

## 생성형 인공지능을 활용한 동영상 제작 툴의 기술적 분석과 발전 방향 연구

강 지 영\*

이화여자대학교 커뮤니케이션·미디어학부 교수

# A Study on the Technical Analysis and Development Directions of Video Production Tools Utilizing Generative AI

Jiyoung Kang\*

Professor, Division of Communication & Media, Ewha Womans University, Seoul 03760, Korea

### [요 약]

본 논문은 생성형 인공지능(AI) 기반 동영상 제작 툴의 기술적 특성과 활용 사례를 분석하여, 동영상 제작 생태계에 미치는 영향을 탐구하였다. Sora, Runway, Pictory를 중심으로 기능적 강점과 한계를 비교·분석한 결과, 이들 툴은 자동화된 영상 제작을 통해 효율성과 창의성을 높이고, 다양한 산업에서 동영상 제작의 접근성을 확대하는 것으로 나타났다. 그러나 프레임 간 일관성 부족, 높은 연산 자원 요구, 세밀한 디테일 구현의 한계와 같은 기술적 문제뿐만 아니라, 저작권 보호와 AI 생성물의 신뢰성 확보와 관련된 윤리적·법적 과제도 여전히 남아 있었다. 이를 해결하기 위해 AI 알고리즘의 고도화, 사용자 맞춤형 기능 확대, 저작권 보호 정책 개선, 생성 과정의 투명성 강화 등의 방안을 제안하였다. 본 연구는 생성형 AI 동영상 제작 툴의 가능성과 한계를 조망하며, 향후 기술 발전과 윤리적 활용을 위한 방향성을 제시하는 데 기여하고자 하였다.

### [Abstract]

This study examines the technical characteristics and use cases of generative artificial intelligence (AI)-based video production tools, exploring their impact on the video production ecosystem. The strengths and limitations of key tools such as Sora, Runway, and Pictory are analyzed and compared. The findings indicate that these tools significantly enhance efficiency and creativity through automated video generation, expanding accessibility across various industries. However, some issues, such as inconsistencies between frames, high computational resource requirements, and limitations in fine-detail rendering, persist, along with ethical and legal challenges related to copyright protection and reliability of AI-generated content. To address these challenges, this study proposes advancing AI algorithms, expanding user customization options, improving copyright protection policies, and increasing transparency in the AI generation process. By assessing both the potential and limitations of generative AI video tools, this study contributes to the discussion on technological advancements and responsible use of AI in video production.

**색인어** : 생성형 인공지능, 동영상 제작 툴, 효율성과 창의성, 저작권과 투명성, 기술적 발전

**Keyword** : Generative AI, Video Production Tools, Efficiency and Creativity, Copyright and Transparency, Technological Advancements

<http://dx.doi.org/10.9728/dcs.2025.26.3.577>



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** 15 January 2025; **Revised** 18 February 2025

**Accepted** 04 March 2025

\*Corresponding Author, Jiyoung Kang

**Tel:** +82-2-3277-2266

**E-mail:** kangjiyoung@ewha.ac.kr

## 1. 서론

### 1-1 연구 배경

2018년, 인공지능(AI)이 제작한 그림 ‘에드몽 드 벨라미(Edmond de Belamy)’가 크리스티 경매에서 약 43만 2천 달러에 낙찰된 사건은 예술계에 큰 반향을 일으켰다[1]. 이는 기존에 인간의 고유한 영역으로 간주되던 예술 창작에 인공지능이 본격적으로 진입했음을 알리는 신호탄이었다. 이후 생성형 인공지능 기술은 그림, 사진, 음악, 영상 등 예술 전반으로 그 영역을 빠르게 확장하며 기술과 예술의 융합을 가속화하고 있다.

특히, 최근 생성형 인공지능(Generative AI)의 발전이 정적인 이미지 생성에서 벗어나 동적인 영상 제작까지 가능하게 하면서 영상 제작의 패러다임을 근본적으로 변화시키고 있다. 단순한 텍스트 입력이나 간단한 스토리보드만으로 고품질의 동영상을 제작할 수 있는 Runway[2], Sora[3], Pictory[4]와 같은 생성형 인공지능 툴들은 제작 시간과 비용을 획기적으로 줄이면서도, 기술적 창작 능력이 부족한 사용자들도 창작에 참여할 수 있도록 한다는 점에서 혁신적이다. 이러한 생성형 인공지능 동영상 툴들은 현재 영화, 애니메이션, 광고 등 다양한 영상 제작 분야에 활용되고 있다.

그중 2023년에 발표된 단편 영화 'The Safe Zone'[5]은 생성형 인공지능의 혁신적 활용을 보여주는 대표적인 사례이다. 이 영화는 대본 작성, 캐릭터 디자인, 시각 효과 구현, 배경 제작에 이르는 모든 과정을 생성형 AI 기술로 처리하여, 몇 달 혹은 몇 년이 걸릴 수 있는 영화 제작 기간을 단 몇 주로 단축하였다. 이를 통해 소규모 제작팀도 고품질의 콘텐츠를 효율적으로 제작할 수 있도록 돕는 새로운 가능성을 제시했다.

또한, 영화 'Everything Everywhere All At Once'는 Runway AI와 같은 첨단 생성형 AI 툴을 활용해 시각적 경계를 확장한 것으로 주목받았다[6]. 특히 인공지능 기술을 활용하여 기존 그래픽 기술로는 많은 시간과 비용이 소요되던 시각효과를 짧은 시간 안에 효율적으로 구현할 수 있었다. 이처럼 AI는 예술적 작업 과정을 효율화하고 창의적 비전을 효과적으로 실현할 수 있는 강력한 도구로 자리 잡고 있다.

그러나 이러한 발전은 단순한 기술적 혁신을 넘어 다양한 도전 과제를 수반한다. AI가 생성한 영상의 품질과 정교함, 저작권 문제, 윤리적 논란, 그리고 인간 창작자의 역할 축소 등의 이슈가 부각되고 있다. 기존 연구들은 AI 기반 예술 창작의 윤리적 문제[7], 인간-기계 협력 모델[8],[9] 저작권 및 법적 규제[10] 등을 논의해 왔으나, 대부분 개별 사례 연구에 초점이 맞춰져 있어 생성형 AI 기술이 영상 제작 생태계 전반에 미치는 구조적 영향에 대한 체계적 분석은 부족한 상황이다.

특히, 기존 연구는 생성형 AI 동영상 제작 툴에 대한 비교 분석 및 실증 연구가 부족하여, 이 기술이 실제 제작 과정에

서 어떻게 활용되고 있는지, 이를 통해 어떠한 정책적 개선이 필요한지에 대한 논의가 제한적이었다. 따라서, 본 연구는 생성형 AI 기술이 영화 및 영상 제작 산업에 미치는 전반적 영향과 활용 가능성을 체계적으로 평가하고, 기술적·정책적 발전 방향을 모색하는 데 초점을 맞춘다.

### 1-2 연구 목적

본 연구의 목적은 생성형 인공지능 기반 동영상 제작 기술이 영화 및 영상 제작 생태계에 미치는 구조적 영향을 심층적으로 분석하고, 이러한 기술이 창작 및 제작 과정에서 제공하는 가능성과 한계를 체계적으로 평가하는 데 있다. 이를 위해 주요 생성형 AI 동영상 제작 툴(Sora, Runway, Pictory)을 중심으로 기능과 활용 사례를 비교·분석하여, 동영상 제작 툴의 기술적 특성과 실질적 효과를 명확히 규명하고자 한다.

더 나아가, 본 연구는 동영상 제작 툴의 기술적 개선 방안과 함께 저작권, 윤리적 논란, 법적 과제와 같은 주요 쟁점에 대한 해결책을 제시함으로써, 생성형 AI 기술의 지속 가능하고 책임 있는 활용 방안을 모색하고자 한다. 이를 통해 창작 생태계 전반에서 효율성과 창의성을 극대화하고, 윤리적·법적 기준을 준수하며, 기술 발전의 사회적·산업적 가치를 극대화하는 전략적 방향성을 제안하고자 한다.

궁극적으로 본 연구는 생성형 AI 기술이 동영상 제작 분야에서 창의적 혁신을 이끄는 도구로 자리 잡는 데 필요한 기술적·정책적 기반을 마련하고, 미래의 콘텐츠 제작 방식에 대한 방향성을 제시하는 데 기여하고자 한다.

## II. 생성형 인공지능 기반 동영상 제작 기술

최근 생성형 인공지능은 딥러닝(deep learning)의 비약적 발전과 함께 예술 및 콘텐츠 제작 분야에서 혁신의 중심이 되고 있다. 특히 동영상 제작에서 생성형 인공지능은 텍스트, 이미지, 또는 기타 입력 데이터를 바탕으로 짧은 클립부터 복잡한 시각 효과를 포함한 고품질 콘텐츠까지 자동으로 생성할 수 있는 기술로 발전하고 있다. 이 과정은 크게 데이터 처리, 모델 학습, 동영상 생성 단계로 나뉘며, 각 단계는 고유한 기술적 메커니즘과 도전 과제를 내포하고 있다.

### 2-1 데이터 처리: 동영상 생성을 위한 기초

생성형 인공지능 기반 동영상 제작 과정에서 데이터 처리는 핵심적인 출발점이다. 생성형 AI는 텍스트, 이미지, 오디오 등 다양한 형태의 데이터를 학습하여 동영상의 구조적 요소와 콘텐츠를 재현하거나 생성할 수 있는 능력을 가진다. 이를 위해 표 1과 같이 대규모 데이터셋의 활용과 정교한 데이터 전처리 과정이 필수적으로 수행된다.

**표 1.** 생성형 인공지능 기반 동영상 제작을 위한 데이터 처리  
**Table 1.** Data processing for generative AI-based video creation

Process	Description	Examples/Tools
Dataset Utilization	Use of large datasets to train AI on scene and action patterns.	ImageNet, Kinetics
Resolution Standardization	Normalize video resolution and frame rates to maintain consistency in input data.	Frame rate adjustments
Labeling and Noise Removal	Annotate objects/actions and remove distorted or noisy frames for accuracy improvement.	Manual/automated labeling tools
Temporal Data Enhancement	Strengthen temporal continuity through interpolation and sequence smoothing.	Interpolation algorithms

먼저, 데이터셋 활용은 AI 모델의 학습에 필수적인 기반이 된다. ImageNet과 Kinetics 같은 데이터셋은 다양한 장면과 동작을 포함하며, AI가 동영상에서 표현되는 객체와 행동 패턴을 학습할 수 있도록 돕는다[7],[8]. 예를 들어, Kinetics 데이터셋은 특정 동작(걷기, 달리기, 점프 등)에 대한 정보를 제공함으로써 AI가 인간의 움직임을 학습하고 이를 동영상 제작에 활용할 수 있게 한다. 이러한 데이터셋은 생성형 AI가 시각적 요소와 시간적 흐름을 이해하는 데 중요한 역할을 한다.

다음으로, 데이터 전처리 과정은 데이터 품질을 높이고 모델 학습의 정확성을 확보하는 데 중점을 둔다. 먼저, 모든 데이터는 해상도와 프레임 비율이 표준화되어야 한다. 이는 데이터의 일관성을 보장하고 학습 과정에서 불필요한 계산 자원을 줄이는 데 기여한다[9]. 또한, 동영상 내 각 장면과 객체에 대해 정확한 라벨링이 이루어져야 하며, 이 과정에서 노이즈가 포함된 데이터(예: 왜곡된 프레임이나 모호한 이미지)는 제거하거나 수정된다. 이러한 과정은 모델이 학습 데이터로부터 정확한 패턴을 학습할 수 있도록 도와준다.

마지막으로, 시계열 데이터의 연속성 강화 작업이 필요하다. 동영상은 시간의 흐름에 따라 연속적인 프레임으로 구성되므로, 누락된 프레임을 보완하거나 연결을 자연스럽게 만드는 작업이 중요하다. 이를 위해 보간(interpolation) 기술을 활용하여 자연스러운 동작의 흐름을 유지하며, AI가 시간적 관계를 명확히 이해할 수 있도록 한다[10]. 이러한 작업은 생성된 동영상에서 부드럽고 자연스러운 동작을 구현하는 데 필수적이다.

**2-2 모델 학습: GAN, VAE, 그리고 Transformer의 역할**

생성형 인공지능 기반 동영상 제작의 핵심은 모델 학습 단계에 있다. 이 과정에서 GAN(Generative Adversarial Networks), VAE(Variational Autoencoder), 그리고 Transformer 기반 모델과 같은 고급 딥러닝 아키텍처는 동영상 생성의 정교함과 효율성을 결정짓는다. 각 모델은 고유

한 학습 방식과 강점을 바탕으로 다양한 요구를 충족시키며, AI 기반 동영상 제작 기술의 발전을 이끌고 있다.

**1) GAN (Generative Adversarial Networks)**

GAN은 생성형 AI 기술 중 가장 널리 활용되는 모델로, 동영상 생성에서 높은 현실감과 정교함을 제공하는 핵심 기술이다. 생성자(Generator)와 판별자(Discriminator)라는 두 개의 신경망이 경쟁적으로 학습하는 구조를 갖추고 있다[11].

생성자는 랜덤 노이즈를 입력받아 점진적으로 사실적인 프레임을 생성하며, 판별자는 생성된 프레임이 실제 데이터와 얼마나 유사한지 평가한다. 두 네트워크는 상호 경쟁하면서 각각의 성능을 개선해 나가며, 이 과정에서 생성자는 점점 더 정교한 결과물을 만들어낸다. 특히, Kinetics와 같은 대규모 동작 데이터셋을 활용하면, GAN은 인간의 움직임과 자연스러운 동작 흐름을 학습할 수 있어 더욱 사실적인 영상 생성이 가능해진다[12].

이러한 특성 덕분에 GAN은 영화의 시각 효과 제작, 게임 캐릭터 애니메이션, 애니메이션에서의 자연스러운 동작 표현 등 다양한 콘텐츠 제작 분야에서 필수적인 도구가 되었다. 또한, GAN은 의료 영상 분석에서도 활용되며, 부족한 데이터 환경에서도 현실적인 의료 영상을 생성하여 연구 및 진단 보조에 기여할 수 있다. 더 나아가, GAN은 가상 현실(VR) 및 증강 현실(AR) 환경 구축에서도 중요한 역할을 하며, 가상의 공간을 보다 사실적으로 렌더링하거나 사용자의 움직임을 실시간으로 반영하는 데 활용된다. 이러한 강력한 생성 능력 덕분에 GAN은 콘텐츠 창작을 넘어 다양한 산업에서 활용되며, AI 기반 콘텐츠 제작의 핵심 기술로 자리 잡고 있다.

**2) VAE: 데이터 압축과 복원의 효율성을 극대화하는 모델**

VAE(Variational Autoencoder)는 데이터를 효율적으로 압축하고 복원하는 생성형 AI 모델로, 잠재 공간(latent space)을 활용하여 입력 데이터를 간결하고 유용한 형태로 변환한다. 이를 통해 데이터의 핵심적인 특징을 보존하면서 불필요한 정보를 제거하고, 복원 과정에서 원본에 가까운 결과를 생성할 수 있도록 한다[13].

VAE는 특히 짧은 시퀀스를 생성하거나 대규모 데이터를 효율적으로 압축하고 복원하는 데 강점을 가진다. 이 특성 덕분에 저장 공간을 절약하고 데이터 처리 속도를 높일 수 있어, 동영상 생성, 이미지 압축, 데이터 복원 등의 분야에서 활용된다. 예를 들어, 동영상 생성에서는 핵심 정보를 효과적으로 처리함으로써 복잡한 연산 과정을 간소화하고, 보다 빠른 결과물을 제공할 수 있다[13].

그러나 VAE는 GAN에 비해 세부적인 시각적 디테일을 재현하는 데 한계가 있다. 이는 VAE가 데이터의 전반적인 일관성과 자연스러움을 유지하는 데 중점을 두는 반면, GAN은 보다 사실적인 디테일을 효과적으로 구현하기 때문이다. 이러한 차이로 인해 VAE는 복잡한 시각적 요소가 필요하지 않은 간단한 구조나 짧은 시퀀스를 생성하는 데 적합하다.

동영상 생성에서 VAE는 데이터 압축과 복원을 통해 효율적인 처리를 지원하며, 특히 빠른 학습이 필요한 경우 강점을 발휘한다. 또한, 단독으로 사용되기보다 GAN이나 Transformer와 결합되어 활용되는 경우가 많으며, 이를 통해 효율성과 디테일을 모두 고려한 콘텐츠 제작이 가능해진다.

**3) Transformer 기반 모델: 동영상 생성의 문맥적 일관성을 강화하는 기술**

Transformer 기반 모델은 자연어 처리(NLP)에서 발전한 아키텍처를 동영상 생성에 적용하여, 텍스트와 시각적 출력 간의 문맥적 일관성을 강화하는 데 중점을 둔다. 이 모델은 입력된 텍스트를 분석한 후 이를 기반으로 시각적 장면을 생성하며, 프레임 간 상호작용을 예측하여 자연스럽게 연속적인 콘텐츠를 제작하는 데 뛰어난 성능을 가지고 있다[13],[14].

Transformer 기반 모델의 핵심은 텍스트를 통해 사용자의 의도를 정확히 파악하고 이를 구체적인 시각적 결과로 변환하는 데 있다. 예를 들어, VideoGPT는 입력된 텍스트를 분석하여 장면의 스타일, 색감, 동작 등을 조화롭게 결합하고, 이를 바탕으로 문맥적 흐름을 유지하는 시각적 콘텐츠를 생성한다. 이러한 접근 방식은 동영상 생성에서 창의적이고 직관적인 결과물을 제공하며, 사용자로 하여금 더욱 풍부한 표현력을 발휘할 수 있도록 돕는다[13].

특히, Transformer 기반 모델은 텍스트-비디오 생성 작업에서 강력한 성능을 발휘한다. 사용자가 간단한 텍스트 명령을 입력하면, 이를 기반으로 장면 전환, 움직임, 조명 효과 등 세부적인 시각적 요소를 자동으로 생성할 수 있다. 이로 인해 VideoGPT와 같은 모델은 영화 제작, 광고 캠페인, 교육 콘텐츠 생성 등 다양한 응용 분야에서 폭넓게 활용되고 있으며, 동영상 생성 기술의 혁신적인 방향성을 제시하고 있다[14].

**2-3 동영상 생성 단계**

동영상 생성 과정은 텍스트, 이미지, 또는 기타 데이터 형태를 입력받아 이를 시각적 콘텐츠로 변환하는 여러 단계의 복합적인 과정을 포함한다. 이 과정은 아래 표 2와 같이 텍스

**표 2.** 동영상 생성 단계의 핵심 기술

**Table 2.** Key technologies in video generation stages

Stage	Technology/Algorithm	Function and Application
Text-to-Video Generation	NLP, Sora, Runway	Converts text input into visual scenes; used in advertising and film production.
Frame Prediction	ConvLSTM, RNN	Sequentially generates subsequent frames from an initial state; ensures smooth and natural movements.
Style Transfer and Detail Enhancement	StyleGAN, Super-Resolution	Applies visual styles and converts low-resolution content to high resolution; improves content quality and detail.

트 기반 생성, 프레임 간 예측, 그리고 스타일 전이 및 디테일 보강의 세 가지 핵심 요소로 구성되며, 최신 생성형 AI 기술의 활용을 통해 고품질의 동영상을 제작할 수 있도록 지원한다. 이러한 단계들은 각각의 기술적 강점을 결합하여, 사용자가 직관적이고 효율적으로 동영상을 생성할 수 있는 기반을 제공한다.

**1) 텍스트 기반 생성**

텍스트 기반 생성은 사용자가 입력한 텍스트를 자연어 처리(NLP) 기술로 분석하여 이를 시각적 장면으로 변환하는 과정을 의미한다. 이 과정은 간단한 텍스트 명령만으로 장면의 스타일, 색감, 조명, 움직임 등을 자동으로 구현할 수 있어, 동영상 제작의 접근성과 효율성을 크게 향상시키는 데 기여한다.

최근 연구에서는 ChatGPT와 같은 언어 모델을 활용하여 텍스트 기반 비디오 콘텐츠 제작이 이루어지고 있으며, Pika Lab AI와 같은 툴을 통해 텍스트 데이터를 시각적 요소로 변환하는 작업이 활발히 진행되고 있다. 이러한 접근법은 비전문가도 손쉽게 고품질 콘텐츠를 제작할 수 있도록 지원하며, 영화 및 광고와 같은 다양한 분야에서 창의적 가능성을 확장하고 있다[15].

특히, Sora와 Runway는 텍스트 기반 동영상 제작의 대표적인 툴로, 광고 및 영화 제작 과정에서 혁신적인 사례를 보여주고 있다. 예를 들어, 간단한 스토리라인이나 특정 시각적 테마를 텍스트로 입력하면, 이를 바탕으로, 자동으로 장면을 구성하고, 디테일한 시각적 요소를 생성하여 창작자의 비전을 실현하는 데 도움을 준다. 이러한 기술은 동영상 제작의 기존 한계를 극복하며, 더 빠르고 효율적인 콘텐츠 제작 환경을 제공하고 있다[15].

**2) 프레임 간 예측**

프레임 간 예측은 초기 상태에서 시작하여 이후 프레임을 순차적으로 생성함으로써 자연스럽게 부드러운 움직임을 구현하는 기술이다. 이 과정은 시계열 데이터를 학습하여 프레임 간의 연속성을 유지하는 데 중점을 두며, ConvLSTM (Convolutional Long Short-Term Memory)과 RNN (Recurrent Neural Network) 같은 알고리즘이 주로 활용된다. 이러한 알고리즘은 과거 프레임 정보를 바탕으로 다음 프레임을 예측하여, 시각적 일관성을 유지하면서 동적인 장면을 생성할 수 있도록 지원한다.

또한, 딥러닝 기반의 비디오 프레임 보간 기술은 연속된 두 프레임 사이의 중간 프레임을 생성하여 영상의 프레임 속도를 증가시키고, 부드러운 재생 품질을 제공한다. 이 기술은 특히 텔레비전, 셋톱박스, 스트리밍 플랫폼 등에서 영상 재생 품질을 개선하는 데 폭넓게 활용되고 있다. 예를 들어, 고속 움직임이 포함된 스포츠 경기나 액션 영화에서 이 기술은 시청자에게 더 자연스럽게 몰입감 있는 시청 경험을 제공할 수 있다[16]. 프레임 간 예측 기술은 동영상 생성뿐만 아니라, 저속 촬영 영상의 보강, 해상도 업스케일링, 그리고 실시간 스트

리밍 콘텐츠의 최적화 등 다양한 응용 분야에서 중요한 역할을 하고 있다.

### 3) 스타일 전이 및 디테일 보강

스타일 전이 및 디테일 보강은 생성된 동영상의 시각적 품질을 향상시키고 창의적 표현력을 강화하기 위해 중요한 역할을 하는 기술이다. StyleGAN과 같은 모델은 특정 예술적 스타일을 영상에 적용하여 독창적인 시각적 콘텐츠를 제작할 수 있다. 예를 들어, 기존의 실사 영상을 애니메이션 스타일로 변환하거나, 특정 유명 예술가의 작품 스타일을 동영상에 적용하는 방식으로 활용된다.

또한, 초해상화(Super-Resolution) 기술은 저해상도 영상을 고해상도로 변환하여 세부적인 디테일을 강화한다. 이를 통해 화면의 선명도를 높이고, 미세한 시각적 요소를 더욱 정교하게 표현할 수 있다. 이러한 기술은 특히 360도 VR 콘텐츠 제작에서 활용되며, 몰입감 높은 고품질 영상을 구현하는데 필수적인 역할을 한다[17].

스타일 전이 및 디테일 보강 기술은 영화, 게임, 광고 콘텐츠 제작 등 다양한 분야에서 널리 사용되고 있으며, 예술적 창의성과 기술적 정교함을 결합한 영상 제작이 가능하도록 지원한다. 이로 인해 시청자에게 더 강렬한 시각적 경험을 제공하며, 디지털 콘텐츠의 가치를 한층 높이는 데 기여하고 있다.

## III. 생성형 인공지능 동영상 제작 툴

생성형 인공지능 기술은 동영상 제작의 효율성을 극대화하고 창의적 표현의 가능성을 확장시키며, 다양한 응용 분야에서 점점 더 중요한 역할을 하고 있다. 이 연구에서는 Sora, Runway, Pictory와 같은 주요 생성형 AI 동영상 제작 툴들을 분석하고자 한다.

### 3-1 Sora

Sora는 OpenAI에서 개발한 텍스트 기반 동영상 생성 모델로, 사용자가 입력한 텍스트를 분석하여 고품질의 동영상을 자동으로 생성한다. OpenAI의 자연어 처리 기술(GPT-4)과 이미지 생성 기술(DALL·E)의 결합을 통해 텍스트 프롬프트에서 사실적이고 연속적인 영상을 생성하는 데 강점을 가진다. 특히, Transformer 기반 아키텍처를 활용하여 텍스트의 의미를 해석하고, 이를 시각적 요소로 변환하는 과정을 수행한다[18].

Sora의 동영상 제작 과정은 다음과 같은 주요 단계를 포함한다. 그림 1과 같이 사용자가 텍스트 입력을 하면 자연어 처리(NLP) 기술을 통해 텍스트의 맥락을 분석하고, 이를 바탕으로 장면 스타일, 색감, 조명, 움직임 등을 자동으로 설정하는 방식으로 진행된다. 이어서 ConvLSTM 및 시계열 데이터 학습 알고리즘을 적용해 프레임 간 연속성을 보장하며, 마지막

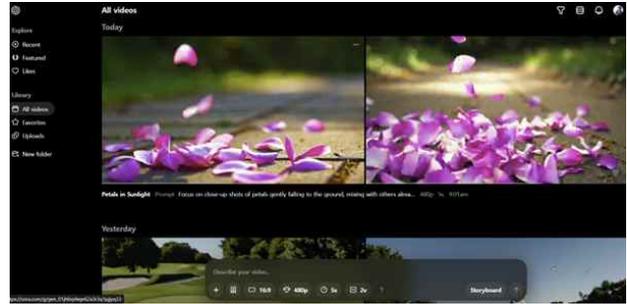


그림 1. 소라의 인터페이스  
Fig. 1. Sora's interface

으로 초해상화(Super-Resolution) 기술을 통해 디테일을 보강하여 최대 60초 길이의 고품질 동영상을 완성한다. 이러한 기술적 접근 방식은 광고, 소셜 미디어, 교육 콘텐츠 등에서 텍스트 기반 효율적인 동영상 제작을 가능하게 한다[19].

그러나 Sora는 제작 가능한 동영상 길이가 60초로 제한되며, 복잡한 동작이나 다중 객체 간 상호작용을 처리하는 데 어려움이 있다. 또한, 고해상도 콘텐츠 생성 시 높은 연산 자원이 요구되므로 하드웨어 성능이 결과물 품질에 큰 영향을 미친다.

### 3-2 Runway

Runway는 AI 기반 동영상 생성 및 편집 플랫폼으로, 텍스트 및 기존 영상 데이터를 활용하여 다양한 기능을 제공한다. 특히, 텍스트-비디오 생성(Text-to-Video Generation) 및 비디오-비디오 변환(Video-to-Video Transformation) 기능을 통해 영화, 광고, 소셜 미디어 콘텐츠 제작에서 널리 활용되고 있다. 최신 모델인 Gen-2는 텍스트 입력만으로 영상 제작이 가능하며, Gen-3는 기존 영상의 스타일 변환과 시각 효과 추가를 더욱 강화하였다[20].

Runway는 그림 2와 같이 사용자가 텍스트 설명 또는 기존 동영상을 입력하면 AI 모델이 이를 분석하여 장면 구성 및 스타일링을 자동 설정하는 방식으로 동영상을 생성한다. 이후, 프레임 간 자연스러운 연결성을 유지하며 Super-Resolution 기술을 적용해 고해상도로 최적화된 결과물을 출력한다. 이러한 기능 덕분에 Runway는 광고 및 영화 제작 현장에서 활용되고 있으며, 볼보(Volvo)와 같은 기업이 이를 활용하여 단기간 내에 고품질의 광고 영상을 제작한 사례가 있다[21].

그러나 Runway의 고급 기능을 활용하려면 높은 성능의 컴퓨팅 자원이 필요하며, 사용자의 하드웨어 성능에 따라 제작 효율성과 결과물 품질이 좌우될 수 있다. 또한, 다양한 기능이 통합된 인터페이스로 인해 초보자가 익숙해지는 데 시간이 필요하고, 무료 플랜에는 기능 제한이 있어 비용 부담이 발생할 수 있다.

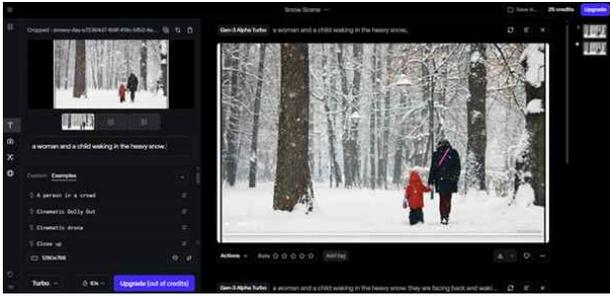


그림 2. 런웨이의 인터페이스  
Fig. 2. Runway's interface

그러나 Runway의 고급 기능을 활용하기 위해 고성능 컴퓨팅을 필요로 하며, 사용자의 하드웨어 성능에 따라 제작 효율성과 결과물의 품질이 달라질 수 있다. 또한, 다양한 기능이 통합된 인터페이스로 인해 초보자가 도구를 익히는 데 시간이 필요하며, 무료 플랜의 기능 제약과 유료 구독 모델의 비용 부담 역시 소규모 제작자들에게는 장벽이 될 수 있다. 그럼에도 불구하고, Runway는 텍스트-비디오 및 비디오-비디오 변환 기술에서 독보적인 성능을 발휘하며, 광고, 영화, 소셜 미디어 콘텐츠 제작 등에서 중요한 도구로 자리 잡고 있다.

### 3-3 Pictory

Pictory는 인공지능(AI)을 활용하여 텍스트 기반 콘텐츠를 동영상으로 변환하는 클라우드 기반 플랫폼이다. 이 도구는 블로그 게시물, 기사, 스크립트 등 텍스트 콘텐츠를 입력받아 자동으로 시각적 요소, 배경 음악, 자막 등을 포함한 동영상을 생성한다. 특히, 텍스트를 동영상으로 변환하는 과정을 자동화하여 사용자가 비디오 편집 기술 없이도 손쉽게 콘텐츠를 제작할 수 있도록 설계되었다[22].

Pictory는 텍스트-비디오 변환 기능을 중심으로 동영상 제작의 효율성을 극대화하고 있다. 그림 3과 같이 사용자가 텍스트를 입력하면, Pictory는 텍스트의 내용을 자연어 처리(NLP) 기술로 분석하여 적합한 이미지를 자동으로 선택하고, 이를 장면화하여 동영상을 생성한다. 생성된 동영상에는 자동 자막 추가 기능이 포함되어 있어 시청자의 접근성과 참여도를 높이며, 다양한 브랜드 맞춤형 템플릿과 스타일을 제공하고 있다.

Pictory의 제작 과정은 크게 텍스트 입력 및 분석, 장면 구성 및 이미지 선택, 배경 음악 및 자막 추가, 템플릿 및 스타일 적용, 그리고 최종 수정 및 출력의 단계로 이루어진다. 텍스트 분석 단계에서 키워드와 주요 주제를 추출한 후, Pictory는 자동으로 시각적 콘텐츠를 구성하고 적절한 이미지를 배치한다. 이후, 텍스트를 기반으로 자막을 생성하고 배경 음악을 추가하여 콘텐츠의 몰입감을 강화한다. 마지막으로, 사용자는 다양한 템플릿과 디자인 스타일을 활용해 세부적인 수정 작업을 거쳐 최종 결과물을 완성할 수 있다. 이러한 자동화 과정은 콘텐츠 제작 시간을 획기적으로 단축하며, 비전문가도 고품질

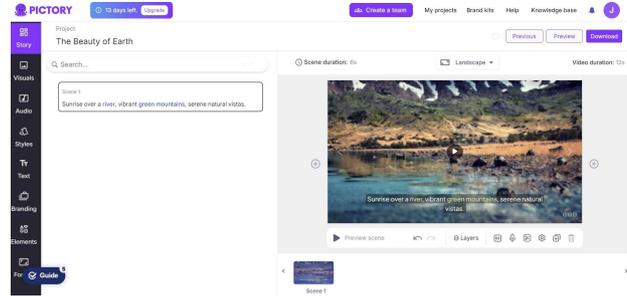


그림 3. 픽토리의 인터페이스  
Fig. 3. Pictory's interface

콘텐츠를 제작할 수 있도록 지원한다[23].

한 출판사는 Pictory를 통해 책의 내용을 요약한 비디오를 제작하여 독자들의 관심을 높였으며, 이는 도서 판매 증가로 이어졌다. 또한, 한 팟캐스트 제작자는 Pictory를 활용해 에피소드의 하이라이트 영상을 제작하여 청취자 수를 확대하였다[24]. 이러한 사례들은 Pictory의 자동화된 비디오 제작 기능이 콘텐츠 제작 시간을 단축시키고, 비전문가도 고품질의 콘텐츠를 제작할 수 있도록 지원함을 보여주었다.

그러나 Pictory는 고급 비디오 편집 소프트웨어에 비해 세부적인 편집 기능이 제한적이며, 생성된 AI 음성이 자연스럽지 않다는 평가가 있다. 그럼에도 불구하고, Pictory는 텍스트 입력만으로 자동 자막 생성, 요약, 시각적 장면 분할을 지원하여 콘텐츠 제작 시간을 크게 단축시키며, 마케팅, 교육, 소셜 미디어 콘텐츠 제작 등 다양한 분야에서 유용한 활용이 가능한 도구로 평가받고 있다.

## IV. 기술적 작동 원리 및 특징

생성형 인공지능 동영상 제작 툴들은 기술적 작동 원리에 따라 크게 텍스트 기반 동영상 제작과 스토리보드 기반 동영상 제작으로 구분된다. 이 두 방식은 콘텐츠 제작 과정에서 각각의 강점과 한계를 가지며, 다양한 응용 분야에서 활용된다. 아래에서는 앞서 소개한 3가지 툴(Sora, Runway, Pictory)의 기술적 작동 원리와 특징을 중심으로 두 방식을 비교 분석한다.

### 4-1 텍스트 기반 동영상 제작

텍스트 기반 동영상 제작은 사용자가 입력한 텍스트를 기반으로 동영상의 스타일, 움직임, 색감 등을 자동으로 설정하고, 이를 시각적 콘텐츠로 변환하는 과정으로 이루어진다. 이 방식은 자연어 처리(NLP) 기술과 컴퓨터 비전 알고리즘의 결합을 통해 동영상 제작의 효율성과 접근성을 크게 향상시키며, 비전문가도 손쉽게 고품질 동영상을 제작할 수 있도록 지원한다.

Sora, Pictory는 모두 텍스트 기반 동영상 제작 툴로, 각

도구는 고유한 기능과 프로세스를 통해 동영상 제작을 지원한다. Sora는 OpenAI의 Transformer 기반 모델을 활용해 광고와 소셜 미디어 콘텐츠 제작에서 뛰어난 성능을 발휘하며, Pictory는 블로그와 기사 등 텍스트 콘텐츠를 동영상으로 변환하는 데 최적화된 도구다.

텍스트 기반 동영상 제작은 텍스트 입력에서 최종 출력까지 아래 표 3과 같이 네 단계로 이루어진다. 먼저, 사용자가 입력한 텍스트를 NLP 기술로 분석하여 주요 키워드와 맥락을 파악하고, 이를 기반으로 장면 스타일과 구성을 자동으로 설정한다. 이후 ConvLSTM 알고리즘을 활용해 프레임 간 연속성을 유지하며 동영상을 생성한 뒤, StyleGAN과 초해상화(Super-Resolution) 기술로 디테일과 해상도를 보강해 최종 고품질 동영상을 출력한다.

표 3. 텍스트 기반 동영상 제작 프로세스

Table 3. Text-based video production process

Step	Description
1. Text Input and Analysis	Users input text, which is analyzed using NLP technology to interpret the context and meaning. Key themes and keywords determine scene composition and style.
2. Scene Creation and Composition	Based on the analyzed text, the style, color, lighting, and motion for each scene are configured. This process converts text into visual data, forming a visual storyline.
3. Frame Generation and Continuity	Using algorithms such as ConvLSTM, initial scenes are connected seamlessly to ensure continuity between frames, creating a cohesive video.
4. Detail Enhancement and Output	Technologies like StyleGAN and Super-Resolution are applied to enhance details and resolution, producing the final high-quality video.

위와 같이 텍스트 기반 동영상 제작은 텍스트 입력만으로 제작 과정을 자동화하여 시간과 비용을 절감할 수 있다는 점에서 큰 강점을 지닌다. 그러나 감정 표현, 복잡한 스토리라인, 상호작용 요소 구현이 필요한 영상 콘텐츠에서는 여전히 한계가 존재한다.

#### 4-2 스토리보드 기반 동영상 제작

스토리보드 기반 동영상 제작은 사용자가 사전에 정의한 스토리보드나 시각적 요소(이미지, 비디오 클립 등)를 입력하여 이를 바탕으로 동영상을 생성하거나 편집하는 방식이다. 이 접근법은 사용자가 각 장면의 구성과 세부 사항을 직접 지정할 수 있다는 점에서 창의적이고 고도의 커스터마이징이 가능하다. 특히, 영화, 광고, 그리고 복잡한 스토리라인이 요구되는 콘텐츠 제작에서 활용도가 높다.

Runway와 같은 도구는 스토리보드 기반 동영상 제작의 대표적인 사례이다. 사용자는 텍스트 명령뿐만 아니라, 스토리보드 형식의 데이터를 사용해 원하는 시각적 결과를 세밀

하게 설계할 수 있다. 예를 들어, 특정 장면의 색감 조정, 배경 교체, 객체 강조 등을 쉽게 구현할 수 있어, 영화 제작이나 고급 마케팅 콘텐츠에서 유용하다[22].

스토리보드 기반 동영상 제작은 사용자가 사전에 정의한 시각적 시나리오와 장면 구성을 입력하면, 이를 기반으로 AI가 동영상을 생성하거나 편집하는 방식으로 이루어진다. 이 프로세스는 사용자가 제공한 스토리보드 데이터를 바탕으로 각 장면의 스타일, 움직임, 색감, 조명 등 세부적인 요소를 설계하며, 이를 통해 고도의 창의성과 맞춤형 제작이 가능하다. 특히 영화, 광고, 고급 마케팅 자료와 같은 분야에서는 복잡한 스토리라인과 정교한 시각적 표현이 필수적이며, 스토리보드 기반 제작 방식은 이러한 요구를 효과적으로 충족할 수 있는 강력한 도구라고 할 수 있다.

아래 표 4는 스토리보드 기반 동영상 제작 프로세스를 체계적으로 정리하여 각 단계의 주요 작업과 역할을 설명한다. 스토리보드 기반 동영상 제작은 스토리보드 입력부터 최종 동영상 출력까지의 과정을 포함하며, 텍스트 설명, 이미지 입력, 장면 구성, 스타일 설정, 세부 조정 등 단계별 세부 작업이 이루어진다. 특히, AI를 활용한 장면 분석, 시각적 콘텐츠 생성, 고해상도 출력과 같은 기술적 요소와 사용자 참여가 필요한 작업 단계를 강조하여, 효율적이고 맞춤형 콘텐츠 제작을 가능하게 한다.

표 4. 스토리보드 기반 동영상 제작 프로세스

Table 4. Storyboard-based video production process

Step	Description
1. Storyboard Input and Setup	Users input a storyboard with text descriptions and images, defining the timeline and key elements (e.g., objects, backgrounds, styles). AI processes this data to start the video creation process.
2. Scene Analysis and Style Setup	AI analyzes the storyboard and sets the style, color palette, lighting, and motion for each scene. Custom visual effects, such as animations or realistic rendering, can be applied based on user preferences.
3. Visual Content Generation	AI generates individual scenes and combines them into a coherent storyline, ensuring frame continuity with algorithms like ConvLSTM, and produces high-resolution visuals.
4. Fine-Tuning and Enhancement	Users review the generated content and make detailed adjustments, such as color corrections, object highlighting, or adding text overlays. Advanced editing tools allow further customization.
5. Final Video Output	The finalized video is exported in high resolution and can be saved in various file formats for different use cases.

V. 생성형 인공지능을 활용한 동영상 결과물의 비교

5-1 동영상 결과물의 정성적 비교

주요 생성형 인공지능 동영상 제작 툴의 성능을 평가하기 위해 동일한 입력 문장을 바탕으로 생성된 동영상의 결과물을 비교하여 정성적, 정량적 평가를 수행하였다. 입력문장은 "A futuristic city at night with flying cars and neon lights."로 정하여 아래 표 5와 같이 Sora, Runway, Pictory의 동영상을 제작하였다. 결과물의 비교 분석을 위해 시각적 품질, 장면 구성 정확도, 부드러움, 색상 및 스타일이라는 네 가지 평가 기준을 적용하였으며, 표 6에 각 결과물의 특징을 정리하였다.

표 5. 생성형 인공지능을 활용한 동영상 결과물  
Table 5. Video outputs using generative AI

Tool	Video
Sora	
Runway	
Pictory	

Sora의 결과물은 고해상도와 포토리얼리스틱한 세부 묘사를 제공하며, "futuristic city," "night," "flying cars," "neon lights" 등의 요소를 정교하게 반영하여 장면 구성의 정확성과 자연스러운 배치를 실현했다. 또한, SF적 분위기를 강조하는 네온 조명과 사실적인 광원 효과가 두드러져, 현실감 있는 비주얼을 구현했다.

Runway는 스타일화된 그래픽과 예술적 필터를 적용하여 창의적인 방식으로 입력 문장을 해석하고 구현했으나, 장면

표 6. 생성형 인공지능을 활용한 동영상 결과물의 질적 비교  
Table 6. Qualitative compison of video outputs using generative AI

Tool	Visual Quality	Scene Composition Accuracy	Smoothness	Color & Style
Sora	High-resolution, photorealistic details	High level of scene arrangement and composition	Smooth motion and consistent frame transitions	Sci-fi ambiance with neon lighting and realistic illumination effects
Runway	Stylized visuals, artistic filters applied	Creative but slightly less structural accuracy	Some scene discontinuities observed	Graphic novel style with strong color contrasts
Pictory	Animation-style, simple effects	Mostly static scene compositions	Limited motion and lower continuity	Bright and clear color palette, simplistic design

구조의 정확성이 상대적으로 낮고 일부 개체 배치가 부자연스러운 경우가 관찰되었다. 모션 측면에서는 일부 불연속성이 발생했지만, 강한 색상 대비와 그래픽 노블 스타일의 시각적 효과를 통해 독창적인 분위기를 강조했다.

Pictory는 애니메이션 스타일의 단순한 효과를 중심으로 동영상을 생성하며, 입력된 요소를 간략하게 표현했다. 정적인 장면 구성과 제한적인 동적 요소 활용이 특징이며, 밝고 명확한 색상을 사용해 직관적인 시각적 효과를 제공했다.

이러한 비교를 통해, 생성형 AI 동영상 제작 도구는 각기 다른 장점과 한계를 가지며, 활용 목적에 따라 적절한 선택이 필요함을 확인할 수 있었다. Sora는 고품질의 사실적인 영상 제작에, Runway는 예술적이고 창의적인 영상 제작에, Pictory는 단순한 애니메이션 및 교육 콘텐츠 제작에 적합한 도구인 것을 알 수 있었다.

5-2 동영상 결과물의 정량적 비교

생성형 인공지능 동영상 제작 툴의 성능을 보다 객관적으로 평가하기 위해 정량적 비교 분석을 수행하였다. 평가 지표로는 프레임 해상도, 프레임 속도, 개체 인식률, 장면 전환 일관성, 색상 정확도 및 스타일 일관성을 선정하여 신뢰성이 검증된 도구와 알고리즘을 활용하여 측정하였다.

1) 프레임 해상도(Frame Resolution)

동영상 제작 도구의 해상도는 콘텐츠의 시각적 품질을 결정하는 중요한 요소이다. 생성형 인공지능 동영상 제작 도구는 사전에 설정된 최대 해상도가 존재하며, 사용자가 이를 직접 변경할 수 없는 경우가 많다. 이에 따라, 본 연구에서는 각 도구의 공식적으로 지원하는 최대 출력 해상도를 기준으로 분석하였다.

표 7. 프레임 해상도 비교

Table 7. Comparision of frame resolution

AI Video Generation Tool	Maximum Resolution
Sora	1920×1080 or 1080×1920
Runway	1920×1080
Pictory	1280×720

분석 결과(표 7), Sora는 1080p 해상도를 공식적으로 지원하며, 이는 고해상도 콘텐츠 제작에 적합함을 시사한다. 반면, Pictory는 최대 720p의 해상도를 제공하여 웹이나 단순한 애니메이션 스타일의 영상 제작에 적합한 것을 알 수 있었다.

2) 프레임 속도 (Frames Per Second, FPS)

프레임 속도(FPS)는 영상의 부드러움과 자연스러운 움직임을 평가하는 핵심 지표이다. 본 연구에서는 ffmpeg 라이브러리[25]를 활용하여 각 동영상의 평균 FPS를 측정하였다.

표 8. 프레임 속도 비교

Table 8. Comparision of frames per second

AI Video Generation Tool	Average FPS
Sora	30~60 FPS
Runway	24 FPS
Pictory	15 FPS

분석 결과(표 8), Sora는 30~60 FPS를 유지하며 자연스러운 움직임을 제공하는 반면, Pictory는 15 FPS로 상대적으로 낮은 프레임 속도를 제공하였다. 이는 Pictory의 영상이 정적인 프레젠테이션이나 간단한 애니메이션 제작에 보다 적합함을 시사한다.

3) 개체 인식률 (Object Recognition Accuracy)

입력 문장에서 명시된 주요 개체("futuristic city," "flying cars," "neon lights")가 생성된 영상 내에서 얼마나 정확하게 구현되었는지를 평가하기 위해 Google Vision AI[26] 및 YOLOv5 모델[27]을 활용하여 개체 인식률을 분석하였다.

표 9. 개체 인식률 비교

Table 9. Comparision of object recognition accuracy

AI Video Generation Tool	Object Recognition Accuracy (%)
Sora	92.5%
Runway	78.3%
Pictory	64.7%

분석 결과(표 9), Sora는 개체 인식률이 가장 높아 입력된 문장을 보다 정밀하게 반영하는 것으로 나타났다. 반면, Runway는 개체 배치의 정확성이 다소 낮았으며, Pictory는 단순화된 개체 표현으로 인해 상대적으로 낮은 인식률을 기록하였다.

4) 장면 전환 일관성 (Scene Transition Consistency)

장면 전환이 얼마나 부드럽고 일관되게 이루어지는지를 분석하기 위해 Optical Flow 분석 방법[28]을 적용하였다. Optical Flow는 연속된 프레임 간의 픽셀 이동을 추적하여 영상 내 모션 흐름의 자연스러움을 평가하는 기법이다.

표 10. 장면 전환 일관성 비교

Table 10. Comparision of scene transition consistency

AI Video Generation Tool	Scene Transition Consistency (%)
Sora	95.4%
Runway	81.2%
Pictory	67.8%

위 표 10와 같이 Sora는 Optical Flow 분석 결과 가장 높은 장면 전환 일관성을 유지하며, 자연스러운 모션을 제공하는 것으로 나타났다. Runway는 비교적 일관된 장면 전환을 유지했으나, 일부 필터 효과로 인해 모션의 연속성이 저하되었다. Pictory는 장면 간 급격한 변화가 발생하여 전환이 부드럽지 못한 경향을 보였다.

5) 색상 정확도 및 스타일 일관성(Color Accuracy & Style Consistency)

색상의 정확성과 스타일 유지 능력을 평가하기 위해 Delta E 색상 차이 계산 방법[29]을 적용하였다. Delta E 값이 낮을수록 색상의 정확도가 높고 스타일이 일관되게 유지됨을 의미한다.

표 11. 색상 정확도 및 스타일 일관성 비교

Table 11. Comparision of color accuracy & style consistency

AI Video Generation Tool	Color Accuracy (%)
Sora	94.3%
Runway	85.6%
Pictory	72.1%

Sora는 색상 정확도가 가장 높아 입력된 설정을 충실히 반영하는 경향을 보였다. 반면, Runway는 예술적 필터 적용으로 인해 색상의 변형이 관찰되었으며, Pictory는 애니메이션 스타일을 반영하여 단순한 색상 구성을 유지하는 특성을 보였다.

종합하면 Sora는 해상도, 프레임 속도, 개체 인식률, 장면 전환 일관성, 색상 정확도에서 가장 우수한 성능을 보였으며, 사실적인 비주얼을 요구하는 콘텐츠 제작에 적합함을 확인하였다. 반면, Runway는 창의적인 스타일 표현과 예술적 필터를 제공하지만, 장면 전환과 개체 정확성에서 상대적으로 낮은 평가를 받았다. 따라서, 예술적 감각이 강조된 콘텐츠 제작에 적합할 것으로 예상된다.

Pictory는 해상도, 개체 인식률, 장면 전환 일관성에서 낮은 평가를 받았지만, 단순한 애니메이션 스타일의 콘텐츠 제

작에는 적합한 성능을 보였다. 따라서, 교육용 자료, 프레젠테이션 영상 등에서 활용 가치가 높을 것으로 판단된다.

5-3 생성형 AI 동영상 제작 도구 비교

동영상 결과물의 비교를 기반으로 하여 각 도구의 기능과 성능을 비교한 결과는 다음 표 12와 같다.

표 12. 생성형 AI 동영상 제작 도구 비교  
Table 12. Comparison of generative AI video creation tools

Feature/Model	Sora	Runway	Pictory
Main Features	Text-to-video generation	Text-to-video & video-to-video transformation	Text-based automated video creation
Technology Base	Transformer, ConvLSTM, Super-Resolution	Gen-2, Gen-3, Super-Resolution	NLP-based image selection & subtitle generation
Video Quality	High-resolution, photorealistic	Stylized graphics, creative filters	Animation-style, intuitive composition
Frame Consistency	High	Medium	Low
Complex Scene Processing	Limited	Relatively superior	Suitable for simple scene composition
User-Friendliness	Usable by non-experts	Requires interface learning	Intuitive usage
Hardware Requirements	High	Medium to high	Low
Use Cases	Advertising, film, social media	Advertising, film, artistic projects	Education, marketing, content summarization

본 연구를 통해 생성형 AI 동영상 제작 도구는 특정 제작 목적과 요구 사항에 따라 선택이 달라질 수 있음을 확인하였다. Sora는 사실적인 영상 제작에, Runway는 창의적이고 예술적인 스타일 구현에, Pictory는 교육 및 간단한 마케팅 콘텐츠 제작에 적합한 것을 알 수 있었다.

VI. 동영상 제작 툴의 발전 가능성과 보완점

5-1 기술적 개선 방안

생성형 인공지능 기반 동영상 제작 툴이 더 널리 활용되고 발전하기 위해서는 현재의 기술적 한계를 넘어 사용자 경험을 혁신적으로 향상시키는 노력이 필요하다. 특히, 동영상 제작 과정에서 사용자가 더욱 효율적이고 직관적으로 작업할

수 있도록 기술적 도구와 인터페이스를 고도화하는 것이 필수적이다. 다음에서는 이를 실현하기 위한 구체적인 기술적 개선 방안을 제시한다.

1) 고도화된 사용자 인터페이스

현재 동영상 제작 툴은 직관적인 인터페이스를 제공하고 있지만, 고급 기능을 활용하는 과정에서 사용자가 여전히 복잡함과 어려움을 느끼고 있다. 초보자와 전문가 모두를 만족시키는 인터페이스 설계는 여전히 중요한 과제로 남아 있으며, 이를 해결하기 위해 AI 기반의 가이드 시스템과 실시간 피드백 기능을 통합할 필요가 있다.

예를 들어, AI가 작업 중 적절한 권장 사항을 제안하거나, 오류를 자동으로 감지하고 수정하는 기능은 사용자가 보다 효율적으로 작업을 완료할 수 있도록 도울 수 있을 것이다. 또한, 다중적 사용자 경험 설계를 도입하여 초보자는 직관적인 워크플로우를, 전문가에게는 세부적인 제어 기능을 제공하는 맞춤형 UI를 구축하는 것이 필요하다.

연구에 따르면, 생성형 AI와 사용자의 실시간 상호작용은 작업 시간을 단축시키고, 결과물의 품질을 향상시키는 데 긍정적인 영향을 미치는 것으로 나타났다[30]. 따라서, 사용자 인터페이스의 개선은 단순한 편리성 향상을 넘어 동영상 제작 툴의 활용도를 확대하고, 보다 다양한 사용자층을 포용할 수 있는 중요한 전략적 접근이 될 것이다.

2) 품질 향상을 위한 알고리즘 개선

현재 생성형 AI 기반 동영상 제작 툴은 프레임 간 일관성 부족과 세부 디테일 구현의 한계로 인해 고품질 콘텐츠 제작에 제약을 보이고 있다. 특히 복잡한 움직임이나 정교한 시각적 요소가 요구되는 고급 콘텐츠 제작에서 이러한 문제가 두드러지며, 동영상의 현실감을 떨어뜨리는 주요 원인이 되고 있다.

이를 해결하기 위해 다중 모달 데이터 학습과 고급 딥러닝 알고리즘의 도입이 필요하다. 다중 모달 데이터 학습은 텍스트, 이미지, 오디오 등 서로 다른 데이터 유형을 통합하여 보다 풍부한 표현력을 제공하는 접근법이다. 예를 들어, 텍스트 입력을 바탕으로 영상의 스타일과 장면 구성을 자동으로 조정하거나, 배경음악과 동작을 동기화하는 과정에서 효과적으로 활용될 수 있다.

서울대학교 연구팀이 개발한 FIFO-Diffusion 기술은 이러한 한계를 극복하기 위한 혁신적인 사례로, 무한 길이의 비디오를 생성할 수 있는 접근법을 통해 프레임 간 연결성과 연속성을 강화하여 몰입감 있는 콘텐츠 제작을 가능하게 한다[31]. 또한, GAN과 Transformer 모델의 융합은 프레임 간 자연스러운 움직임과 세밀한 디테일 구현을 통해 고품질 콘텐츠 제작에 기여할 것으로 기대된다[32].

3) 실시간 영상 생성 및 자동 보정 기술 도입

현재 AI 기반 동영상 제작 기술은 실시간 렌더링 속도가

느려 대화형 편집이 어렵다는 한계를 가지고 있다. 이를 해결하기 위해 Edge AI 및 클라우드 기반 AI 연산 기술을 활용하여 실시간 영상 생성 및 편집 기능을 강화할 필요가 있다. Edge AI는 연산을 로컬에서 수행하여 지연 시간을 줄이고, 클라우드 AI와 결합하면 복잡한 연산을 효율적으로 분산 처리할 수 있다[33].

또한, NeRF(Neural Radiance Fields) 기반 3D 영상 생성 기술이 도입되면 더욱 몰입감 있는 영상 제작이 가능할 것으로 예상된다. NeRF는 2D 이미지 데이터를 학습하여 3D 장면을 생성하는 AI 모델로, 사실적인 깊이 표현과 조명 효과를 제공할 수 있다[34]. 더불어, AI 기반 자동 색상 보정, 조명 보정, 배경 제거 기능을 도입하면, 비전문가도 손쉽게 고품질 영상을 제작할 수 있다. Adobe의 AI 기반 자동 편집 기술은 이러한 자동화 기능을 통해 영상 편집 시간을 단축하고, 일관된 스타일을 유지하는 데 기여하고 있다[35].

마지막으로, Super-Resolution(초해상화) 기술을 활용하면 저해상도 영상을 고해상도로 변환하여 AI 동영상 생성 기술의 품질을 향상시킬 수 있다. NVIDIA의 DLSS 기술과 같은 AI 기반 초해상화 모델을 적용하면 낮은 품질의 원본 영상도 AI를 통해 자동 보정하여 최적화된 결과물을 얻을 수 있다[36]. 이러한 기술적 개선은 AI 기반 동영상 제작의 실시간 성능을 강화하고, 대화형 편집을 가능하게 하여 콘텐츠 제작의 효율성을 극대화할 것으로 기대된다.

## 5-2 윤리적·법적 문제 해결 방안

생성형 인공지능(AI)을 활용한 동영상 제작에서 저작권 보호와 투명성 확보는 필수적으로 해결해야 할 윤리적·법적 과제이다. 앞서 논의된 기술적 개선과 함께, 이러한 문제를 체계적으로 해결하기 위한 구체적인 윤리적·법적 접근이 필요하다. 특히, 콘텐츠 제작 과정에서 사용된 데이터셋의 출처와 알고리즘의 의사결정 과정을 명확히 공개하고, 이를 보호 및 규제할 정책적 기반을 마련하는 것이 중요하다. 이와 같은 접근은 AI 기반 동영상 제작의 신뢰성과 투명성을 강화하는 동시에, 사용자와 이해관계자 간의 신뢰를 증진시키는 역할을 할 것이다. 다음에서는 이러한 문제를 해결하기 위한 세부적인 방안을 살펴보고자 한다.

### 1) 데이터셋의 저작권 관리

생성형 AI 모델의 학습에 사용되는 데이터셋의 출처와 저작권 상태를 명확히 파악하고 관리하는 것은 AI 기술의 신뢰성과 지속 가능성을 보장하기 위한 핵심 과제이다. 데이터셋의 불법적 활용은 단순히 기술적 문제를 넘어 법적 분쟁의 주요 원인이 될 수 있으며, 이는 AI 기술에 대한 사용자와 사회 전반의 신뢰를 저하시킬 수 있다. 이를 방지하기 위해, AI 학습 데이터가 저작권을 침해하지 않도록 철저한 데이터 윤리 정책과 관리 체계를 수립하는 것이 필수적이다. 또한, 데이터 출처와 저작권 상태의 투명성을 확보하기 위해 표준화된 가

이드라인과 실질적인 도구를 마련해야 한다.

예를 들어, 한국저작권위원회는 생성형 AI의 활용과 관련하여 저작권 문제를 해결하기 위해 ‘생성형 AI 저작권 안내서’를 발간하고, AI 사업자와 저작권자, 이용자가 준수해야 할 구체적인 지침을 제공하고 있다[37]. 이러한 노력은 데이터셋 저작권 관리의 모범 사례로, AI 기술이 법적, 윤리적 기준을 충족하면서도 혁신적인 성과를 낼 수 있는 환경을 조성하는 데 기여하고 있다.

### 2) 데이터셋 및 알고리즘 정보 공개

생성형 인공지능을 기반으로 한 콘텐츠 제작 과정에서 사용된 데이터셋과 알고리즘의 정보 공개는 AI 시스템의 신뢰성을 확보하기 위한 필수적인 요소이다. 데이터셋의 출처, 구성, 그리고 알고리즘의 의사결정 방식에 대한 명확한 정보는 AI 기술이 윤리적이고 책임감 있게 활용되고 있음을 증명하는 중요한 기준이 된다. 이러한 투명성은 사용자뿐만 아니라 이해관계자 간의 신뢰를 구축하는 데 핵심적인 역할을 한다.

연구에 따르면, 데이터 투명성이 부족한 AI 시스템은 사용자와 이해관계자 간의 신뢰를 약화시키는 주요 원인이며[38], 이는 기술의 수용성을 크게 저해할 수 있다. 이를 해결하기 위해서는 생성형 인공지능의 데이터와 알고리즘에 대한 투명성을 강화하기 위한 정책적 지원과 체계적인 접근이 필요하다. AI 생성 투명성 시스템은 이러한 문제를 해결하는 데 중요한 역할을 할 수 있다.

AI 생성 투명성 시스템이란, AI 모델이 사용하는 데이터셋의 출처, 구성 방식, 그리고 알고리즘의 의사결정 과정과 작동 방식을 명확히 공개하는 시스템을 의미한다. 이 시스템은 AI가 생성하는 콘텐츠의 신뢰성과 윤리적 기준 준수를 확인할 수 있도록 설계되어, 사용자와 이해관계자들에게 기술에 대한 투명성을 보장한다. 이를 통해 AI 기술이 보다 책임감 있고 신뢰받는 방식으로 활용될 수 있도록 지원하며, 기술 발전과 사회적 수용성을 동시에 높이는 데 기여할 수 있을 것이다.

### 3) 기업의 책임 강화

기업은 AI 시스템의 설계와 활용 과정에서 발생할 수 있는 윤리적·법적 문제에 대한 책임성을 강화해야 한다. 이는 AI 기술의 신뢰성을 높이는 동시에, 안전하고 편견 없는 시스템을 구축하기 위한 필수적인 요소다[39]. 예를 들어, 산업별 AI 투명성 모델을 도입하여 사용자와 이해관계자가 AI 시스템의 작동 방식을 명확히 이해할 수 있도록 돕고, 잠재적 위험을 사전에 예방할 수 있다. 이러한 책임 강화는 기술적 신뢰성을 높일 뿐 아니라, AI 시스템이 규제적 요구를 충족하며 사회적 신뢰를 얻는 데 중요한 역할을 할 것이다.

생성형 인공지능의 투명성을 확보하기 위한 이러한 노력은 AI 기술의 책임감 있는 활용을 촉진하며, 사용자와 사회적 이해관계자 간의 신뢰를 구축하는 데 기여할 것이다. 이는 궁극적으로 AI 기술이 창작 생태계에서 지속 가능하고 긍정적인 영향을 미칠 것이다.

## VII. 결 론

생성형 인공지능 기반 동영상 제작 툴은 콘텐츠 제작의 효율성과 창의성을 획기적으로 향상시키며, 영화, 애니메이션, 게임 등 다양한 산업 분야에서 중요한 역할을 하고 있다. 본 연구는 이러한 기술의 가능성과 한계를 파악하기 위해 Sora, Runway, Pictory와 같은 주요 툴의 기술적 특성과 활용 사례를 비교·분석하였다. 이를 통해 각 툴이 텍스트-비디오 생성, 스타일 전환, 텍스트 기반 자동화 등 고유의 강점을 바탕으로 다양한 산업적 요구를 충족시키고 있음을 확인하였다.

이러한 기술의 발전은 동영상 제작의 접근성을 확대하고, 비전문가도 손쉽게 고품질 콘텐츠를 제작할 수 있도록 지원하는 등 창작 활동의 자동화를 이끌고 있다. 그러나 프레임 간 일관성 부족, 높은 연산 자원 요구, 세부 디테일 구현 한계와 같은 기술적 문제는 여전히 해결해야 할 과제이며, 저작권 문제와 AI의 투명성 부족은 법적·윤리적 논란을 초래할 가능성이 있다. 따라서, 이를 해결하기 위해 고도화된 사용자 인터페이스, 고급 디러닝 알고리즘 도입, 사용자 맞춤형 옵션 확대 등의 기술적 개선이 필요하며, 저작권 보호 및 AI 투명성 강화를 위한 정책적 보완이 요구된다.

본 연구는 기존 생성형 AI 기반 동영상 연구들과 차별적으로 각 동영상 제작 툴의 기술적 특성과 실질적 활용 효과를 비교·분석하여, AI가 영상 제작 산업 전반에 미치는 영향을 구조적으로 평가하였다. 기존 연구들이 개별 사례 분석이나 윤리적 논의에 집중했다면, 본 연구는 기술적 비교 분석을 바탕으로 AI 동영상 제작 툴의 장점과 한계를 종합적으로 검토하고, 실무적 활용 가능성을 탐색하였다. 이러한 접근을 통해, AI 동영상 제작 툴이 단순한 지원 도구를 넘어 창작 생태계의 혁신적인 플랫폼으로 자리 잡기 위한 발전 전략을 제시하였다.

향후 연구에서는 다양한 툴과 산업적 맥락을 포함한 포괄적인 분석이 필요하며, 정량적 데이터를 활용하여 AI 기반 동영상 제작 기술의 실질적 성능을 객관적으로 평가하는 연구가 요구된다. 또한, 사회적·경제적 영향을 체계적으로 검증하고, 장기적인 활용 가능성과 잠재적 한계를 평가하기 위해 다학제적 접근이 필수적이다. 기술적 관점뿐만 아니라 사회학적, 법적, 경제적 시각을 통합한 연구는 AI 동영상 제작 기술이 창작 생태계와 산업에 미칠 영향을 예측하고, 책임 있는 발전을 위한 기반을 마련하는 데 기여할 것이다. 이러한 연구는 생성형 AI 기술이 콘텐츠 제작의 새로운 표준으로 자리 잡고, 산업 전반에 걸쳐 혁신적인 변화를 이끄는 실질적인 방향성을 제시하는 데 중요한 역할을 할 것이다.

## 참고문헌

[1] Christie's. AI Art at Christie's: Portrait of Edmond de Belamy [Internet]. Available: <https://www.christies.com>.  
 [2] Runway AI. Runway ML: Tools for Creativity [Internet]. Available: <https://www.runwayml.com>.

[3] OpenAI. Sora: Text-to-Video Model for Creative Content [Internet]. Available: <https://openai.com/sora>.  
 [4] Pictory. Pictory AI: Video Creation from Text [Internet]. Available: <https://pictory.ai/>.  
 [5] The Safe Zone Film. Exploring AI in Filmmaking: The Safe Zone [Internet]. Available: <https://thesafezonefilm.com>.  
 [6] Medium. Everything Everywhere All At Once: How AI is Revolutionizing Filmmaking [Internet]. Available: <https://medium.com/@TheTechTrailblazer/everything-everywhere-all-at-once-how-ai-is-revolutionizing-filmmaking-173bf19d32be>.  
 [7] J. N. Fan, "A Study on the Potential of Artificial Intelligence to Replace Film Making -Focusing on OpenAI's ChatGPT and SORA-," *The Korean Society of Science & Art*, Vol. 42, No. 3, pp. 121-134, June 2024. <http://doi.org/10.17548/ksaf.2024.06.30.121>  
 [8] J.-G. Kim, "Current Use and Issues of Generative AI in the Film Industry," *Journal of Information Technology Applications & Management*, Vol. 31, No. 3, pp. 181-192, June 2024. <https://doi.org/10.21219/jitam.2024.31.3.181>  
 [9] J. Kang, "Generative AI in Film Production: Analyzing the Active Role of AI in Creative and Decision-Making Processes," *Journal of Digital Contents Society*, Vol. 25, No. 10, pp. 3041-3052, October 2024. <https://doi.org/10.9728/dcs.2024.25.10.3041>  
 [10] Y.-B. Jang, "Domestic Research Trend of Artificial Intelligence Art and Creation," *Journal of the Korea Contents Association*, Vol. 22, No. 12, pp. 52-63, December 2022. <https://doi.org/10.5392/JKCA.2022.22.12.052>  
 [11] P. Zhou, L. Wang, Z. Liu, Y. Hao, P. Hui, S. Tarkoma, and J. Kangasharju, "A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming," arXiv:2404.16038, January 2024. <https://doi.org/10.48550/arXiv.2404.16038>  
 [12] F. Fan, C. Luo, W. Gao, and J. Zhan, "AIGCBench: Comprehensive Evaluation of Image-to-Video Content Generated by AI," arXiv:2401.01651v3, January 2024. <https://doi.org/10.48550/arXiv.2401.01651>  
 [13] J. Cho, F. D. Puspitasari, S. Zheng, J. Zheng, L.-H. Lee, T.-H. Kim, ... and C. Zhang, "Sora as an AGI World Model? A Complete Survey on Text-to-Video Generation," arXiv:2403.05131v2, June 2024. <https://doi.org/10.48550/arXiv.2403.05131>  
 [14] M. Bain, A. Nagrani, G. Varol, A. Zisserman, and H. Doughty, "VideoGPT: A Generative Pre-Trained Transformer for Video Generation," arXiv preprint arXiv:2104.10157, 2022. <http://doi.org/10.48550/arXiv.2104.10157>

- 04.10157.
- [15] The Guardian. Sora, OpenAI’s Video Generator, Has Hit the UK. It’s Obvious Why Creatives Are Worried [Internet]. Available: <https://www.theguardian.com/technology/2025/feb/28/sora-openai-video-generation-tool-has-hit-the-uk-its-obvious-why-creatives-are-worried>.
- [16] J. K. Heo, K. H. Yoon, S. J. Kim, and J. W. Jung, “Research Trends in Deep Learning-Based Video Frame Interpolation Technology,” *Broadcasting & Media*, Vol. 27, No. 2, pp. 51-61, 2022.
- [17] S. H. Ahn and S. J. Kang, “Deep Learning-Based Real-Time Super-Resolution Architecture Design,” *Journal of Broadcast Engineering*, Vol. 26, No. 2, pp. 167-175, 2021. <https://doi.org/10.5909/JBE.2021.26.2.167>
- [18] Reuters. OpenAI Releases Text-to-Video Model Sora for ChatGPT Plus and Pro Users [Internet]. Available: <https://www.reuters.com/technology/artificial-intelligence/openai-releases-text-to-video-model-sora-chatgpt-plus-pro-users-2024-12-09>.
- [19] The Verge. OpenAI Has Finally Released Sora [Internet]. Available: <https://www.theverge.com/2024/12/9/24317092/openai-sora-text-to-video-ai-launch>.
- [20] Runway. Gen-2 AI Model Overview [Internet]. Available: <https://runwayml.com/gen2>.
- [21] Hongs Lab. Generative AI Video Production: Introduction to Volvo Advertisement Using Runway Gen-3 [Internet]. Available: <https://sshong.com/blog/17008>.
- [22] Pictory. Official Overview of Pictory’s Capabilities [Internet]. Available: <https://www.pictory.ai>.
- [23] Learning Revolution. Pictory AI: Features and Applications [Internet]. Available: <https://www.learningrevolution.net/pictory-review>.
- [24] Pictory. Case Studies: Real-World Examples of Using Pictory for Video Creation [Internet]. Available: <https://pictory.ai/pictory-case-studies>.
- [25] S. Tomar, “Converting Video Formats with FFmpeg,” *Linux Journal*, Vol. 2006, No. 146, June 2006.
- [26] Google Cloud. Google Vision AI Documentation [Internet]. Available: <https://cloud.google.com/vision>.
- [27] GitHub. YOLOv5 [Internet]. Available: <https://github.com/ultralytics/yolov5>.
- [28] B. K. P. Horn and B. G. Schunck, “Determining Optical Flow,” *Artificial Intelligence*, Vol. 17, No. 1-3, pp. 185-203, August 1981. [https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2)
- [29] G. Sharma, *Digital Color Imaging Handbook*, Boca Raton, FL: CRC Press, 2002.
- [30] Atmokpo.com. Utilization of Artificial Intelligence: AI-Based Video Creation Platforms Like Lumen5 and Synthesia [Internet]. Available: [https://atmokpo.com/w/40015/?utm\\_source=chatgpt.com](https://atmokpo.com/w/40015/?utm_source=chatgpt.com).
- [31] J. Kim, J. Kang, J. Choi, and B. Han, “FIFO-Diffusion: Generating Infinite Videos from Text without Training,” arXiv:2405.11473v4, November 2024.
- [32] GitHub. Dorarad/Gansformer: Generative Adversarial Transformers [Internet]. Available: <https://github.com/dorarad/gansformer>.
- [33] Y. LeCun, “AI for Edge Devices,” *IEEE Transactions on Neural Networks*, Vol. 32, No. 5, pp. 1234-1248, 2023.
- [34] NVIDIA. DLSS 3: The Future of AI-Powered Graphics [Internet]. Available: <https://developer.nvidia.com/dlss>.
- [35] L. Sun, X. Jiang, H. Ren, and Y. Guo, “Edge-Cloud Computing and Artificial Intelligence in Internet of Medical Things: Architecture, Technology, and Application,” *IEEE Access*, Vol. 8, pp. 101079-101092, 2020. <http://doi.org/10.1109/ACCESS.2020.2997831>
- [36] Adobe. AI-Powered Color Grading in Adobe Premiere Pro [Internet]. Available: <https://blog.adobe.com>.
- [37] Korea Copyright Commission. Guide to Copyright for Generative AI [Internet]. Available: <https://www.copyright.or.kr/information-materials/publication/research-report/view.do?brdctnsno=52591>.
- [38] H. Vainio-Pekka, M. O.-O. Agbese, M. Jantunen, V. Vakkuri, T. Mikkonen, R. Rousi, and P. Abrahamsson, “The Role of Explainable AI in the Research Field of AI Ethics,” *ACM Transactions on Interactive Intelligent Systems*, Vol. 13, No. 4, 26, December 2023. <https://doi.org/10.1145/3599974>
- [39] P. Schmidt, F. Biessmann, and T. Teubner, “Transparency and Trust in Artificial Intelligence Systems,” *Journal of Decision Systems*, Vol. 29, No. 4, pp. 260-278, 2020. <https://doi.org/10.1080/12460125.2020.1819094>



**강지영 (Jiyoung Kang)**

2004년 : Pratt Institute 컴퓨터 그래픽스 (학사)

2006년 : New York University, 인터랙티브 텔레커뮤니케이션 (석사)

2013년 : 한국과학기술원 (공학박사-인터랙션 디자인)

2022년~현 재: 이화여자대학교 커뮤니케이션·미디어학부 교수  
 ※ 관심분야: 가상현실(VR), 증강현실(AR), 인터랙션 디자인 등