직접 연동함으로써 원작 충실도와 예술적 재현성 측면에서 유의미한 성능 향상을 입증하였다. 이는 본 연구가 설정한 다층적 인코딩 메커니즘이 단순한 이미지 생성을 넘어, 사용자가 고전시가의 문화적 심층(기의)에 더욱 깊이 몰입하도록 돕는 유효한 설계임을 시사한다. 특히 일반 사용자 그룹은 원작 충실도와 상호작용성에서 높은 만족도를 보였으며, 전문가 그룹은 예술적 재현성 항목에서 다소 보수적인 평가를 내렸으나, 인코딩된 정서 및 구조의 시각/청각 반영도 측면에서는 맞춤형 플랫폼이 우수하다는 점을 인정하였다.

주관적 평가를 보완하고 제안하는 인코딩-디코딩 메커니즘의 효용성을 정량적으로 입증하기 위해 CLIP Score를 주요 지표로 활용한 A/B 테스트(A/B Test)를 수행하였다. CLIP Score는 이미지와 텍스트 간의 의미론적 유사성을 측정하는 지표로, 생성된 시각 콘텐츠가 원시 텍스트의 의미를 얼마나 정확하게 반영하는지를 객관적으로 평가한다. 본 연구에서는 CLIP Score를 통한 이미지-텍스트 의미론적 일치도 분석의 신뢰성과 직관성을 높이기 위해, 자체 개발한 CLIP 기반 시각화 웹 애플리케이션(AI Image Semantic Evaluation)을 활용하였다. 이 애플리케이션은 업로드된 이미지의 CLIP 특징 벡터를 추출하고 다양한 텍스트 프롬프트와의 유사도를 계산하여 정량적 점수 및 시각화된 데이터로 제공한다.

평가 지표는 생성된 콘텐츠가 원시 텍스트를 얼마나 정확히 반영하는지를 측정하기 위해 주제 부합성, 정서 표현력, 이미지 재현성, 시적 분위기, 문화적 맥락의 5대 차원으로 구성되었으며, 분석 결과는 오각형 방사형 그래프로 시각화하여 제시된다.

비교 실험은 총 20편의 고전시가를 대상으로 본 연구의 구조화된 인코딩을 적용한 그룹 B(실험군)와 원시 텍스트를 그대로 입력한 그룹 A(대조군)로 구분하여 진행되었다. 각 그룹의 콘텐츠에서 추출한 5개의 대표 프레임에 대한 CLIP Score 평균값을 산출하여 성능 차이를 비교 분석하였으며, 측정 결과는 <표 6>과 같다.

[표 6] CLIP Score 기반 그룹 간 이미지-텍스트 의미론적 일치도 비교 분석 결과

지표 (Metric)	그룹 A (원시 텍스트 입력)	그룹 B (본 연구 제안 방식)
평균 CLIP Score	0.23	0.82
표준 편차 (Std Dev)	0.07	0.03

[그림 8] Group A: 원시 텍스트 입력 (대조군 일부)

[그림 9] Group B: 구조화된 인코딩 적용 (제안 모델 일부)

분석 결과, 구조화된 인코딩을 적용한 그룹 B(제안 모델)는 평균 CLIP Score 0.82(표준 편차 0.03)를 기록하여, 단순 텍스트 기반인 그룹 A(대조군: 0.23, 편

차 0.07) 대비 현저한 성능 우위를 보였다. 이러한 결과는 본 시스템의 인코딩-디코딩 메커니즘이 고전시가의 복합적인 의미론적 정보를 시각 콘텐츠로 더욱 정확하고 충실하게 구현함을 객관적으로 보여준다. 나아가, 본 연구의 구조화된 인코딩 규칙이 범용 생성 모델의 한계를 극복하고, 고전시가 고유의 이미지와 정서, 그리고 시적 구조를 시각적으로 정밀하게 제어하는 데 효과적임을 입증한다.

4-4. 생성 결과의 한계 및 오류 분석

본 연구에서 제안하는 시스템은 고전시가 멀티모달 생성의 품질을 향상시켰으나, 이는 단순한 기술적 오류라기보다, 대규모 학습 데이터의 문화적 편향성과 시각 재현 방식의 구조적 제약에서 기인하는 한계로 이해할 수 있다. 다음은 생성 과정에서 나타난 대표적인 오류 사례에 대한 분석이다.

[그림 10] 제안 모델의 시각적 오류 및 한계

'송대 시인' 프롬프트를 입력했음에도 <그림 10A>처럼 서구적 이미지가 생성되는 '문화적 환각'이 나타난다. 이는 모델이 서구·현대 중심의 방대한 데이터로 학습되어 동양 고전 문화 개념을 정확히 분별하지 못하고, 시각적으로 유사한 다른 문화권 이미지를 출력하기 때문이다.이러한 현상은 텍스트-이미지 생성 모델이 학습 데이터의 통계적 분포에 의존한다는 점에서 비롯되며, 특정 문화권의 시각 자료가 상대적으로 과잉 대표(over-representation)되는 반면, 동양 고전 문화 데이터는 저대표(under-representation)되는 구조적 불균형과 관련된다. 이는 고전시가의 시각화 과정에서 문화적 맥락이 왜곡될 가능성을 내포하는 '시각적 본질 제한'으로 작용한다.

또한 여러 이미지를 연결해 영상을 구성할 때 프레임 간 시각적 불일치가 발생한다. <그림 10B>처럼 '눈 덮인 지붕' 장면에서 이전 해변 프레임의 파도 디테일이 잔존해 지붕이 파도처럼 렌더링 되는 오류가 대표적이다. 더불어 AI 모델은 역사적 정확성보다 조형적 완결성을 우선해 시대 배경과 충돌하는 요소를 생성하기도 한다. <그림 10C>처럼 '고대 운하 도시'임에도 현대적 조명과 시설물이 혼재되는 사례가 이에 해당한다.

군중 이미지 생성에서도 <그림 10D>처럼 서로 다른 시대·문화권의 복식이 뒤섞이는 고증 오류가 반복된다. 이는 모델이 특정 시대의 복식 규범과 맥락적 일관성을 충분히 제어하지 못하기 때문이다. 이러한 사례들은 현재 텍스트-이미지 생성 모델이 훈련 데이터의 문화적 편향성으로 인해 문화적·논리적·시공간적 일관성 유지에 근본적 한계를 지님을 보여준다. 비록 기저 학습 데이터에 문화적 편향성이 존재하더라도, 본 연구에서 제안한 다층적 인코딩 메커니즘(B그룹)은 단순 텍스트 입력(A그룹) 대비 이러한 문화적 환각을 통계적으로 유의미하게 억제하고 의미론적 일치도를 대폭 향상시켰다(0.82 vs 0.23). 이는 본 시스템 설계 개입의 필요성과 유효성을 명확히 증명한다. 나아가, 궁극적인 문화적 재현의 정합성을 온전히 확보하기 위해서는 향후 도메인 특화 데이터(Domain-specific Data) 구축을 통한 모델 미세 조정(Fine-tuning)과 사후 고증 프로세스를 통합하는 하이브리드 접근이 후속 과제로 요구된다.

5. 결론

본 연구는 AIGC 기술을 매개로 고전시가의 미학적 체계와 멀티모달 커뮤니케이션 설계를 통합하는 연구 프레임워크를 제안하였다. 디지털 전환기에 인공지능이 고전시가의 텍스트성을 시청각적 실체로 구체화하고, 이를 새로운 문화 향유 방식으로 확장할 수 있는 가능성을 이론적·실증적으로 규명하는 데 연구의 목적을 두었다.

연구는 문헌 고찰 및 이론 체계 구축, 사례 기반 실증 창작, 인터랙티브 플랫폼 프로토타입 설계·구현, A/B 테스트 및 전문가 평가를 통한 타당성 검증의 4단계 연구 설계에 따라 수행되었다. 기술적 실행 차원에서는 '텍스트 분석-스토리보드-시청각 생성-동적 구현-상호작용'의 5단계 AIGC 창작 프로세스를 적용하였다. 신기질(辛棄疾)의 〈청옥안·원석(靑玉案·元夕)〉과 황진이(黃眞伊)의 〈동짓달 기나긴 밤을〉을 실증 대상으로, 텍스트에 내재된 정서와 시적 심상(心象, poetic imagery)이 규칙 기반 매핑과 잠재 확산 모델(Latent Diffusion Model, LDM)을 통해 조형 언어와 음률로 변환되는 과정을 고찰하였다. CLIP Score 기반 정량적

평가 결과, 제안된 구조화된 인코딩 방식은 원시 텍스트 입력 방식에 비해 의미론적 정확성과 일관성을 유의미하게 개선하였다(평균 CLIP Score: 0.82 vs. 0.23).

본 연구의 학술적 의의는 다음 세 가지 측면에서 제시될 수 있다. 첫째, 이론적 측면에서 언어학·기호학·생성 AI 이론을 접목하여, 고전시가의 구조를 기계 연산이 가능한 기호 데이터로 치환하고 이를 다시 인간의 감각적 경험으로 환원하는 멀티모달 인코딩-디코딩 프레임워크를 정립하였다. 둘째, 방법론적 측면에서 대형 언어 모델(LLM)과 영상 생성 모델을 연결하는 파이프라인을 구축하여, 시적 정서의 다층적 시청각 재구성을 위한 구체적 기술 경로를 제안하였다. 셋째, 응용적 측면에서 인간 참여형(Human-in-the-Loop) 구조가 적용된 인터랙티브 플랫폼 프로토타입을 구현함으로써, 사용자가 창의적 주도권을 갖고 AI와 협업하는 새로운 고전시가 창작 모델의 실천적 가능성을 제시하였다.

다만 본 연구는 다음과 같은 한계를 내포하고 있으며, 이는 후속 연구의 과제로 남겨진다. 첫째, 인코딩 규칙이 중·한 고전시가 체계를 중심으로 설계되어, 여타 언어권의 상이한 운율 구조와 문화적 은유를 포괄하는 데 제약이 있다. 둘째, 시·공간적 제약으로 인해 다양한 연령·문화적 배경의 대규모 사용자 집단을 대상으로 한 광범위한 실증 검증이 이루어지지 못하였다. 셋째, LDM 기반 디코딩은 구체적 시각 묘사에는 효과적이었으나, 고도로 추상적인 철학적 개념이나 중의적(重意的) 표현의 시각화 과정에서 의미 축소 또는 왜곡이 발생할 수 있다. 넷째, CLIP Score는 이미지-텍스트 간 의미론적 유사성을 정량화하는 데 유효하나, 예술 작품의 주관적 미감과 정서적 깊이를 수치로 온전히 포착하는 데에는 본질적 한계가 있다.

이를 극복하기 위한 후속 연구의 방향은 세 가지로 제시된다. 첫째, 멀티모달 코퍼스를 다국어 및 다양한 문학 장르로 확충하여, AI의 문화 맥락 해석 능력을 범용적으로 고도화해야 한다. 둘째, 다양한 사용자 집단을 대상으로 한 광범위한 실증 연구를 수행하고, 정성적·정량적 평가가 상호 보완되는 다차원적 평가 모델을 개발해야 한다. 셋째, 현재의 프로토타입을 발전시켜 사용자 피드백 기반의 적응형 생성 모델(Adaptive Generative Model)을 구축함으로써, 글로벌 디지털 환경에서 고전시가의 지속 가능한 확산을 실현하고자 한다.

참고문헌

1. 강병규, ‘AI의 중국 古典詩歌 창작 —詩語의 학습과 생성’, 중국문학, 2019.
2. 김선영, ‘대학교육에서 멀티모달 기반 생성형 AI를 활용한 콘텐츠 제작 프로그램 사례 연구’, 한국콘텐츠학회논문지, 2024.
3. 박라미, 유진, 최유리, 오효정, ‘AI를 활용한 멀티모달 분석 기반 디자인 언어 이해: 가우디 건축 사례를 중심으로’, 디지털콘텐츠학회논문지, 2025.
4. 손주현, 박진숙, ‘아이덴티티디자인 커뮤니케이션 과정 중의 오류 발생원인에 관한 연구’, Archives of Design Research, 2012.
5. 송영훈, 김남기, 정경용, ‘객체 탐지와 텍스트 융합 기반 멀티모달 인터랙티브 시스템’, 한국정보기술학회논문지, 2025.
6. Cai Xinyuan, Chen Qiuchan, Zhang Jian, ‘AI gongchuang de Zhongguo gushi wenhua VR tiyan xitong sheji yu shixian’, PUBLISHING JOURNAL, 2023.
7. Guo Zhipeng, Xiaoyuan Yi, Maosong Sun, Wenhao Li, Cheng Yang, Jiannan Liang, Huimin Chen, Yuhui Zhang, Ruoyu Li, 『Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations』, Association for Computational Linguistics, 2019.
8. Hall, S., 『Culture, Media, Language』, Hutchinson, 1980.
9. Jianli Zhao, Hyo Jong Lee, ‘Automatic Generation and Evaluation of Chinese Classical Poetry with Attention-Based Deep Neural Network’, Applied Sciences, 2022.
10. Liwen Yang, Zhidong Zhang, Kaipeng Niu, Sitian Pan, Weiping Zhu, Chao Ma, 『2024 IEEE Smart World Congress (SWC)』, 2024.
11. Siyi Zhou, Yiquan Zhou, Yi He, Xun Zhou, Jinchao Wang, Wei Deng, Jingchen Shu, ‘IndexTTS2: A Breakthrough in Emotionally

Expressive and Duration-Controlled Auto-Regressive Zero-Shot Text-to-Speech', arXiv, 2025.

12. Zheng, Junlin, 'Research on Visualization Strategies for Classical Poetry Short Videos Based on AI Technology: A Case Study of 'Ode to the Ages", New Media Research, 2025.
13. www.chinaqw.com
14. mp.weixin.qq.com