

생성형 AI 기반 K-판타지 애니메이션 제작 파이프라인 연구

이영화¹, 김다루², 권오주³, 문채영⁴, 박구만⁵, 최영희^{6*}

¹서울과학기술대학교 나노IT융합디자인대학원 정보통신미디어공학과 학생, ²서울과학기술대학교 스마트ICT융합공학과 학생,

³서울과학기술대학교 국방융합과학대학원 정보통신미디어공학과 학생, ⁴서울과학기술대학교 문예창작학과 학생,

⁵서울과학기술대학교 스마트ICT융합공학과 교수, ⁶서울과학기술대학교 문예창작학과 교수

A Study on K-Fantasy Animation Production Pipeline Based on Generative AI

Younghwa Lee¹, Daru Kim², O JU Kwon³, Chaeyoung Moon⁴,
Gooman Park⁵, Younghee Choi^{6*}

¹Student, Dept. of Information Technology and Media Engineering, The graduate School of Nano IT Design Fusion, Seoul National University of Science and Technology

²Student, Dept. of Smart ICT Convergence Engineering, Seoul National University of Science and Technology

³Student, Dept. of Information and Communication Media Engineering, Graduate School of Defense Convergence Science, Seoul National University of Science and Technology

⁴Student, Dept. of Creative Writing, Seoul National University of Science and Technology

⁵Professor, Dept. of Smart ICT Convergence Engineering, Seoul National University of Science and Technology

⁶Professor, Dept. of Creative Writing, Seoul National University of Science and Technology

요약 본 연구는 생성형 인공지능을 활용한 K-판타지 애니메이션 제작 파이프라인을 제안하고, 단편 애니메이션 「괴국의 안리수」로 실효성을 검증한다. 전통적 제작 방식은 높은 비용과 인력 의존도로 독립 창작자와 소규모 제작사에게 높은 진입장벽이 된다. 이를 해결하기 위해 Midjourney·Runway Gen-4·Higgsfield 등 상용 AI 도구와 LoRA 파인튜닝 기법을 결합한 통합 파이프라인을 구성하였다. 캐릭터와 배경 모델을 분리 파인튜닝하는 전략으로 프레임 간 일관성을 확보하고, 한국 전통 건축의 구조적 비례와 괴이국 특유의 색채 체계를 반영한 전용 데이터셋을 구축해 기존 AI의 한국적 미감 표현 한계를 극복하였다. 전문가 평가에서 시각적 완성도, 연출 일관성, 미학적 표현력, 제작 효율성에서 긍정적인 결과를 확인하였으며, 전통 방식 대비 제작 기간과 인력이 절감되어 생성형 AI 기반 K-콘텐츠 제작의 실질적 대안 가능성을 입증하였다.

주제어 : 생성형 인공지능, K-판타지, LoRA 파인튜닝, 캐릭터 일관성, 한국적 미감, AI 콘텐츠 제작 파이프라인

Abstract This study proposes a K-fantasy animation production pipeline utilizing generative AI and verifies its effectiveness through the short animation "Anrisu in Wonderland." Traditional production methods pose high entry barriers for independent creators and small studios due to their high costs and heavy reliance on human labor. To address this, an integrated pipeline combining commercial AI tools like Midjourney, Runway Gen-4, and Higgsfield with LoRA fine-tuning techniques was developed. A strategy of separately fine-tuning character and background models ensured consistency between frames. A dedicated dataset reflecting the structural proportions of traditional Korean architecture and the unique color system of the "Weird Country" was constructed, overcoming existing AI limitations in expressing Korean aesthetics. Expert evaluations confirmed positive results in visual completeness, production consistency, aesthetic expressiveness, and production efficiency. Compared to traditional methods, this approach reduced production time and manpower, proving its potential as a practical alternative for generative AI-based K-content production.

Key Words : Generative AI, K-Fantasy, LoRA Fine-tuning, Character Consistency, Korean Aesthetic Sensibility, AI contents Production Pipeline

이 연구는 서울과학기술대학교 교내 일반과제 연구비 지원으로 수행되었습니다.

*교신저자 : 최영희(chanwch@seoultech.ac.kr)

접수일 2025년 12월 24일 수정일 2026년 02월 06일 심사완료일 2026년 02월 19일

1. 서론

OTT 플랫폼의 급성장과 디지털 미디어 환경의 확산을 배경으로, K-드라마·K-팝·K-웹툰에 이어 K-애니메이션까지 K-콘텐츠가 글로벌 시장에서 두드러진 성과를 거두고 있다[1]. 특히 한국 전통 설화, 무속 신앙, 고유 세계관을 현대적으로 재해석한 'K-판타지' 장르의 독창적 소재와 높은 서사적 잠재력을 바탕으로 국제 시장에서 차별화된 경쟁력을 지닌다. 그러나 이러한 콘텐츠를 글로벌 수준의 완성도로 시각화하기 위해서는 고도화된 연출력과 기술 역량이 동시에 요구된다.

문제는 전통적인 애니메이션 제작 방식이 이러한 수요를 감당하기 어렵다는 구조적 한계에 있다. 기획, 작화, 채색, 촬영에 이르는 전 과정이 노동 집약적으로 운영되며, 분 단위 콘텐츠 제작에도 막대한 인력·시간·비용이 소요된다[2]. 이는 자본력이 부족한 독립 창작자와 중소기업 제작사에게 높은 진입장벽으로 작용하며, 빠르게 변화하는 시장 수요에 민감하게 대응하기 어렵게 만드는 주요 원인이 되고 있다[3].

이러한 제작 환경의 구조적 비효율성을 해소할 대안으로, 최근 생성형 인공지능을 활용한 새로운 제작 패러다임이 부상하고 있다[4]. 확산 모델(Diffusion Model)[5]의 발전에 힘입어 Midjourney[6], DALL-E 3[7], Stable Diffusion[8] 등 텍스트-투-이미지(Text-to-Image, T2I) 도구가 상용화되었고, Runway Gen-4[9], Higgsfield[10], Kling[11], OpenAI Sora[12], Google Veo3[13] 등 영상 생성 AI의 등장으로 프롬프트 기반의 영상 제작이 현실화되었다. 이러한 도구들은 창작의 진입장벽을 낮추고 제작 효율을 획기적으로 향상시킬 잠재력을 지닌다.

그러나 기존의 범용 생성형 AI 모델을 K-판타지 애니메이션 제작에 그대로 적용하기에는 두 가지 핵심적인 기술적·표현적 한계가 존재한다.

첫째, 캐릭터 일관성 문제이다. 애니메이션은 연속된 프레임 전체에서 동일 캐릭터의 외형·색채·표정이 일관되게 유지되어야 한다. 그러나 상용 AI 도구는 동일한 프롬프트를 반복 입력하더라도 프레임마다 캐릭터의 얼굴 형태나 의상이 미묘하게 변화하는 현상이 발생하며, 이는 애니메이션의 서사적 몰입을 저해하는 근본적인 문제로 작용한다.

둘째, 한국적 미감 표현의 한계이다. 생성형 AI 모델들은 주로 서구 중심의 대규모 데이터셋으로 학습되어, 한국 전통 복식, 건축 양식, 오방색 색채 체계 등 한국 고유의 미학적 요소를 정확하고 지속적으로 재현하는 데

뚜렷한 제약이 있다. 한국 전통 건축의 처마 곡선이나 한옥의 비례 체계와 같은 요소는 생성 과정에서 왜곡되거나 중국·일본 건축 양식과 혼용되는 경우가 빈번히 관찰된다.

이에 본 연구는 이러한 한계를 극복하기 위한 K-판타지 애니메이션 전용 제작 파이프라인을 제안한다. 캐릭터 모델과 배경 모델을 독립적으로 학습하는 분리형 LoRA(Low-Rank Adaptation)[14] 파인튜닝 전략과 Midjourney, Runway Gen-4, Higgsfield 등 상용 도구를 유기적으로 결합하여, 캐릭터 일관성과 한국적 미감 표현 두 가지 과제를 동시에 달성하는 방법론을 제시한다.

제안한 파이프라인의 실효성은 K-판타지 단편 애니메이션 「괴이국의 안리수」의 직접 기획·제작을 통해 실증적으로 검증하였다. 캐릭터 일관성은 애니메이션 캐릭터 특성에 특화된 CCIP(Contrastive Character Image Pretraining) 지표로 정량 평가하였으며, 영상 제작 전문가 12인을 대상으로 한 정성 평가에서 시각적 완성도, 연출 일관성, 미학적 표현력, 제작 효율성 전반에 걸쳐 평균 4.37점(5점 만점)의 긍정적 결과를 확인하였다. 또한 전통 제작 방식 대비 제작 기간 약 64% 단축, 참여 인력 50% 절감이라는 효율성 향상을 정량적으로 입증하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 생성형 AI 이미지·영상 생성 모델의 기술적 원리와 LoRA 파인튜닝 기법에 대한 이론적 배경을 제시한다. 3장에서는 캐릭터·배경 분리형 파인튜닝 전략과 캐릭터 일관성 평가 방법론을 기술한다. 4장에서는 「괴이국의 안리수」 제작 사례를 통해 AI 기반 콘텐츠 제작 전체 파이프라인을 실증적으로 분석한다. 5장에서는 전문가 집단 정성 평가 결과를 분석하고, 6장에서는 전통 제작 방식과의 효율성 비교를 제시한다. 마지막으로 7장의 결론에서 연구 결과의 시사점과 향후 연구 방향을 논의한다.

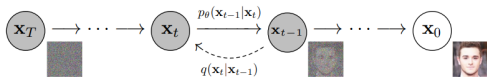
2. 이론적 배경

2.1 생성형 AI와 이미지/영상 생성

생성형 인공지능은 학습한 데이터의 패턴을 바탕으로 새로운 콘텐츠를 자동으로 생성하는 AI 기술이다[14]. 이 중 텍스트 프롬프트를 입력하여 이미지나 영상을 생성하는 텍스트-투-이미지(Text-to-Image, T2I)와 텍스트-투-영상(Text-to-Video, T2V) 모델들이 최근 급속

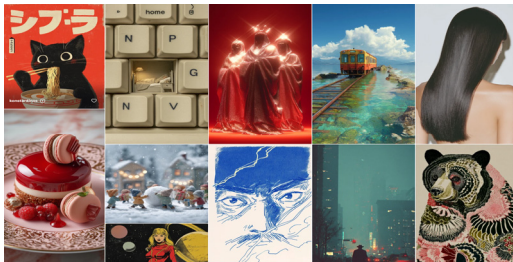
히 발전하고 있다.

텍스트-투-이미지 생성의 기술적 기초는 확산 모델(Diffusion Model)에 있다. [그림 1]과 같이 이 모델은 노이즈로부터 시작하여 단계적으로 노이즈를 제거하며 이미지를 합성하는 방식으로 작동한다. 사용자가 입력한 텍스트 프롬프트는 CLIP[15] 같은 텍스트 임베딩 모델을 통해 벡터로 변환되고, 이 벡터를 조건으로 하여 이미지 생성 모델이 작동한다.



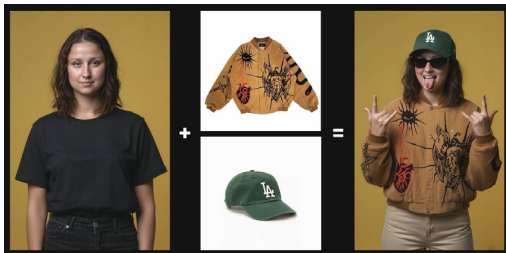
[Fig. 1] Conceptual diagram of the diffusion model

대표적인 상용 T2I 모델로는 Midjourney, DALL-E 3, Stable Diffusion, NanoBanana 등이 있다. Midjourney는 미적 완성도와 스타일 일관성으로 평가 받고 있으며, [그림 2]와 같이 이미지 제작에 폭넓게 활용되고 있다.



[Fig. 2] Image generated by Midjourney

NanoBanana는 구글 딥마인드의 제미니 2.5(Gemini 2.5) 플랫폼에 통합된 멀티모달 모델로, [그림 3]과 같이 기존 이미지의 스타일과 구도를 유지하면서 특정 인물이나 장면을 일관되게 편집 또는 변형하는데 특화되어 있다.



[Fig. 3] Example of NanoBanana image synthesis

텍스트-투-영상 생성은 T2V 기술을 시간 차원으로 확장한 것으로, 3D 시공간(spatio-temporal) 어텐션 메커니즘을 활용하여 프레임 간 시간적 일관성을 유지한다 [16]. 대표적인 상용 도구로는 OpenAI의 SORA, Runway Gen-4, Kling, Higgsfield, Midjourney, Google Veo3 등이 있다.

Runway Gen-4의 경우 aleph 기능을 통해 하나의 이미지로 여러 시점의 장면을 생성할 수 있으며, Act-Two 기능과 립싱크(lip-sync) 기능을 통해 인물의 대화와 입 모양을 동기화할 수 있어 캐릭터 대화 장면 제작에 활용된다.

Higgsfield는 역동적인 카메라 워크와 인물 동작 표현에 강점을 지닌 영상 생성 플랫폼이다. 모션 템플릿 기능을 활용하면 카메라 이동에 따른 명암 대비를 자연스럽게 유지할 수 있으며, 피사계 심도(Depth of Field)와 렌즈 왜곡(Optical Distortion) 제어를 통해 시각적 깊이감을 강화할 수 있다. 팬(pan), 틸트(tilt), 줌(zoom), 돌리(dolly) 등 영화 촬영 기법을 구현하여 시네마틱 영상 제작이 가능하다.

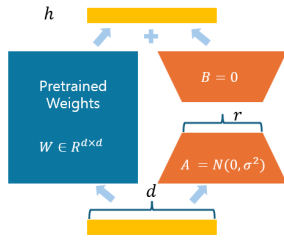
Midjourney의 영상 생성은 생성된 이미지를 기반으로 5초 내외의 짧은 영상을 만들어주며, 지속모션과 고속모션으로 움직임 제어를 할 수 있다. 또한 영상확장(extend) 기능을 통해 영상의 길이를 연장할 수 있다.

Google Veo3는 Transformer[17] 구조의 확산모델로, 오디오까지 통합되어 생성되며, Flow를 통해 최대 148초 길이의 영상까지 확장할 수 있다.

2.2 파인튜닝 기법

기존의 대규모 생성형 AI 모델은 일반적인 패턴만 학습하기 때문에, 특정 캐릭터나 한국적 미감을 지속해서 재현하는 데 한계가 있다. 이를 극복하기 위해 파인 튜닝(Fine-tuning)기법이 활용된다. 파인튜닝 기법이란 소규모의 해당 캐릭터 데이터나 원하는 스타일 데이터를 이용해 모델을 추가 학습시킴으로써 특정 속성에 모델을 맞춤화하는 방법이다. DreamBooth[18], LoRA, StyleAlign[19] 등의 방법이 있으며, 그중에서 LoRA는 대규모 모델의 특정 부분만 소규모 매개변수로 조정하는 경량 파인튜닝 기법이다. [그림 4]의 LoRA 기본 원리에서 볼 수 있듯 기본 아이디어는 모델의 원본 가중치를 전부 업데이트하지 않고, 대신 일부 계층에 소수의 저차원 행렬을 덧붙여 학습시키는 것이다. 이렇게 하면 학습시켜야 할 파라미터 수가 크게 줄어들고, 훈련 속도와 메모리 효율이 향상된다. 본 연구에서는 LoRA 파인튜닝 기

법을 활용하여 시각적 특성을 독립적으로 제어할 수 있도록 설계하였다.



[Fig. 4] The fundamental principles of LoRA

2.3 생성형 AI 기반 영상/애니메이션 파이프라인

최근 텍스트-투-이미지(T2I), 텍스트-투-비디오(T2V) 확산 모델을 활용해, 시나리오 작성-장면별 이미지/영상 생성-편집-후반 작업으로 이어지는 생성형 AI 기반 영상 파이프라인이 제안되고 있다. 기존 연구는 주로 여러 샷에 걸쳐 색채·라이팅·카메라 구도와 캐릭터 외형을 일정 수준 유지할 수 있도록, 프롬프트 템플릿, 레퍼런스 이미지, 제한적 파인튜닝을 조합한 워크플로를 설계하는 데 초점을 두고 있다[20]. 그러나 이러한 접근은 특정 장르나 문화적 미감보다는 일반적인 영화·영상 제작 현장 적용에 중점을 두며, 상용 T2I/T2V 도구와 LoRA를 통합해 K-판타지 장르에 특화된 애니메이션 제작 파이프라인을 구성하고 전문가 평가로 그 효과를 검증한 사례는 드물다. 본 연구는 Midjourney, Runway Gen-4, Higgsfield와 LoRA 파인튜닝을 연계한 파이프라인을 통해, 캐릭터·배경 일관성과 한국적 미감 표현을 동시에 달성하는 K-판타지 애니메이션 제작 절차를 제안한다는 점에서 이러한 선행 연구들과 차별성을 지닌다.

3. 제안하는 기법

본 연구에서는 Midjourney v7을 기본 생성 플랫폼으로 설정하고, Flux_dev[21] 기반의 LoRA 파인튜닝 모델을 병행 적용하였다. 이러한 결합을 통해 프롬프트의 언어적 지시와 학습된 스타일 정보가 상호 보완적으로 작용하도록 설계하였다.

3.1 LoRA 파인튜닝 기법

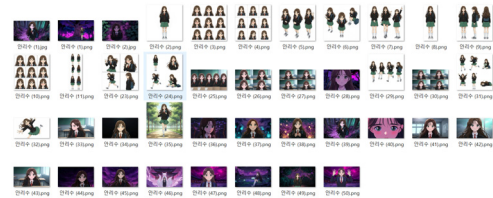
3.1.1 캐릭터용 파인튜닝 모델 데이터셋

캐릭터 일관성을 위한 캐릭터 파인튜닝용 데이터셋은

주인공 안리수의 외형, 표정, 복장, 조명 톤의 일관성 확보를 목표로 <표 1>과 같은 구성 요소로 구축하였다. [그림 5]와 같이 총 50장의 이미지가 사용되었으며, Midjourney 기반 프롬프트로 생성된 시안과 생성된 이미지 중 제작진의 수정·보정 작업이 수행된 이미지를 포함하였다.

<Table 1> Anrisu Character Dataset Structure

Category	Number of images	Key components
face	29	Basic expression, no emotion
Expression of emotions	5	Expressing various emotions such as surprise, fear, determination, and smile
Upper body full body	7	Various postures (jumping, running, sitting, raising hands, etc.)
Lighting changes	9	Natural light, strange purple lighting



[Fig. 5] Dataset image for Anrisu character

3.1.2 배경용 파인튜닝 모델 데이터셋

한국적 미감이 드러나는 배경 생성을 위해 배경 파인튜닝용 데이터셋은 한국 전통 건축의 구조적 비례와 외국 특유의 색채 체계 재현을 목표로 구축하였다. 한국 전통 건축은 자연과의 조화를 중시하는 배치 원리 위에 기단(基壇), 기둥, 공포(栱包), 지붕이라는 위계적 구조를 형성한다. 특히 처마의 완만한 곡선과 추녀 끝의 반전(反轉)은 한국 전통 건축을 중국·일본 건축과 구별 짓는 고유한 특징이라고 할 수 있다.

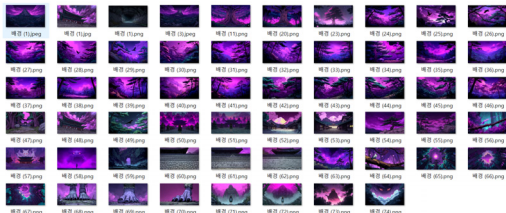
괴이국의 공간은 '현실과 환상의 경계'라는 서사적 설정을 가지므로, 전통 양식의 실제적 재현에서 나아가 초현실적 색채와 조명 변형을 함께 포함하여 데이터셋을 구성하였다. 괴이국의 색채 체계는 전통 오방색을 기반으로 하되, 그 채도와 명도를 신비롭고 몽환적인 느낌이 들도록 하기 위해 의도적으로 강한 마젠타 톤으로 왜곡하는 방식으로 설계되었다. 괴이국 배경용 데이터셋 구조는 <표 2>와 같다.

[그림 6]과 같이 총 74장의 이미지가 사용되었으며 Midjourney 기반 프롬프트로 생성된 시안과 생성된 이

미지를 제작진이 수정·보정한 이미지를 포함하여 구축하였다.

〈Table 2〉 Background Dataset Structure

Category	Number of images	Key components
Hanok Structure	15	Roofline, eave curve, traditional Korean house style
Korean pine	7	Blue, Red, Yellow, Green Saturation Contrast
Forest of the Strange Kingdom	45	grain direction, reflected light
Monster Appears Background	7	illumination structure, contrast ratio



〔Fig. 6〕 Kwaeiguk Background Use Dataset Images

3.1.3 파인튜닝 모델의 파라미터 설정

본 연구에서의 파인튜닝 모델은 Replicate[22] API를 통해 flux_dev의 사전 가중치 위에서 LoRA 구조를 이용하여 독립적으로 훈련하였다. 파인튜닝은 Replicate 클라우드 환경에서 수행되었으며, GPU는 NVIDIA A100 (40GB VRAM) 인스턴스를 사용하였다. LoRA의 경우 학습률은 1×10^{-4} , 스텝 수는 1000, LoRA rank는 16, optimizer는 adamw8bit로 설정하였다.

3.2 파인튜닝 모델의 캐릭터 일관성 성능 평가

일반적으로 캐릭터 일관성 지표로 사용하는 ArcFace[23]는 실사 인물 얼굴 데이터셋으로 학습되었기 때문에, 애니메이션 캐릭터 특유의 양식화된 얼굴 특징(큰 눈, 단순화된 코와 입, 비현실적 비율 등)을 효과적으로 인코딩하지 못할 가능성이 있다. 따라서 본 연구에서는 캐릭터 일관성 유지의 성능 평가를 위해 CCIP(Contrastive Character Image Pretraining)[24] Distance를 사용하였다.

CCIP는 CLIP에서 영감을 받아 개발된 대조 학습 기반 모델로, 애니메이션 캐릭터 이미지 간의 정렬에 특화되어 있다. 약 3,982명의 캐릭터로부터 수집된 약 240,000장의 이미지로 학습되었으며, CCIP의 출력값은 비유사도(Distance) 점수로, 값이 낮을수록 두 이미지의

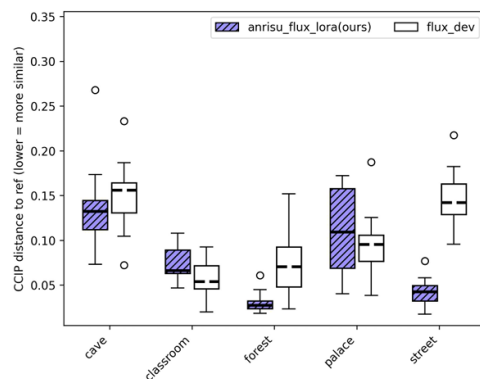
캐릭터가 동일 인물일 가능성이 높다. flux_dev 모델에 LoRA기법을 적용한 Ours 모델과 LoRA 모델을 적용하지 않은 Baseline 모델을 비교 평가하였다. 총 5개의 씬, 씬 당 20장의 데이터셋을 생성하여, 총 100장의 테스트 셋을 구축하여 실험을 진행하였다.

〈표 3〉은 LoRA 모델을 사용한 ours 모델과 LoRA 모델을 적용하지 않은 Baseline 모델의 씬별 측정 결과를 정리한 표이다.

〈Table 3〉 CCIP Face Distance

Scene	Proposed (Ours)	Baseline (flux_dev)
cave	0.1360	0.1507
classroom	0.0731	0.0578
forest	0.0296	0.0752
palace	0.1089	0.0934
street	0.0419	0.1465
ALL	0.0779	0.1047

〔그림 7〕은 CCIP 결과를 씬별 박스플롯으로 나타낸 그래프이다. forest, street 이 두 씬은 K-판타지 애니메이션의 핵심 배경으로, 캐릭터 LoRA가 해당 맥락에서 특히 효과적으로 작동함을 보여준다. 반면 classroom 씬에서 Baseline 모델이 더 좋은 성능을 보이는 것으로 나타나고 있다. 이는 classroom이 현대적이고 일상적인 환경으로, 범용 모델이 학습한 데이터 분포에 가깝기 때문으로 해석할 수 있다.



〔Fig. 7〕 CCIP segmented character cosine similarity box plot

〈표 4〉는 Ours 모델과 Baseline 모델의 이미지 예시이다. LoRA기법을 적용한 Ours 모델로 생성한 이미지에서 캐릭터가 명확히 유지되는 것을 확인할 수 있다.

<Table 4> Test data image

Scene	Proposed (Ours)	Baseline (flux_dev)
cave		
classroom		
forest		
palace		
street		

4. 제작 파이프라인

4.1 「괴이국의 안리수」 작품 개요

「괴이국의 안리수」는 루이스 캐럴(Lewis Carroll)의 「이상한 나라의 앨리스(Alice's Adventures in Wonderland)」[25]를 모티브로 하여, 한국 전통 설화와 민담 속 괴물 및 신화를 현대적 감각으로 재해석한 한국형 판타지 단편 애니메이션이다. 본 작품은 현실과 환상의 경계를 넘나드는 주인공 '안리수'의 시점을 통해, 일상의 공간이 초현실적 세계로 확장되는 과정을 그린다. 이야기는 중학생 안리수가 시험을 망치고 실의에 빠진 채 한강공원 산책로를 걷다 우연히 '금색 고양이'를 쫓아 터널 속으로 들어가며 시작된다. 터널은 현실과 비현실의 경계를 잇는 매개로 작용하며, 이곳을 통과한 안리수는 괴물과 요괴가 공존하는 이계(異界) - 「괴이국(怪異國)」에 진입하게 된다. 그곳에는 한국 전통 설화에 등장하는 도깨비불, 어둡서니, 천년호 등우나 괴물들이 존재한다. [그림 8]은 작품에 등장하는 캐릭터들의 이미지이다.

작품의 서사는 전통 설화의 '이계 체험 구조'를 따르

면서, 현대 도시적 정서를 접목한다. 강변 산책로, 휴대폰, 동네 고양이 등 일상적 사물이 괴이국에 들어서면 각각의 정령적 존재로 변형되며, 이는 디지털 사회 속 개인의 내면을 투사하는 장치로 기능한다. 주인공이 만나는 괴물들은 단순한 공포의 존재가 아니라, 두려움과 외로움을 품은 또 다른 자아로 묘사된다. 작품의 결말에서 안리수는 현실로 돌아오지만, 경험의 흔적은 일상의 사물 속에 남으며 현실과 환상의 경계가 완전히 사라지지 않은 채 열린 결말로 마무리된다.

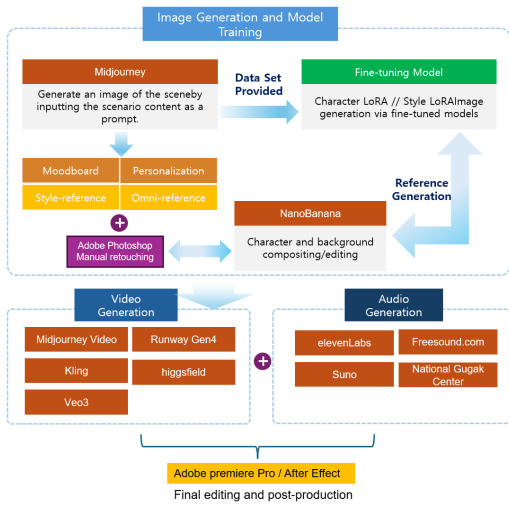


[Fig. 8] Characters in 'Anrisu in Wonderland'

4.2 전체 제작 파이프라인 구조

본 연구에서 구현한 「괴이국의 안리수」의 제작 파이프라인은 5단계로 구성된다. 이 파이프라인은 텍스트 → 이미지 → 영상 → 사운드 → 최종 통합 출력의 순서로 구성되어 있다.

AI 이미지 생성 단계에서는 Midjourney를 기반으로 한 인물·배경 이미지 생성을 수행하고, Flux 기반의 파인 튜닝 모델(LoRA)을 적용하여 캐릭터 일관성과 한국적 미감의 배경 일관성을 확보하였다. 일관성을 확보한 이미지는 Midjourney의 레퍼런스 기능을 활용하여 재생성하였으며, NanoBanana를 통해 캐릭터와 배경을 합성, 편집하는 과정을 거치며 이야기를 이끌어가는 이미지를 생성하였다. 영상 생성 단계에서는 Midjourney Video, Runway Gen-4, Higgsfield 등의 상용도구를 이용하여 장면별 모션 시퀀스를 생성하고 카메라 워크, 조명, 심도 등을 제어하였다. 편집 및 후보정 단계에서는 Adobe Premiere Pro[26]와 After Effects[27]를 사용하여 컷 편집, 색 보정, 시각 효과 합성을 진행하였다. 사운드 디자인 및 믹싱 단계에서는 ElevenLabs[28] 기반 음성합성과 SUNO[29]의 음악 생성 기능을 활용해 배경음 및 효과음을 제작하였다. [그림 9]를 통해 AI 기반 콘텐츠 제작 전체 파이프라인을 확인할 수 있다.



[Fig. 9] AI-Based Content Creation Full Pipeline

4.3 AI 이미지 생성

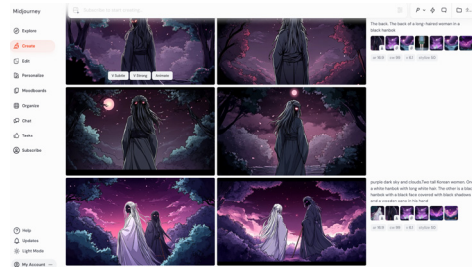
본 연구에서는 Midjourney v7을 기본 생성 플랫폼으로 설정하고, Flux_dev 기반의 LoRA 파인튜닝 모델을 병행 적용하였다. 이러한 결합을 통해 프롬프트의 언어적 지시와 학습된 스타일 정보가 상호 보완적으로 작용하도록 설계하였다.

Midjourney의 옴니 레퍼런스를 활용하여 주인공 캐릭터의 얼굴 형태와 표정의 일관성을 유지하였으며, [그림 10]과 같이 Midjourney Edit의 Generative Fill 및 Inpainting을 사용하여 수작업으로 배경의 노이즈나 단절된 오브젝트를 자연스럽게 보완하였다.



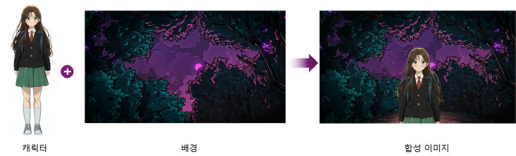
[Fig. 10] Image before and after applying Midjourney Edit

참고 이미지의 색감, 질감, 조명, 분위기 등 시각적 스타일을 추출해 새로운 이미지에 일관되게 적용하는 Midjourney의 스타일 레퍼런스 기능을 활용하여 침부한 기존 이미지와 유사한 이미지를 생성하여 다양한 구도로 활용하였다. [그림 11]은 미드저니 이미지 생성에 레퍼런스 이미지들이 적용된 사례이다.



[Fig. 11] Image Generation Using Midjourney Option Features

마지막으로 파인튜닝을 통해 확보된 인물-배경 각각의 시각적 일관성은 이후 NanoBanana 플랫폼에서 통합, 재생성되었다. NanoBanana를 활용하여 [그림 12]와 같이 생성된 인물 이미지와 배경 이미지를 분리 입력받아 색조 조정, 조명 톤 매칭, 공간 구도 정렬 등을 자동으로 수행하여 이미지를 생성하였다.

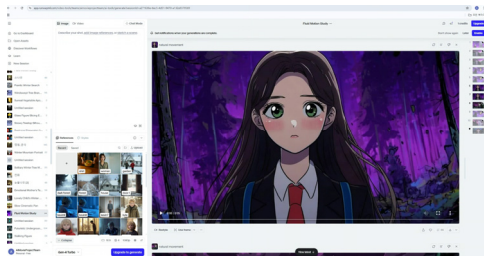


[Fig. 12] Example of NanoBanana Synthesis

결과적으로 본 연구의 모델 선택 전략은 캐릭터와 배경을 통합적으로 학습시키는 대신, 각각 특화된 파인튜닝 모델을 생성해 독립적 품질을 확보한 뒤 후처리 합성 단계에서 두 요소의 감정선과 색채 조화를 완성하는 방식으로 설정되었다.

4.4 AI 영상 생성

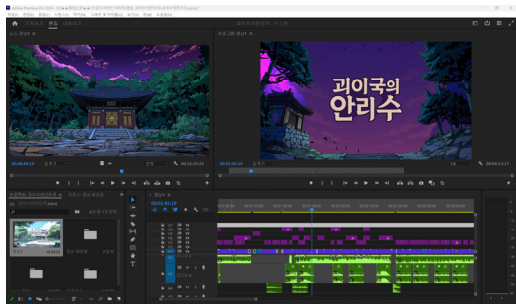
ChatGPT를 활용하여 각 장면의 핵심 서사 단위를 대화형 스크립트 형태로 변환한 후, Midjourney Video, Runway Gen-4, higgsfield, Veo3 등 영상생성 모델의 프롬프트 입력값으로 사용하였다. 프롬프트에는 인물의 동작, 시선, 감정선, 카메라의 이동 방향과 렌즈 속성이 포함되었으며, AI가 자동으로 장면의 흐름을 해석하여 키프레임 간의 자연스러운 전환을 생성할 수 있도록 구성되었다. 특히 Runway Gen-4의 Acttwo기능과 립싱크 기능은 인물의 대사를 처리하기 위해 사용하였고, higgsfield는 다양하고 다이내믹한 카메라 워킹을 통해 '괴이국'에서의 긴장감과 인물의 심리 표현을 강화하였다. [그림 13]은 Runway Gen4를 활용한 영상생성 과정을 보여준다.



[Fig. 13] Video Generation Process Using Runway Gen4

4.5 편집 및 사운드 믹싱

본 단계에서는 Adobe Premiere Pro, After Effects, Topaz Video Enhance AI 등의 후처리 도구를 사용하여 컷 편집, 색보정(Color Grading), 시각 효과(VFX) 및 업스케일링을 수행하였다. [그림 14]는 Adobe Premiere Pro 편집 장면이다.



[Fig. 14] Adobe Premiere Pro editing scene

인물의 대사 음성은 ElevenLabs Voice AI를 통해 합성하였다. 주인공 안리수는 밝고 명료한 여성 음색으로 설정되었으며, 금묘아기와 수전박사는 각각 장난스럽고 전자음이 섞인 음색으로 설계되었다. AI가 생성한 음성은 Adobe Premiere Pro에서 채널별로 합성하였다.

배경음악은 한국 전통 음악, 국악을 기반으로 한 사운드 톤을 구현하기 위해 Suno를 중심으로 제작하였다. 그러나 Suno와 같은 글로벌 음악 생성 모델은 한국 전통 악기에 대한 학습 데이터가 매우 제한되어 있어 국립국악원에서 공개한 국악기 음원 데이터를 직접 확보하여 학습 레퍼런스로 활용하였다. 그 결과, 괴이국의 주요 장면에서는 해금의 떨림과 대금의 호흡감, 장구의 리듬감을 살린 AI 생성 국악 사운드를 구현할 수 있었다.

5. 전문가 집단 정성 평가

AI 기반 콘텐츠 제작 파이프라인으로 완성된 작품 「괴이국의 안리수」를 대상으로, 영상 제작 전문가 집단의 정성적 평가를 수행하였다. 평가는 본 작품이 AI 기반으로 제작되었음을 명시적으로 고지한 상태에서 진행되었으며, 영상의 시각적 완성도, 연출 일관성, 캐릭터-배경 일관성, 미학적 표현력, 기술적 안정성, 제작 효율성에 대해 리커트 척도(1점: 매우 낮음, 5점: 매우 높음) 및 서술형 응답을 통해 평가하였다. 평가는 총 12명의 영상 제작 전문가를 대상으로 진행되었다. <표 5>와 같이 참여자 선정 기준은 영상 제작 분야 경력 5년 이상, 단편 애니메이션 또는 영상 콘텐츠 제작 프로젝트 참여 경험 보유자로 설정하였다.

<Table 5> Expert Panel Composition

Field	Number of participants	ratio
Video Producer	1	8.3%
Graphic Designer	5	41.7%
Video Director (Video Technology)	3	25.0%
Post-production editing and colour grading specialist	3	25.0%
Total	12	100%

설문은 6개의 평가 항목을 제시했으며, 각 항목당 5개의 문항, 총 30문항으로 진행하였다. 전문가 집단 평가 항목 및 질문은 <표 6>과 같다.

<Table 6> Expert Panel Evaluation Item Composition and Questions

Evaluation Criteria	Key Evaluation Questions Examples
Visual quality	<ul style="list-style-type: none"> The color palette and lighting in the video felt natural and harmonious. The visual consistency was maintained as there was little difference in resolution or texture between frames.
Direction and Scene Composition Skills	<ul style="list-style-type: none"> The character's movement and viewpoint transitions were maintained stably. The overall direction effectively guided the audience's gaze and heightened immersion.
Character and Background Consistency	<ul style="list-style-type: none"> Even as the scene changed, the characters' appearance and proportions remained consistently stable. The character's style (artistic style, linework, color palette, etc.) was consistent with the background's aesthetic tone.
Aesthetic expressiveness	<ul style="list-style-type: none"> Traditional aesthetic elements (e.g., Korean color palette, the beauty of negative space, restrained composition) are well expressed.

	<ul style="list-style-type: none"> Symbolic elements within the scene (light, fog, background structures, etc.) enhanced the narrative's emotional tone.
Technical Stability	<ul style="list-style-type: none"> There was almost no occurrence of resolution degradation or color distortion. No afterimages or noise interfered with the viewing experience during the AI frame interpolation process.
Production Efficiency	<ul style="list-style-type: none"> I found the AI-based production method to be more efficient than traditional production. The speed of incorporating requested revisions was fast, and the quality of the incorporation was high.

〈표 7〉는 전문가 집단 정성적 평가 항목과 평균 점수, 그리고 주요 서술형 의견 내용이다. 전문가 평가단은 총 30개의 문항에서 평균 4.37점의 점수를 주었고 AI 기반 제작 파이프라인으로 완성된 단편 애니메이션 작품의 완성도와 제작 효율성을 높이 평가하였다. 특히 캐릭터와 배경의 일관성 부분에서 평균 4.26점을 기록하였는데 이는 캐릭터와 배경을 독립적으로 분리 적용한 LoRA 기법을 통해 캐릭터 일관성과 한국적 미감이 잘 유지되고 있음을 나타낸다고 할 수 있다.

〈Table 7〉 Expert Panel Evaluation Scores and Key Opinions

Evaluation Criteria	Average score (5 points out of 5)	Key Narrative Opinions
Visual quality	4.36	"The color palette and lighting beautifully captured the work's atmosphere, and the texture rendering was also very delicate, resulting in a high level of overall refinement."
Direction and Scene Composition Skills	4.37	"The scene transitions and camera angles flow seamlessly, creating a smooth rhythm and stable pacing, resulting in an exceptionally high level of overall production quality."
Character and Background Consistency	4.26	"The harmonious blend of color tones and spatial depth between the figures and background creates a unified visual composition, with the subjects standing out naturally and impressively."
Aesthetic expressiveness	4.39	"The aesthetic atmosphere and emotional delivery of the work blended harmoniously, allowing for natural immersion, and the delicate portrayal of emotional flow was particularly impressive."
Technical Stability	4.49	"The frame quality is stable and the colors are rendered consistently, giving a strong sense of technical excellence."
Production Efficiency	4.39	"The AI-based production process significantly boosts work efficiency, enabling rapid handling of repetitive editing or correction tasks, making it highly practical and valuable for real-world applications."
average	4.376	

전문가 평가는 AI 기반 제작 방식이 실험적 단계를 넘어 창작 효율성과 미학적 성취를 동시에 확보할 수 있는 실질적 제작 방식으로 자리 잡고 있음을 보여준다. 이는 본 연구의 AI 애니메이션 제작 파이프라인이 단순 기술 실험을 넘어, 향후 애니메이션 제작 방식의 구조적 변화를 제안하는 중요한 사례로 평가될 수 있음을 의미한다.

6. AI 기반 제작 파이프라인의 효율성 분석

전통적인 영상 제작 방식과 AI 기반 영상화 파이프라인을 비교함으로써, AI 기술이 '이미지 생성-영상화-편집-사운드'에 이르는 단편 애니메이션 제작의 전 과정에 미치는 효율적 영향을 정량적으로 검증하였다.

〈표8〉은 1분짜리 애니메이션제작에 소요되는 평균 제작시간, 참여인력을 기존 제작 방식과 AI 기반 제작 파이프라인을 비교하였다.

〈Table 8〉 Comparison of Efficiency Between Traditional Production Methods and AI-Based Production Pipelines

Category	Traditional animation production	AI-based production pipeline
Average production time (per minute of content)	Approximately two weeks	Approximately 5 days
Participating personnel	8 people (storyboards, character design, coloring, editing, sound, etc.)	4 people (Prompting, Visualization, Editing, Sound)

시간 효율성에 있어 AI 파이프라인은 이미지 생성, 시퀀스 변환, 영상 편집까지 대부분의 절차가 자동화되어 전통적 2주 제작 기간[30,31,32]을 5일(약 64% 단축)로 줄였다. 특히 렌더링과 합성 단계가 AI 시퀀스 생성으로 통합되며 전체 공정이 단일 워크플로우로 단축되었다.

인력 효율성 부분은 전통적으로 8명이 수행하던 제작 과정[30,31,32]을 프롬프트 엔지니어, 영상화 담당, 편집자, 사운드 디자이너 등 4인 체제로 축소하였다. 인력 절감률은 약 50%, 인적 자원 운영과 커뮤니케이션 부담이 절반 이하로 줄어들었다.

수정 용이성 부분은 AI 파이프라인에서는 장면 수정이 필요한 경우, 프롬프트 수정이나 모델 파라미터 조정으로 즉시 재생성이 가능하다. 수정 빈도가 다소 높더라도, 수정 과정이 간단하고 빠르기 때문에 실제 제작 효율성은 오히려 향상되었다고 볼 수 있다.

7. 결론

본 연구는 생성형 인공지능을 활용하여 K-전통 판타지 단편 애니메이션을 제작하는 과정 전반을 체계적으로 분석하고, 파인튜닝 모델 개발, 영상화, 전문가 평가로 이어지는 통합적 제작 파이프라인을 제시하였다. 이러한 구조는 기존의 고비용·고인력 중심 제작 방식에서 벗어나, AI 기반의 효율적 창작 환경을 실제로 구현할 수 있다는 가능성을 확인하기 위해 설계된 것으로, 본 연구의 핵심 목적인 생성형 AI 기술이 예술적·기술적 완성도를 갖춘 영상 제작을 수행할 수 있는지를 실증하는 데 있었다.

또한 생성형 AI 기반 K-판타지 애니메이션 제작에서 가장 큰 기술적 난제였던 '일관성 유지 문제'를 해결하기 위해 캐릭터용 모델과 배경용 모델을 분리하여 파인튜닝하는 방식의 새로운 제작 접근법을 제안하였다. 이는 실제 작품 「괴이국의 안리수」 제작 과정에서 효과적으로 작동하였으며 캐릭터 외형의 안정성, 배경 스타일의 지속성, 장면 간 미학적 균형을 유지하는 데 중요한 역할을 수행하였다. CCIP Distance 지표를 통한 정량 평가에서 제안 모델(Ours)은 전체 평균 0.0779를 기록하여, LoRA를 적용하지 않은 베이스라인 모델(0.1047) 대비 캐릭터 일관성이 유의미하게 향상되었음을 확인하였다.

전문가 평가 결과는 시각적 완성도(4.36), 연출 및 장면 구성력(4.37), 캐릭터·배경 일관성(4.26), 미학적 표현력(4.39), 기술적 안정성(4.49), 제작 효율성(4.39) 등 전 항목에 걸쳐 평균 4.37점(5점 만점)으로 긍정적인 평가를 받았으며, 이는 생성형 AI가 단순 참조용 도구가 아닌 실질적 창작 도구로 기능할 수 있음을 뒷받침하였다. 또한 전통적 제작 방식과의 효율성 비교에서 분당 제작 기간이 약 2주에서 5일로 약 64% 단축되었으며, 참여 인력 역시 8인에서 4인으로 약 50% 절감되어 AI 파이프라인의 실용적 가치를 정량적으로 입증하였다.

그러나 본 연구는 단편 애니메이션 제작을 중심으로 수행되었기 때문에, 장편·시리즈 제작에서 요구되는 대규모 장면 관리, 복잡한 서사 구조 유지, 캐릭터의 동세(動勢)·표정 변화 일관성, 고해상도 영상 안정성 등은 여전히 해결해야 할 과제로 남아 있다. 아울러 씬 유형에 따른 LoRA 파인튜닝 효과의 편차를 고려할 때, 향후 연구에서는 학습 데이터셋의 구성과 씬 특성 간 정합성을 보다 정밀하게 설계할 필요가 있다.

결론적으로, 본 연구는 생성형 AI가 단순 보조 도구를 넘어 창작의 중심 요소로 기능할 수 있음을 실증하였으며, AI-인간 협업 기반의 새로운 애니메이션 제작 패러

다임의 가능성을 제시하였다.

REFERENCES

- [1] Korea Creative Content Agency, "Analysis Report on Content Industry Trends for the Second Half of 2023 and the Full Year," KOCCA, 2024.
- [2] S. H. Kim and J. W. Park, "Analysis of Production Efficiency in Traditional vs. Digital Animation," Journal of Korea Multimedia Society, Vol. 21, No. 8, pp. 956-964, 2018.
- [3] Lee Young-sook, "A Study on Structural Problems and Improvement Measures in the Korean Animation Industry," Journal of Comics and Animation Studies, Vol. 56, pp. 123-145, 2019.
- [4] J. Epstein et al., "Art and the Science of Generative AI," Science, Vol. 380, No. 6650, pp. 1110-1111, 2023.
- [5] Ho, J., Jain, A., and Abbeel, P., "Denosing Diffusion Probabilistic Models," Advances in Neural Information Processing Systems (NeurIPS), 2020.
- [6] Midjourney, Midjourney Official Website, Available: <https://www.midjourney.com>
- [7] OpenAI, "DALL·E 3," OpenAI, 2023, Available: <https://openai.com/dall-e-3>
- [8] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B., "High-Resolution Image Synthesis with Latent Diffusion Models," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, Available: <https://github.com/CompVis/stable-diffusion>
- [9] Runway, Runway Gen-4, Available: <https://runwayml.com/gen4>
- [10] Higgsfield, Higgsfield AI Video Generation, Available: <https://www.higgsfield.ai>
- [11] Kuaishou, Kling AI Video Generation, Available: <https://klingai.com>
- [12] OpenAI, Sora, Available: <https://openai.com/sora>
- [13] Google, Google Veo, Available: <https://ai.google.dev/veo>
- [14] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W., "LoRA: Low-Rank Adaptation of Large Language Models," arXiv preprint, arXiv:2106.09685, 2021.
- [15] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I., "Learning Transferable Visual Models From Natural Language Supervision," arXiv preprint, arXiv:2103.00020, 2021.
- [16] Zhou, Y., et al., "Spatio-temporal Attention for Video Generation," Proceedings of the European Conference

on Computer Vision (ECCV), pp. 123-140, 2023.

[17] Vaswani, A., et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[18] Ruiz, N., Li, Y., Jampani, V., et al., "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation," *arXiv preprint, arXiv:2208.12242*, 2022.

[19] Hao, Y., Lai, Z., Song, J., Chen, Q., and He, Y., "StyleAlign: Teaching Diffusion Models with Style Consistency for Effective Image Generation," *arXiv preprint, arXiv:2310.03045*, 2023.

[20] J. Doe, A. Smith, and B. Lee, "AI Pipelines for Visual and Character Coherence in Film," *Proceedings of the ACM International Conference on Multimedia*, pp. 1-10, 2024.

[21] Stability AI, FLUX: A Fully Open, High-Performance, Large-Scale Text-to-Image Model, *Stability AI Technical Report*, 2024.

[22] Replicate, Replicate API and Model Deployment Platform, Available: <https://replicate.com>

[23] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019

[24] https://deepghs.github.io/waifuc/main/advanced_guides/ccip/index.html

[25] Carroll, Lewis, *Alice's Adventures in Wonderland*, Macmillan, 1865.

[26] Adobe, "Adobe Premiere Pro," Available: <https://www.adobe.com/products/premiere.html>

[27] Adobe, "Adobe After Effects," Available: <https://www.adobe.com/products/aftereffects.html>

[28] ElevenLabs, ElevenLabs AI Voice Platform, Available: <https://elevenlabs.io>

[29] Suno, Suno AI Music Generation, Available: <https://www.suno.ai>

[30] Pixune Studio, "How Long Does 2D Animation Take? Production Timeline," *Pixune Blog*, 2026.

[31] Animation Apprentice, "The Animated Production Pipeline - Part 2," *YouTube*, 2025.

[32] ToonsMag, "Traditional Animation vs. Digital Animation: The Ultimate Battle of Creativity," *ToonsMag*, 2024.

이영화(Younghwa Lee)

[정회원]



- 2002년 2월 : 서울과학기술대학교 매체공학과 (공학사)
- 2005년 2월 : 서울과학기술대학교 매체공학과 (공학석사)
- 2001년 8월 ~ 2021년 12월 : 현대홈쇼핑 방송그래픽팀 종합편 집 감독
- 2022년 2월 ~ 현재 : 서울과학기술대학교 산학협력단 IoT융복합연구소 책임연구원

<관심분야>

생성형 인공지능, 컴퓨터 그래픽스, AI 콘텐츠 제작

김다루(Daru Kim)

[준회원]



- 2020년 3월 ~ : 서울과학기술대학교 스마트ICT융합공학과

<관심분야>

AI미디어

권오주(O JU Kwon)

[준회원]



- 2025년 3월 ~ : 서울과학기술대학교 국방융합과학대학원 정보통신미디어공학과 석사과정
- 2025년 2월 : 동아방송예술대학교 방송기술학과 (예술학사)
- 2024년 2월 : 동아방송예술대학교 미디어창작학부 방송기술계열 (예술전문학사)

- 2023년 3월 ~ 2024년 11월 : DIMA OBVAN 중계기술팀

<관심분야>

AI 미디어 콘텐츠 제작, VFX

문 채 영(Chaeyoung Moon) [준회원]



- 2003년 2월 : 상명대학교 예술대학 만화학과(예술학사)
- 2002년 11월 ~ 2018년 9월 : 애니메이터, 일러스트레이터 (동화)
- 2018년 ~ 현재 : SF/F 소설가, 동화작가, 청소년소설가
- 2024년 9월 ~ 현재 : 서울과학기술대학교 일반대학원 문예창작과 석사과정

<관심분야>

SF/F, 동화, 청소년소설, 여성서사

최 영 희(Younghee Choi) [정회원]



- 1999년 8월 : 서울과학기술대학교 문예창작학과(문학사)
- 2002년 2월 : 경희대학교 국어국문학(문학석사)
- 2012년 8월 : 고려대학교 국어국문학(문학박사)
- 2019년 3월 ~ 2021년 2월 고려대학교 문화창의학부 초빙교수
- 2021년 3월 ~ 현재 : 서울과학기술대학교 문예창작학과 부교수

<관심분야>

생성형 AI 영화 제작, 영화 시나리오, 동아시아 영화 및 TV 드라마, 미디어와 영상문학

박 구 만(Gooman Park) [정회원]



- 1984년 2월 : 한국항공대학교 전자공학과
- 1986년 2월 : 연세대학교 전자공학과 공학석사
- 1991년 2월 : 연세대학교 전자공학과 공학박사
- 1991년 3월 ~ 1996년 9월 : 삼성전자 신호처리연구소 선임연구원
- 2016년 1월 ~ 2017년 12월 : 서울과학기술대학교 나노IT디자인융합대학원
- 1999년 8월 ~ 현재 : 서울과학기술대학교 스마트ICT융합공학과 교수
- 2006년 1월 ~ 2007년 8월 : Georgia Institute of Technology Dept.of Electrical and Computer Engineering, Visiting Scholar

<관심분야>

컴퓨터비전, 지능형실감미디어