

AI 기반 영상 콘텐츠 저작도구의 유형 및 동향

이 세 희¹ · 이 진^{2*}

¹한양대학교 문화콘텐츠학과 박사과정

²한양대학교 ERICA 문화콘텐츠학과 조교수

Types and Trends in Artificial Intelligence-Based Video Content Authoring Tools

SeHui Yi¹ · Jin Lee^{2*}

¹Ph.D. Course, Department of Culture Contents, Hanyang University, Seoul 04763, Korea

²Assistant Professor, Department of Culture Contents, Hanyang University ERICA, Ansan 15588, Korea

[요 약]

ChatGPT가 공개된 이후, AI가 적용된 다양한 콘텐츠 저작도구가 등장했다. 이러한 시점에서 본 연구는 생성형 AI를 메타미디어(meta-medium)으로 규정하고, AI 기반 영상 콘텐츠 저작도구의 양상을 통시적·공시적으로 고찰하는 것이 목적이다. 영상 저작도구의 통시적 발전 양상을 살피고, 영상 제작 단계와 입력·출력 개념을 기준으로 AI 기반 영상 콘텐츠 저작도구를 범주화함으로써 관련 동향에 대한 거시적이고 통합적인 접근을 시도했다. 그 결과 AI 영상 저작도구는 크게 통합체(syntagma) 구조와 계열체(paradigm) 구조로 진화하고 있었다. 통합체 측면에서의 변화는 영상 제작 효율 상승뿐만 아니라, 영상 창작 주체의 확장과 기존 영상 콘텐츠 제작 방식의 다양화를 촉진한다. 계열체 측면에서의 변화는 영상 제작 시간과 비용을 절감하여 영상 제작 효율의 상승을 중심으로 이루어진다. 이 같은 변화는 지금까지 통용되었던 영상 제작 리터러시의 변화가 이루어질 수 있음을 시사한다.

[Abstract]

Since the release of ChatGPT, various artificial intelligence (AI)-enabled content authoring tools have emerged. At this point, this study aims to define generative AI as a meta-medium, and to examine the aspects of AI-based video content authoring tools from a diachronic and synchronic perspective. We analyzed the development of video authoring tools over time and categorized AI-based video content authoring tools by production stages and input-output concepts to provide an integrated overview of current trends. Therefore, we found that AI video authoring tools are evolving into syntagma and paradigm structures. The change in the syntagma aspect improves video production efficiency and promotes the expansion of video creation entities and diversification of existing video content production methods. The paradigm shift centers on increasing the efficiency of video production by reducing the time and cost of video production. These changes suggest that the video production literacy that has been prevailing until now may change.

색인어 : 인공지능 콘텐츠, AI 창작 도구, AI 영상 프로그램, 인공지능 영상 유형, 생성형 AI

Keyword : AI Content, AI Creation Tools, AI Video Program, AI Video Types, Generative AI

<http://dx.doi.org/10.9728/dcs.2024.25.6.1589>



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 26 April 2024; Revised 30 May 2024

Accepted 11 June 2024

*Corresponding Author; Jin Lee

Tel: +82-31-400-5436

E-mail: zzin2024@hanyang.ac.kr

1. 서론

본 연구는 생성형 AI를 메타미디어(meta-medium)으로 규정하고, AI 기반 영상 콘텐츠 저작도구의 양상을 기호학의 통합체·계열체 개념을 통해 고찰한다. 이를 위해 영상 저작도구의 통시적 발전 양상을 살피고, 영상 제작 단계와 입력, 출력 개념을 기준으로 AI 기반 영상 콘텐츠 저작도구를 범주화함으로써 관련 동향에 대한 거시적이고 통합적인 접근을 시도한다.

2022년 하반기 OpenAI에서 ChatGPT가 출시된 이후, 인공지능에 관한 관심이 폭발적으로 증가했다. ChatGPT뿐만 아니라, Midjourney, NovelAI 같은 이미지 생성 AI, MusicGen 등의 사운드 생성 AI, Runway Gen-2, Sora 등의 영상 생성 AI도 주목받으면서 텍스트, 이미지, 사운드, 영상 등 다양한 양식의 콘텐츠 창작에 변화가 일고 있다.

레프 마노비치는 모든 창작 산업 부문에서 호환성 있는 저작도구를 사용하는 제작 환경과 작업흐름이 채택되면서 미디어 디자인 영역에 많은 변화가 나타났다고 주장했다. 예시로 1990년대 후반에 애프터 이펙트 및 관련 소프트웨어의 채택이 새로운 애니메이션 비주얼을 의미하는 전문 용어인 ‘모션 그래픽’의 채택으로 이어졌다[1]. 도구였던 컴퓨터는 단순 범용 기계가 아닌 문화적 미디어가 된 것이다. 오늘날의 모든 미디어 및 문화콘텐츠 생산의 공통된 특성이 미디어 저작과 관련된 응용 소프트웨어에 의해 만들어졌다는 점을 간과할 수 없다. 이는 미디어 연구에 소프트웨어 관점이 필요한 이유이다. 따라서 저작도구인 소프트웨어의 진화 양상을 살펴보는 것은 영상 콘텐츠의 변화를 조망하기 위한 필수 과정이다.

소프트웨어 쪽으로 확장하면 레프 마노비치가 진화의 개념을 활용하여 소프트웨어를 분석한 연구가 있다. 마노비치는 진화의 개념 중 하나인 ‘혼종화(hybridization)’를 새로운 미디어의 발명에 대한 핵심적 메커니즘으로 보고 소프트웨어와 미디어를 분석했다. 혼종적 미디어는 유사 개념인 멀티미디어와 비교하면, 다중적 미디어 유형의 결합이라는 공통점을 갖고 있으나 궁극적으로는 방향성이 다르다[1]. 멀티미디어는 다른 미디어 유형이 병렬적으로 위치하는 응용 프로그램이나 전자문서를 뜻한다. 파워포인트나 웹페이지처럼 다중적 미디어 유형을 특정 명령어를 통해 삽입하거나 첨부하는 방식이다. 따라서 삽입된 콘텐츠가 병렬적으로 분리되어 있고, 고유한 언어를 유지하며 인터페이스와 기술이 분리되어 있어 미디어의 고유 구조에는 영향을 미치지 않는다. 반면, 혼종적 미디어는 모든 요소가 결합 되어 새로운 미디어가 발생한다. 예시로 구글 어스의 경우, 위성사진인 표상물과 3D 인터페이스가 결합하여 새로운 표상물과 인터페이스를 만들었다. 이처럼 미디어 언어는 혼종화를 통해 새로운 속성을 획득하고 재구조화를 통해 더욱 풍부해진다[1]. 콘텐츠 분야에서 혼종화의 대표적인 예시로는 웹툰을 들 수 있다. 만화라는 양식이 컴퓨터, 인터넷과의 혼종화를 통해 새로운 속성을 가진 웹툰이 된 것이다. 케이와 골드버그는 이처럼 질적으로 다르고 역사적으로

전례가 없는 변화를 일으키는 그 무엇을 “메타미디어”이란 용어로 표현했으며 컴퓨터를 메타미디어로 정의했다[2]. 지금까지 컴퓨터와 스마트폰이 하드웨어로써 메타미디어이라면, 인공지능은 소프트웨어로써 메타미디어로 볼 수 있다. 따라서 인공지능 시대에 AI 저작도구를 연구하는 것은 미디어와 콘텐츠의 변화를 예측하고 분석하는 데 매우 중요하다.

영상 콘텐츠 저작도구와 관련된 선행 연구의 경우 그 변화를 직접적으로 조망한 연구는 없었으나, 권승태가 대표적인 영상 저작도구 중 하나인 파이널컷프로X의 기호학적 분석을 시도한 연구가 있다[3]. 해당 연구는 프로그램의 시각 정체성을 도출하는 데에 의미는 있으나 동향 측면에서 통합적 관점을 제시하지는 않았다. 정다영, 김승인은 디자인 업무 유형에 따라 그래픽 저작도구의 사용 현황을 분석했지만, 인공지능이나 영상 저작도구에 관한 내용은 없었다[4].

본 논문에서는 다양한 분야의 AI 저작도구 중에서도 영상 콘텐츠 저작도구에 주목한다. 텍스트와 이미지, 사운드, 모션 등이 결합된 복합 양식인 영상은 스마트폰을 통한 저작도구의 대중화와 함께, 만들어진 영상을 공유하여 소비할 수 있는 플랫폼인 유튜브, 틱톡, 인스타그램 등과 맞물려 폭발적인 성장 중이다[5]. 이미 영상 콘텐츠는 시각적, 청각적 만족을 비롯해 정보전달 역할까지 확장하며[6] 우리 일상에서 매우 큰 부분을 차지하고 있다. 점점 더 편리해지는 영상 저작도구와 공유 플랫폼의 대중화는 창작과 소비의 경계를 무너뜨리고 있다. AI 영상 저작도구 Vrew는 영상 제목만 입력해도 영상 내용을 작성하고, 이를 토대로 데이터베이스에 있는 영상 클립과 생성 이미지를 조합하여 몇십초 만에 1분짜리 영상을 만든다. Runway Gen-2와 Sora는 장면에 관한 설명을 텍스트로 입력하면, 이미지나 영상 소스 없이 AI가 텍스트 지문을 영상으로 바꿔준다. 창작자가 영상에 관한 리터러시 없이도 영상 콘텐츠 제작이 가능하게 되었다. 이 같은 AI가 가져온 영상 콘텐츠 창작의 변화는 장기적으로 콘텐츠 산업에 큰 영향을 미칠 것이다.

2024년 3월을 기준으로 AI를 활용한 영상 콘텐츠 저작도구는 저마다 다른 방법으로 다채로운 형태의 영상을 제작할 수 있다. 인공지능이 영상에 미친 메타미디어 측면에서의 변화를 통찰하기 위해서는 현시점에서 동향 분석을 통한 체계적인 정리가 필요하다. 현재 일부 생성 AI 저작도구의 사례 연구[7]와 상대적으로 먼저 개발이 이루어졌던 사운드 AI 분야의 동향 연구[8]는 있었지만, 통합적 관점에서 AI 영상 저작도구를 유형화한 동향 연구는 없었다.

본 연구의 목적은 AI 기반 영상 콘텐츠 저작도구의 동향을 파악하여 AI로 인한 영상 제작 방식의 변화를 통합적으로 조망하는 것이다. 이를 통해 인공지능이 영상 콘텐츠 산업 분야에서 메타미디어으로써 어떤 변화를 불러올지 고찰하고자 한다.

영상 콘텐츠는 이미지, 텍스트, 사운드 등 여러 요소를 활용하는 복합 양식이다. 따라서 AI 영상 콘텐츠 저작도구의 제작 방식은 주로 입력 소스에 따라 다른 양상으로 나타난다. 인공지능이 등장하기 전의 영상 저작도구는 촬영된 영상 소

스 위주로 편집하는 방식이었으나, 인공지능이 등장한 이후에는 영상이 아닌 이미지나 사운드, 텍스트 등의 소스로 영상을 제작할 수 있게 되었다. 이에 본 연구에서는 입력(input), 출력(output)을 핵심 기준으로 삼아 AI 영상 콘텐츠 저작도구의 동향을 파악하고자 한다. 이를 선별한 기준은 입력, 출력물에 따른 활용 기능 및 적용 기기에 따라 영상 제작 방식에 드러나는 특징이 있는지를 중점적으로 보았다. 또 인공지능을 활용한 기능의 다양성과 편의성을 고려하여 선정하였다. 마지막으로 보다 많은 수의 사용자층이 접할 수 있다는 측면을 염두에 두어, 전면 유료 서비스보다는 체험이 가능한 부분 유료나 무료 소프트웨어를 우선순위로 고려하였다.

본 연구에서는 거시적이고 통합적으로 영상 제작 소프트웨어를 조망하기 위해 2장에서는 영상 콘텐츠 저작 소프트웨어가 어떻게 AI까지 오게 되었는지 통시적 관점에서 영상 제작 도구 패러다임의 발전 양상을 살핀다. 발전 양상은 시대별 창작 주체의 변화 및 주된 패러다임에 따라 구분하여 정리한다. 3장에서는 공시적 분석 관점에서 2024년 3월까지 출시된 주요 AI 영상 콘텐츠 저작도구를 영상 제작 단계와 입력(input), 출력(output) 개념을 적용하여 사례별 유형화를 시도한다. 유형화된 저작도구는 기초화적 개념인 통합체와 계열체를 통해 분석한다. 이를 통해 AI 영상 저작도구를 조망할 수 있는 통합적 관점을 제시할 것이다. 4장에서는 AI를 활용한 영상 저작도구의 동향 분석을 통한 시사점 및 연구의 한계, 산업 전망 등을 논의한다.

II. 영상 저작도구의 통시적 발전 양상

마셜 매클루언은 저서 <미디어의 이해>에서 “우리는 도구를 만들며, 그러고 나면 도구는 우리를 만든다”고 표현했다[9]. 이는 도구가 미치는 영향력이 상상 이상으로 크다는 것을 의미한다. 레프 마노비치는 저서 <소프트웨어가 명령한다>에서 미디어의 변화에 소프트웨어의 중요성을 강조하며, 소프트웨어는 단순히 도구를 넘어 현대 미학에 의도하지 않은 훨씬 더 광범위한 영향을 미친다고 했다[1]. 인공지능이 메타미디어으로써 영상 저작도구에 미칠 영향을 통합적으로 전망하려면, 먼저 통시적 관점에서 영상 저작도구의 발전 양상과 시기별 메타미디어에 따른 제작 방식의 변화, 혼종적 미디어의 출현, 콘텐츠 소비, 창작 주체의 변화 등을 살펴볼 필요가 있다.

영상 제작 초기에는 필름을 직접 잘라서 붙이는 아날로그적이고 선형적인 방식의 편집 방식이었다. 1990년대에 들어서야 컴퓨터를 활용한 디지털 방식의 영상편집 소프트웨어가 보편화되었다[10]. 영상 저작도구의 역사는 편집 위주의 소프트웨어부터 시작하며, 기술의 발전에 따라 다양한 변화를 거쳐왔다[3]. 본 논문에서는 시기별 메타미디어와 함께 전문화, 대중화, 자동화의 키워드를 중심으로 영상 저작도구의 기술적 변화와 창작 주체의 변화, 발전 양상을 살핀다.

2-1 PC 메타미디어: 디지털 편집, 영상 편집의 전문화

메타미디어으로써 PC가 본격적으로 보급되기 시작하며 영상편집도 컴퓨터화가 진행됐다. 이 시기에는 비디오테이프를 저장 매체로 활용한 카메라로 영상을 촬영했다. 따라서 촬영한 아날로그 데이터를 디지털 데이터로 변경하는 과정을 거친 후에야 비로소 컴퓨터를 활용하여 영상을 편집하는 시기였다. 1980년대 후반에 PC 영상 제작 소프트웨어인 Avid사의 Media Composer가 출시되었다. 이 시대의 영상 제작 소프트웨어는 주로 PC에서 작동하였으며, Adobe Premiere, Final Cut Pro, Avid Media Composer 이외에 Sony Vegas 등이 있었다. 이때 출시된 소프트웨어 중 Adobe Premiere Pro와 Final Cut Pro는 2020년대에도 널리 사용되며 디지털 영상 저작도구의 양대 산맥이 되었다. 두 프로그램 모두 영상을 시간에 따라 라인 형태로 펼쳐 놓은 타임라인 형식의 영상편집 방식을 취한다. 타임라인 형식은 소프트웨어가 존재하지 않던 시기에 영상 필름을 펼쳐 놓고 잘라서 편집했던 방식을 디지털로 옮겨 놓은 형태이다. 시간에 따라 길게 펼쳐진 영상 파일을 자르고 붙이는 인터페이스로 현재도 대다수의 영상 저작도구에서 활용된다.

1990년대에 컴퓨터 영상 소프트웨어가 디지털 방식으로 크게 발전했던 반면, 촬영 장비는 일부 아날로그 방식이 남아 있었다. 과거에 비해 센서 등 디지털 방식의 다양한 부품들이 적용되어 카메라의 크기가 작아지고 편리해졌지만, 저장 매체는 아날로그 형태였다. 이 시기에는 miniDV(6mm), Video8(8mm), VHS(16mm) 등의 비디오테이프를 저장 매체로 사용하는 캠코더나 비디오키메라를 이용해 촬영을 진행했다. 촬영을 마치면 데크(Deck)라는 기계를 이용해 아날로그 데이터를 디지털 데이터로 변환하여 컴퓨터에서 후반 편집 작업을 수행했다. 당시 카메라를 비롯해 데크 같은 영상 장비는 굉장히 고가였기 때문에, 영상 창작은 전문가 집단 위주로 전문화된 제작 및 편집이 이루어졌다는 특징이 있다. 제작된 영상 콘텐츠는 주로 TV, 영화 등의 레거시 미디어에서 소비되었다.

2-2 모바일 메타미디어: 촬영 데이터의 디지털화, 스마트폰 보급과 영상 제작 대중화

2000년대 중반부터는 편집뿐만 아니라 촬영 및 데이터 저장 방식도 디지털화가 시작됐다. 기존 캠코더 및 카메라와 더불어, 고화질 사진을 찍던 DSLR 카메라나, 콤팩트한 크기의 디지털카메라로 영상을 촬영하는 경우가 많아졌다. 영상 장비의 가격도 과거보다 많이 저렴해져서 일반인 영상 제작에 대한 진입장벽이 허물어지는 시기다. 영상 저장 매체도 비디오테이프에서 디지털 방식인 SD카드 형태로 변화하여, 데이터 저장 및 보관이 용이해졌다.

특히 2000년대 후반에는 초고속 인터넷과 스마트폰 카메라의 발달로 인해 영상 콘텐츠의 소비와 제작이 훨씬 간편해

졌다. 유튜브 같은 영상 스트리밍 서비스가 등장하며 전문가가 아닌 일반인의 영상 제작이 활발해졌고, UCC(User Created Contents)라는 신조어가 등장했다. 창작 주체가 전문가에서 일반인으로 이동하며 전공자 이외의 영상편집 수요가 늘어났고, 이 시기에 출시되었던 영상편집 소프트웨어는 전문가용이 아니라 비전문가를 대상으로 한 무료 PC 영상편집 소프트웨어가 많아졌다. 대표적으로 Shotcut, Avid Media composer First 등이 있다. 하지만 일반인의 영상은 전문가 수준의 영상 퀄리티에 미치지 못했고, 주로 스낵 컬처의 형태로 소비되었으며 대부분 PC에서 인터넷 커뮤니티 사이트를 통해 감상할 수 있었다.

2010년 이후에는 메타미디어으로써 스마트폰이 빠르게 보급되었다. 모바일 기기로 인해 생활에 큰 변화가 생겼으며 영상편집 또한 스마트폰으로 수행할 수 있는 어플리케이션이 등장하기 시작했다. 2010년에 출시된 아이폰 기본 앱 iMovie를 비롯해, 다양한 영상편집 어플이 출시되었다. 눈에 띄는 점은 PC 기반 영상 저작도구가 모바일 버전으로도 출시되었다는 점이다. Adobe Premiere Pro는 아이패드와 아이폰용 영상 프로그램으로 Premiere Rush를 출시했고, Vrew도 모바일 버전을 별도로 출시했다.

따라서 창작자는 시간, 공간의 제약을 받지 않고 언제 어디서든 스마트폰으로 영상을 촬영하고 편집할 수 있게 되었다. 2000년대에 전문가에서 일반인으로 창작 주체가 이동했다면, 2010년대에는 창작 주체의 이동뿐 아니라 시공간의 제약마저 사라지게 됐다. 소프트웨어의 사용법 또한 기존 프로그램보다 월등하게 쉬워져서 남녀노소 누구나 보편적으로 영상을 만들 수 있게 됐다. 영상 제작의 진정한 대중화가 이루어진 셈이다. 게다가 영상 길이가 짧아지고 먹방, 브이로그 등 새로운 형태의 콘텐츠가 늘어났다. 제작한 영상 콘텐츠는 유튜브, 틱톡, 인스타그램 등에 올려 수익 창출 또한 가능해지면서 상업적 측면에서의 변화도 이루어졌다.

이 시기에 출시된 모바일 영상 저작도구는 PC 소프트웨어와 비교해서 기능이 축소되었다는 특징이 있다. 이는 대부분 하드웨어적 문제에서 발생한다. 모바일 기기의 성능이 PC와 큰 차이가 있을 뿐만 아니라, 디바이스 화면 크기 또한 PC에 비해 작기 때문이다. 기능이 축소되다 보니, 숏폼 동영상, 여행 브이로그, 뮤직비디오 등 특정 목적에 맞는 영상 콘텐츠만 제작할 수 있는 모바일 어플리케이션이 다수 출시되었다.

2-3 AI 메타미디어: AI를 활용한 영상 제작 자동화

2010년대부터 영상 제작에 부분적으로 인공지능을 활용한 사례는 존재한다. 대표적으로 음성인식을 통한 자동 자막 기능을 제공한 유튜브가 있다. 이후 2018년에 국내 스타트업 VoyagerX가 AI 음성인식을 통해 자동 생성된 자막을 편집할 수 있는 소프트웨어인 Vrew를 출시했지만, 지금까지 영상 편집에 관한 다양한 기능은 없었다. 영상에 인공지능을 본격적으로 활용하기 전에는 다양한 이미지 생성 AI가 공개되었

다. 2021년에 이미지 생성 AI 모델 <DALL-E>가 출시된 이후, 짧은 기간 안에 이미지 생성 인공지능의 성능이 향상되었다. 2022년 7월에 후속 모델인 <DALL-E2>, 같은 해 8월에 텍스트 to 이미지 모델 <Stable Diffusion> 오픈소스가 공개되며 이미지 생성 AI가 대중에게 주목받았다. 영상에서 큰 비중을 차지하는 요소가 이미지인 만큼, 이미지 인공지능의 발전과 함께 영상도 인공지능 시대로 접어들었다.

메타미디어으로써 본격적인 인공지능 시대의 시작은 ChatGPT가 출시된 2022년이다. 이 시기부터 다양한 소프트웨어 기업에서 텍스트로 영상을 생성하는 저작도구를 공개했다. 기존 영상 제작 방식과 달리, 다양한 AI 기능을 활용하여 영상편집 및 제작 과정을 자동화하기 시작했다.

다음 장에서는 AI 기반 영상 콘텐츠 저작도구를 영상 제작 단계와 입력, 출력 개념에 따라 유형화한다. 그리고 기호학적 개념을 활용하여 유형별 사례를 분석하여 AI 기반 영상 콘텐츠 저작도구의 동향을 파악할 것이다.

III. AI 기반 영상 콘텐츠 저작도구의 유형

본 논문에서는 레프 마노비치가 주장한 ‘미디어’이란 개념에서 착안한 입력(Input), 출력(Output) 개념과 일반적인 영상 콘텐츠 제작 단계를 활용하여 AI 기반 영상 콘텐츠 저작도구의 유형화를 시도해보고자 한다.

‘미디어’이라는 용어는 사전적 의미로는 중간이라는 의미를 가지고 있으며 일반적으로는 매개, 매체라는 뜻으로 통용된다. 예술 쪽에서는 표현의 수단이나 소재라는 뜻으로도 쓰인다. 마노비치가 주장하는 ‘미디어’이란 개념은 ‘알고리즘’과 ‘데이터 구조’의 결합을 의미한다. 알고리즘과 데이터 구조는 컴퓨터 프로그래밍의 두 가지 근본적인 요소이다. 마노비치는 이 두 가지 요소가 응용 프로그램이 이용자에게 어떻게 제시되는지, 응용 프로그램을 사용할 때 무엇을 할지에 대한 이용자의 이해를 규정한다고 주장했다. 따라서 응용 프로그램이 규정하고 이용자가 경험하는 미디어는 특정한 데이터 구조, 그리고 그런 구조로 저장될 콘텐츠를 만들고 편집하고 보기 위한 알고리즘, 그 두 가지가 짝을 이루는 무엇이라고 설명했다[1]. 이 개념에 따르면 AI 기반 저작도구도 미디어의 하나라고 볼 수 있다. 데이터 구조는 입력물의 형식으로 볼 수 있으며, 인공지능이 알고리즘에 해당한다. 하지만 저작도구의 세부적인 알고리즘을 분석하는 것은 기술적으로 불가능하므로 알고리즘을 적용한 결과물인 출력물의 형태로 분석하고자 한다. 본 논문에서는 AI 기능 분류를 위해 영상 콘텐츠 제작 단계를 활용했다. 영상 콘텐츠 제작 공정은 기획과 준비 단계인 프리 프로덕션(Pre-Production), 제작 단계인 프로덕션(Production), 후반 편집 및 가공 단계의 포스트 프로덕션(Post-Production) 단계로 나눌 수 있다[11]. 그림 1은 김기범이 요약한 일반적인 영상 콘텐츠의 제작 공정이다.

본 논문에서는 이를 재구성하여 영상 콘텐츠 제작 단계를

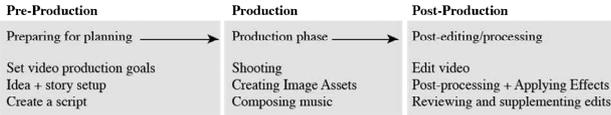


그림 1. 일반적인 영상 콘텐츠의 제작 공정[12]
Fig. 1. Production process for typical video content[12]

크게 기획, 제작(촬영), 편집(가공)의 3단계로 분류한다. 여기서 기획 단계는 영상을 제작하기 전에 작성하는 기획안과 시놉시스, 대본 등을 의미하므로 결과물은 텍스트가 된다. 이 글은 AI 기반 영상 콘텐츠 저작도구의 동향을 파악하는 목적을 가지고 있다. 따라서 아래 그림 2와 같이 기획 단계에 해당하는 AI는 텍스트만 산출할 수 있으므로 유형 분석에서 배제하고, 결과물이 영상으로 나올 수 있는 제작(촬영)과 편집(가공) 단계 중심으로 AI가 어떻게 적용되었는지 분석한다.



그림 2. 영상 콘텐츠 제작 단계별 AI 활용 분류
Fig. 2. Categorizing AI for video content creation

3-1 제작(촬영) 단계: 형상화 AI

본 논문에서는 영상 제작(촬영) 단계에 사용되는 AI는 형상화 AI라고 정의한다. 형상화 AI는 영상 이외의 입력(텍스트, 이미지, 사운드)을 받아 영상으로 출력한다. 종종 형상화와 변형을 서로 유사한 개념으로 보고 혼용하는 경우가 있기에, 본 논문에서는 그림 3과 같이 입력물과 출력물의 형태를 기준으로 분류한다. 즉, 입력물이 영상이 아니고, 출력물이 영상일 경우에만 형상화 AI로 보는 것이다. 입력물과 출력물이 모두 영상인 경우는 편집 단계의 변형 AI로 본다.

형상화 AI는 기존 영상편집 방식에서 볼 수 없었던 인공지능만의 독특한 영상 제작 방식이다. 텍스트, 이미지 등 영상이 아닌 매체로 영상을 형상화할 수 있는 유형으로, 영상 제작에 대한 지식 없어도 영상을 만들 수 있다. 따라서 인공지능 시대에 영상 제작 방식과 산업에 큰 변화를 야기할 수 있으며, 향후 영상 콘텐츠 분야에서 새로운 장르를 형성할 수도 있기에 학술적 의의를 가진다.

현재 AI를 활용한 영상 콘텐츠 저작도구에서 활용되는 입

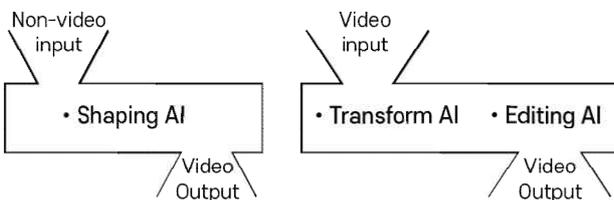


그림 3. 입력물과 출력물에 따른 영상 AI 분류
Fig. 3. Categorizing video AI based on inputs and outputs

력물은 텍스트, 이미지, 사운드로 나눌 수 있다. 이를 텍스트 to 영상, 이미지 to 영상, 사운드 to 영상의 소분류로 나누어 영상 형상화 AI의 사례를 파악하고자 한다.

1) 텍스트 to 영상

텍스트를 영상으로 바꿔주는 영상 저작도구는 생성되는 요소의 종류에 따라 ①지문 생성형과 ②대사 생성형, ③캐릭터 생성형으로 나눌 수 있다.

지문 생성형은 대본에서 지문에 해당하는 프롬프트를 영상으로 표현하는 유형이다. 장면에 관한 설명과 연출을 텍스트로 받아 영상으로 만드는 형태로 Runway Gen-2, Sora, Moonvalley, Pika, Genmo 등이 있다. 2024년 3월 기준, 대부분의 저작도구는 한 번에 긴 영상을 생성하진 못하고 한 프롬프트 당 샷(Shot, 영상을 구성하는 가장 작은 단위[13])과 유사한 형태의 초 단위 영상을 만들어 준다. 이를 보완하기 위해 가장 최근 공개된 Sora의 경우는 1분 분량의 고화질 비디오를 생성할 수 있다고 자체 홈페이지에 설명되어 있다[14]. Sora가 다른 저작도구와 비교해서 긴 영상을 생성하고 성능이 좋다는 평가를 받는 이유는 Diffusion 모델에 트랜스포머 아키텍처를 결합하여 저차원 latent 공간으로 압축하고 시공간 패치로 분해하는 등의 기술을 적용했기 때문이다. 이를 도식으로 표현하면 그림 4와 같다.

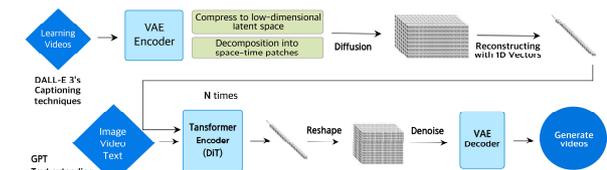


그림 4. Sora의 동영상 생성 AI 추정 아키텍처[15]
Fig. 4. Sora's video generation AI estimation architecture[15]

Runway Gen-2의 경우, 영상 생성 시 카메라 모션과 스타일을 선택할 수 있는 기능이 있어 일부 연출도 가능하다. 하지만 네 프로그램 모두 별도의 자막이나 텍스트, 사운드는 생성하지 않고, 오로지 입력한 지문을 표현한 영상 그 자체만 출력한다. 그림 5처럼 생성 기능에 초점을 맞춘 대신, 영상을 자

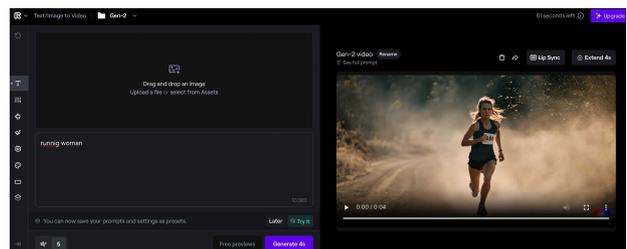


그림 5. Runway Gen-2의 영상 생성 장면 스크린 캡처
Fig. 5. Screen capture of Runway Gen-2's video generation scene

르고 붙이는 편집 기능은 없다는 특징이 있다. 기존 영상 제작 도구와는 전혀 다른 인터페이스도 눈에 띈다. 영상을 입력물로 사용하지 않으므로 타임라인 방식의 인터페이스가 없고, 텍스트와 이미지를 입력하는 공간이 메인으로 설정되어 있다. 이는 인공지능이 메타미디엄으로써 초래한 변화라고 볼 수 있다.

대사 생성형은 입력한 텍스트를 대사(나레이션)로 보고 영상을 생성한다. 기본적으로 입력한 텍스트를 영상의 자막으로 변환하여 자막 기반의 씬(Scene)을 생성한 뒤, 해당 씬에 알맞은 이미지를 생성하거나, 데이터베이스에 있는 영상 클립을 조합하여 시퀀스(Sequence) 형태의 영상을 생성한다. 생성한 영상은 자막에 있는 정보를 반영한다는 특징이 있다. 대사 생성형은 영상 자체를 생성하지 않고 이미지에 효과를 주거나 이미 있는 클립을 조합하는 형태로, 지문 생성형보다 더 빠른 시간에 긴 길이의 영상을 출력할 수 있다. 출력된 영상에는 자막과 사운드, 이미지, 나레이션(더빙), 배경음악 등이 포함된다. 소프트웨어마다 영상 소스와 영상의 조합 방식이 조금씩 다르다는 특징이 있으며 대부분의 소프트웨어가 영상 편집 기능을 동반한다. Vrew, InVideo, Wondershare Filmora, Pictory 등 다수의 제품이 출시되어 있다. 본 논문에서는 대표적으로 Vrew와 InVideo를 사례로 든다.

국내 스타트업 보이저엑스에서 출시한 Vrew는 텍스트의 정보를 바탕으로 생성한 AI 이미지에 모션그래픽 형태의 움직임을 주고, 미리 활용되어 저장된 저작권 프리 영상 클립과 배경음악, 인공지능 성우의 목소리를 조합하여 영상을 제작한다. InVideo의 경우, AI 이미지보다는 고품질의 영상 클립 소스와 실제 촬영된 이미지, 배경음악, 그래픽 효과를 중심으로 영상 출력물을 만든다. 빠른 시간에 캐주얼한 정보전달 영상을 만들고자 하는 경우는 브루가 유용하게 쓰였고, 시간이 더 걸려도 역동적이고 감각적인 영상을 만들고자 한다면 템플릿과 그래픽적 요소가 많은 InVideo가 적합했다. 눈에 띄는 점은 그림 6처럼 두 저작도구 모두 텍스트 to 영상 기능을 활용할 때 편집 인터페이스가 타임라인 방식이 아닌 텍스트 중심의 편집 방식을 취한다는 것이다. Vrew의 경우는 텍스트를 행 단위로 편집하는 반면, InVideo는 씬 단위의 편집 방식을

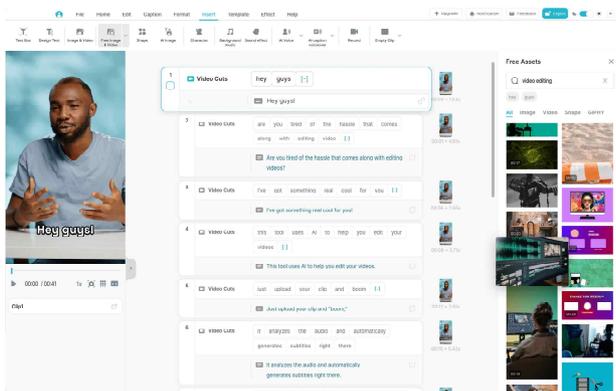


그림 6. Vrew의 텍스트 기반 영상편집 화면[16]
Fig. 6. Vrew's text-based video editing screen[16]

취했다. 이는 유튜브 등 스트리밍 플랫폼의 대중화로 영상의 정보전달 기능이 커졌다는 점, 자연어처리 AI가 발전하며 Input, Output으로 텍스트를 널리 활용하는 점에서 생긴 변화라고 추측하며 AI가 메타미디엄으로써 초래한 변화라고 볼 수 있다.

캐릭터 생성형 저작도구는 입력한 텍스트를 가상 캐릭터가 말하는 형태의 영상으로 생성한다. 대사를 기반으로 영상이 만들어지는 것은 대사 생성형과 유사하나, 캐릭터 생성형은 대사 생성형 영상처럼 다양한 컷을 만드는 것이 아니다. 그림 7처럼 사람 형태로 만들어진 3D 모델이 입력한 텍스트를 정확한 발음으로 읽어주는 것이 핵심이다. 발음에 따라 입 모양 또한 움직이므로 잘 만들어진 모델은 실제 사람이 말하는 것처럼 보인다. 이에 해당하는 저작도구는 D-ID와 DeepBrain AI, synthesia, elai 등이 있다. 아나운서가 정보를 알려주는 형태의 영상이나, 제작자의 신상이 드러나지 않게 가상의 캐릭터로 영상을 제작하고 싶은 경우, 교육용 동영상 등에 유용하게 사용할 수 있다. AI가 3D 캐릭터와 결합한 캐릭터 생성형 저작도구는 혼종적 미디어라고 볼 수 있는 버추얼 휴먼 콘텐츠의 창작을 활성화할 수 있다.

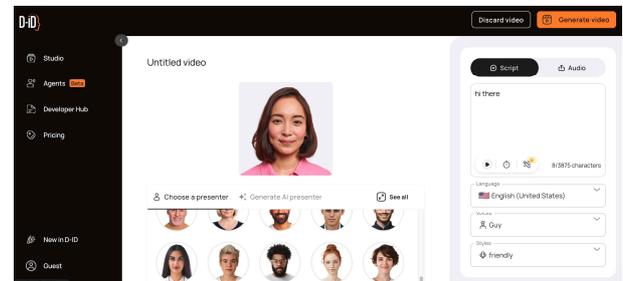


그림 7. D-ID의 저작도구 스크린 캡처
Fig. 7. Screen capture of D-ID's authoring tool

2) 이미지 to 영상

이미지를 영상으로 만들어주는 AI는 입력한 이미지의 내부적 움직임을 표현하는 ①애니메이션 생성형과, 이미지 외부에 모션이나 템플릿 등 디자인 요소를 추가하여 그래픽 효과를 주는 ②모션그래픽 생성형이 있다. 애니메이션과 모션그래픽의 개념은 상당히 유사하므로 혼동의 우려가 있다. 본 논문에서는 입력한 요소와 출력한 요소의 그래픽요소를 기준으로 구분한다. 즉, 그림 8처럼 출력물에 입력한 이미지 이외에 추

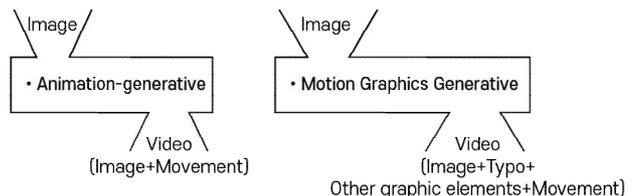


그림 8. 출력물에 따른 애니메이션, 모션그래픽 생성형 구분
Fig. 8. Differentiate between animation and motion graphics based on output

가된 그래픽적 요소가 없을 경우를 애니메이션 생성형이라고 보았고, 출력물에 입력한 이미지 이외에 타이포그래피, 아이콘 등의 그래픽적 요소가 추가되었을 경우 모션그래픽 생성형이라 정의했다.

애니메이션 생성형 AI는 2차원 이미지 내부에 있는 요소를 부분적으로 움직이거나, 3차원의 깊이감을 주거나, 카메라 앵글의 효과 등을 통해 애니메이션 영상을 만들어준다. 이에 해당하는 소프트웨어는 Runway Gen-2, LeiaPix 등이 있다. 아직 움직임이 상업적으로 활용할 정도로 자연스럽지는 않으나, 정지된 이미지를 영상처럼 만들 수 있다는 점에서 활용이 기대된다. 동일한 작업을 어도비 애프터 이펙트에서 수행하려면 굉장히 오랜 시간이 걸리는 것에 비해 애니메이션에 관한 지식 없이 손쉽게 이미지에 움직임을 줄 수 있다는 점이 제작 효율성을 높인다. 애니메이션 생성형 저작도구는 애니메이션의 강도와 움직임을 조절할 수 있는 기능은 있으나, 영상을 자르고 편집하는 기능이 없다. 향후, 자르고 편집할 수 있는 기능이 생긴다면 단순히 이미지가 움직임을 갖는 것을 넘어 이미지 수집만으로도 하나의 영상 콘텐츠가 제작될 수 있을 것이다.

모션그래픽 생성형 소프트웨어는 이미지 내부에 움직임을 주지 않고 이미지 전체가 동일한 좌표로 움직이거나 분할된다. 여기에 이미지 외부에 시스템 내부에 저장되어 있던 그래픽 심볼이나 템플릿을 활용하여 역동적인 영상을 만들어준다. 대표적으로 InVideo, Vita, CapCut 등이 있다. 모션그래픽 생성형은 애니메이션 생성형과 달리 독립적인 저작도구가 아니라, 종합적으로 영상편집을 할 수 있는 프로그램에 추가되어 있다는 특징이 있다. 이러한 프로그램을 사용하면 제작 과정이 까다로운 편이라 전문가의 영역이었던 모션그래픽도 손쉽게 활용할 수 있으므로, 모션그래픽의 대중화를 촉진할 것이다.

3) 사운드 to 영상

사운드 to 영상과 관련된 인공지능은 음성을 분석하여 텍스트로 변환 후 영상으로 만드는 ①음성 인식형과, 음성 이외의 소리를 인식하여 영상으로 만드는 ②소리 인식형으로 나눌 수 있다. 음성 인식형은 영상과 관련된 AI 중 비교적 오랜 역사를 가지고 있는 반면에, 소리 인식형은 이미지나 텍스트 등을 활용한 다른 AI 저작도구에 비해 나중에 개발되었다는 특징이 있다.

음성 인식형 AI는 음성을 분석하여 자막을 생성하고 자막 내용에 맞는 영상을 생성한다. 이에 해당하는 저작도구는 Vrew가 있다. 이외에도 Adobe Premiere Pro, CapCut, Vita도 목소리를 분석하여 자동으로 자막을 생성하는 기능이 있다. 전통적인 영상편집 프로그램에 추가되는 대표적인 인공지능 기능 중 하나이다.

소리 인식형 AI는 2024년 3월 시점으로 아직 초기 개발 단계에 해당하여 출시된 상품은 없지만, 포스텍 김성빈 등 4명의 연구진이 개발한 Sound2Scene Model이 개발되어 있

다[17]. Sound2Scene Model은 입력된 오디오를 시각적으로 변환한 다음, 사전 학습된 생성기를 활용하여 영상을 만든다. 다만 아직 상업적으로 활용하기에는 생성된 영상 출력물의 품질이 떨어진다는 한계가 있으나[18], 발전 가능성이 큰 분야다. 사운드 to 영상의 경우, 기존 영상 제작 과정에서 영상편집 후에 적절한 사운드를 추가하는 방식과는 정반대 순서의 제작 방식이다. 따라서 이 유형에서만 볼 수 있는 독특한 형태의 영상이 제작되거나, 음악을 시각적으로 표현할 수 있는 혼종화된 미디어가 나타날 가능성이 있다.

3-2 편집 단계: 변형 AI와 편집 AI

편집 단계에 사용되는 AI는 변형 AI와 편집 AI로 정의한다. 이 단계에서의 입력물은 영상이며 출력물도 영상이다. 변형 AI는 입력한 영상에 적용된 기능에 따라 ①스타일 전이와 ②얼굴 합성, ③조절 및 보정으로 나누며, 편집 AI는 ①제거, ②추가, ③조합으로 분류할 수 있다. 변형 AI와 편집 AI는 기존 영상편집에서 반복적으로 자주 쓰이는 작업의 효율성을 향상하는 방향으로 개발되어 영상 제작의 자동화를 촉진한다.

1) 변형 AI

변형 AI는 입력된 영상 소스를 AI로 인식하여 스타일이나, 효과 등에 변형을 준다. 편집 AI가 입력한 영상 소스를 삭제, 추가하여 영상 길이가 줄어들거나 늘어나는 것과 달리, 변형 AI는 입력한 영상 길이와 출력한 영상 길이가 동일하다는 특징이 있다. 변형 AI는 기능에 따라 ①스타일 전이, ②얼굴 합성, ③조절 및 보정으로 나눌 수 있다.

스타일 전이는 입력한 영상의 스타일을 유명 화가의 작품이나 원하는 스타일로 바꿔주는 AI라고 할 수 있다. 일반 실사 영상을 애니메이션이나, 수채화 등 다양한 형태로 변형할 수 있어, 신기함을 강조하는 쇼츠 영상에 많이 활용된다. 대표적인 소프트웨어로 EbSynth, Kaiber가 있다.

얼굴 합성은 영상에 나오는 얼굴을 다른 사람 몸에 합성하거나, 표정 등을 바꿀 수 있다. 딥페이크(Deepfake)라고도 알려져 있으며, 영상뿐 아니라, 오디오, 사진, 문서 등 다양한 산업 현장에서 널리 활용되고 있다[19]. DeepFaceLab, Faceswap, Deepswap, Facemagic 등의 소프트웨어가 이에 해당한다. 얼굴 합성 AI는 저렴한 비용으로 특수효과 재현이 가능한지만, 무분별한 남용 또는 악의적 목적으로 제작된 콘텐츠는 영상의 진위를 알 수 없어 사회적으로 문제가 될 수 있으므로, 활용 시 법적 검토와 더불어 향후 관련 제도 정비 등이 필요하다.

조절 및 보정은 변형되는 범위가 스타일 전이와 얼굴 합성에 비해 미미하다. 따라서 영상 가공 시 후반작업에 많이 활용된다. 기능적 유사성 때문에 편집 AI와 혼동할 수 있지만 편집 AI와 달리, 기능 적용 이후에도 입력물의 영상 길이에 변함이 없다는 특징이 있다. 조절 기능에는 영상 리프레임, 사운드 리믹스, 세부 색상 조절, 자동 더킹(대화화 배경 오디오

가 잘 어우러지게 하는 기능) 등이 있다. 보정 기능으로는 인물의 체형 보정, 자동 색상 일치, 색감 보정, 음성 강화 등의 기능이 있다. Adobe Premiere Pro, CapCut 등 전통적인 영상 저작도구에 지속적으로 추가되는 AI 기능이다.

2) 편집 AI

편집 AI는 자동화라는 키워드 안에서 ①제거, ②추가, ③조절 기능으로 분류할 수 있다. 영상의 스타일 보다는 작업의 효율성을 높이고자 하는 목적이 크다.

제거는 영상에 필요 없는 부분을 자동으로 제거하는 것이다. Adobe Premiere Pro에서는 컷 편집에서 대화가 끊긴 부분의 공백을 제거하고, “어”와 “음” 같은 추임새의 단어를 감지하여 삭제하는 기능이 있다. CapCut에서는 불필요한 영상 배경을 자동으로 제거해주시기도 한다.

추가는 그래픽요소나 템플릿을 활용한 자동 편집이 대표적이다. 이미 만들어진 템플릿에 입력한 소스 영상을 넣어서 빠른 시간에 역동적인 영상을 만든다. 모션그래픽 생성형과 유사하나, 모션그래픽 생성형은 인풋이 jpg, png 등의 이미지이고, 편집 AI에서 추가 기능은 인풋이 avi, mp4 등의 동영상이다. 추가 기능이 있는 소프트웨어는 Capcut, Vita 등이 있다. 영상 소스를 모션그래픽 템플릿에 적용하면 손쉽게 자동으로

편집된 영상이 나오기 때문에, 영상 제작 초보자 들이 모바일 어플리케이션에서 자주 활용한다.

조합은 입력한 영상 소스에서 불필요한 컷과 필요한 컷을 구분하여 추출, 조합하여 장면을 자동으로 편집한다. 장면 편집, 자동 편집이라고도 하며 Adobe Premiere Pro의 플러그인 AutoPod, Adobe Premiere Pro의 장면 편집 감지 기능이 이에 해당한다.

편집 AI에 해당하는 제거, 추가, 조절 기능 모두 영상편집 시, 반복되는 작업을 인공지능이 자동으로 처리함으로써 후가공 작업에 들어가는 시간을 줄여준다는 특징이 있다. 이 같은 자동화는 영상 제작 인력의 구성을 단순화할 수 있으며, 나아가 영상 제작 산업 구조를 간소화할 수 있을 것이다.

본 연구에서 제시한 개념을 유형화하여 AI 기반 저작도구를 분석한 결과는 표 1과 같다. 2024년 3월 기준 가장 주목 받는 분야는 형상화 AI였다. 특히 텍스트 to 영상 분류에 해당하는 소프트웨어의 출시가 이슈였다. 이는 인공지능과 영상 제작이 결합한 독특한 특징으로 향후 AI 기반 영상 저작도구가 보편화된다면, 기존 제작 방식과는 다른 발상의 영상이 늘어날 것이며 효율성 향상은 물론 일반인도 중간 이상의 품질의 영상 제작이 가능해지므로 표현의 기회가 확장될 것이다. 표현의 기회 확장은 창작자의 진입장벽 해소뿐만 아니라, 실

표 1. 영상 제작 단계에 따른 AI 영상 저작도구 유형 및 특징

Table. 1. Types and features of AI video authoring tools at different stages of video production

Production (Shooting) *Non video input				
Large	Medium	Small	Features	Programs
Shaping AI	Text to Video	Action line Generative	Output only the video itself that the text describes	RunwayGen-2, Moonvalley, Pika, Genmo, Sora
		Dialogue Generative	Converts the text you enter into subtitles and outputs a video with a combination of text, sound, images, and more	Vrew, InVideo, Filmora, Pictory
		Character Generative	Output a video of a fictional character speaking your typed text	D-ID, DeepBrain, synthesia, elai
	Image to Video	Animation Generative	Give movement to the inside of an entered image	RunwayGen-2, LeiaPix
		Motion Graphics Generative	Add motion + graphic elements outside the image you entered	InVideo, Vita, Capcut
	Sound to Video	Voice-activated	Analyze speech to generate subtitles and output video that matches the subtitle content	Vrew, CapCut, Vita
Sound-activated		Transform non-speech sounds into visuals to create video	Sound2Scene Model	
Editing (Processing) *Video input				
Large	Medium		Features	Programs
Transform AI		Style transition	Transforms the style of the footage you enter into the style of a famous painter or your favorite style	Ebsynth, Kaiber
	*Same input/output video length	Face synthesis	Compositing a face onto someone else's body or altering facial expressions(Video authenticity is difficult to determine)	DeepFaceLab, Faceswap, Deepswap, Facemagic
		Adjustment and Retouching	Adjustments: video reframing, sound remixing. detailed color adjustments, auto ducking Retouching: body shape correction, auto color matching, color tone correction, voice enhancement, etc.Added to traditional video editing software	Adobe Premiere Pro, CapCut
Editing AI		Removal	Blank spaces, deleting after detecting a filler word, automatically removing video backgrounds, and more	Adobe Premiere Pro, CapCut
	*Different input/output video length	Addition	Automated editing with graphic elements or templates. Create dynamic videos in less time	CapCut, Vita
		Combination	Extracting and combining the necessary and unnecessary cuts from a video source. Also known as scene editing or auto-editing.	Adobe Premiere Pro, AutoPod(Plug-in)

제 영상 산업 현장에서도 활용할 수 있다. 이미 영화제작 시 AI 기능을 활용한 사례가 있으며[20], 사전시각화(Pre-visualization) 단계에도 활용할 수 있다. 예시를 들면 주로 이미지로 작업 되었던 콘티나 스토리보드 대신 AI 저작도구를 활용한 프로토타입의 영상으로 썸 구성을 제작하여 실제 촬영 시 제작진 간의 소통을 더 원활하게 할 수 있을 것이다. 또한, 기존의 전통적 영상 저작도구에는 작업의 효율성을 높일 수 있도록 변형 AI의 조절 및 보정 분류와 편집 AI의 전 분류에 해당하는 AI 기능이 꾸준히 추가되고 있다. 이는 영상 제작 인력의 간소화와 같은 변화를 야기할 수 있으며, 주로 팀으로 제작되는 영상 제작사의 구조가 개인화될 수 있음을 시사한다.

나아가 AI 영상 콘텐츠 저작도구의 보편화가 지속되면, 적은 비용으로도 영상과 다른 기술과의 융합 및 연계를 촉진하여, 증강현실, 인터랙티브 스토리텔링 등 기존 영상과는 다른 새로운 영상 경험을 만드는 데 유용하게 활용할 수 있을 것이다.

AI가 메타미디어으로써 가져올 변화는 편집 단계보다는 제작 단계의 형상화 AI에서 시작될 가능성이 높다. 형상화 AI는 기존 영상 제작 방식과는 전혀 다른 방법으로 영상을 제작할 수 있게 개발된 반면에, 편집 단계에 해당하는 저작도구는 기존에 없던 새로운 형태의 변화를 가져오기보다는 기존 방식으로 수고롭게 작업했던 양식의 콘텐츠를 손쉽게 만들 수 있게 제작되었기 때문이다. 이는 AI 영상 콘텐츠 저작도구가 기존 제작 방식 대비 어떠한 효율을 주는지, 효과, 시간, 비용 측면에서 정리한 표 2에서도 드러난다.

편집(가공) 단계의 변형 AI와 편집 AI가 일부 인건비와 후가공 비용을 절감할 수 있지만, 영상 제작 단계의 형상화 AI

는 획기적으로 영상 제작 단계에서 소모되는 시간과 비용을 절약할 수 있다. 또한 형상화 AI의 대량생성형 저작도구는 제작 단계뿐만 아니라 기획 및 편집 단계 역할까지 수행할 수 있는 AI 기능을 갖추고 있다. 이는 주로 편집과 후가공에서 활용되던 영상 저작도구의 쓰임새가 영상 제작 파이프라인의 전 단계로 확대되었다는 것을 의미한다. 주요 저작도구의 성능을 비교하면 표 3과 같다.

성능이 공개된 AI 영상 저작도구의 기능 중 최대 영상 길이, 영상 연장 여부, 카메라 컨트롤, 모션 컨트롤로 나누어 성능을 비교한 결과 텍스트 to 영상에 해당하는 Runway Gen-2와 Pika 1.0의 경우 생성되는 영상의 길이는 초 단위로 짧았다. 하지만 영상을 연장하고 카메라와 모션 컨트롤이 가능했다. 반면에 Deepbrain AI, Pictory 등 생성되는 영상의 길이가 분 단위로 긴 저작도구는 영상을 연장하거나 카메라, 모션의 컨트롤이 불가능하다는 특징이 나타났다[15].

표 3. 주요 AI 영상 저작도구의 성능 비교

Table 3. Compare the performance of leading AI video authoring tools

Tools	Maximum video length	Extend video	Camera controls	Motion controls
Runway Gen-2	18sec	O	O	O
Pika 1.0	15sec	O	O	O
Genmo ai	6sec	X	O	O
Synthesia	10min	X	X	O
Deepbrain AI	20min	X	X	X
InVideo	40min	X	O	X
Pictory	60min	X	X	X

표 2. 영상 제작 단계별 AI 영상 저작도구의 효용 분석

Table 2. Analyzing the effectiveness of AI video authoring tools at each stage of video production

Video Creation Steps	Categorize	Utility analysis	
Production (Shooting)	Shaping AI	Effects	<ul style="list-style-type: none"> - Enables people with no video production literacy to create videos by simply entering text or images. Prototype videos can be created for high-cost content production, and can be used in the pre-visualization phase of production. such as contingency or storyboarding. - Create new types of videos that are different from traditional production methods depending on the input. - Can be used as a next-generation authoring tool with automated video content production. - Dialog generation tools for text-to-video categorization can also generate scripts and edit videos, enabling tasks not only in the production phase but also in the planning and editing phases - Increasing importance of text in stylized video editing (text-centric interfaces are emerging)
		Time	Dramatically reduced video production time
		Cost	Reduce production costs for content, storyboarding, set construction, original artwork, special effects, and more
Editing (Processing)	Transform AI	Effects	<ul style="list-style-type: none"> - Automate repetitive effects on video, such as body shape correction - Can be used for shorts, commercials, and more, with the ability to change the look and feel of live-action videos
		Time	The input video has to be created separately, saving only post-production time
		Cost	Reduce special effects and post-production costs
	Editing AI	Effects	<ul style="list-style-type: none"> - Automate simple repetitive tasks like removing white space and editing cuts - Automate motion graphics with templates
		Time	Can save some video editing time
		Cost	More footage can be edited with the same number of people, reducing editing costs

이 같은 저작도구의 변화 양상은 기호학의 통합체 (syntagma), 계열체(paradigm) 모델을 도입하여 분석할 수 있다. 에드거와 세즈윅은 기호학을 광범위한 인간 행위와 생산물의 유의미한 관계를 지각하고 분석할 수 있는 힘을 지니고 있어, 문화연구 영역에서 중요하게 이용할 수 있는 이론적 도구들의 모음이라고 했다[21]. 이런 점에서 AI라는 기호가 영상 저작도구를 통해 어떻게 콘텐츠를 생성하는지 분석하는 기호학적 접근은 적절한 분석 틀이라고 할 수 있다. 통합체는 기호의 배열이고 계열체는 기호의 집합이라고 할 수 있다. 이 개념은 구조주의 언어학자 페르디낭 드 소쉬르의 언어학에서 출발했다[22]. 에드거와 세즈윅에 따르면 ‘계열체’는 메시지가 구성되는 의미 있는 단위를 말하는데, 이러한 계열체로부터 하나의 의미 있는 전체를 구성하는 기호들의 조합을 ‘통합체’라고 정의했다[21]. 버거에 따르면 모든 텍스트는 사건이 전개되는 순서에 의해 통합체 구조를 갖고, 텍스트 속에서 발견되는 숨겨진 대칭에 의해 계열체 구조를 가지며 이 두 가지 방식으로 의미를 생성한다[23]. AI 영상 저작도구 유형 역시 통합체 구조와 계열체 구조로 분석할 수 있다. 이를 그래픽으로 표현하면 그림 9와 같다.

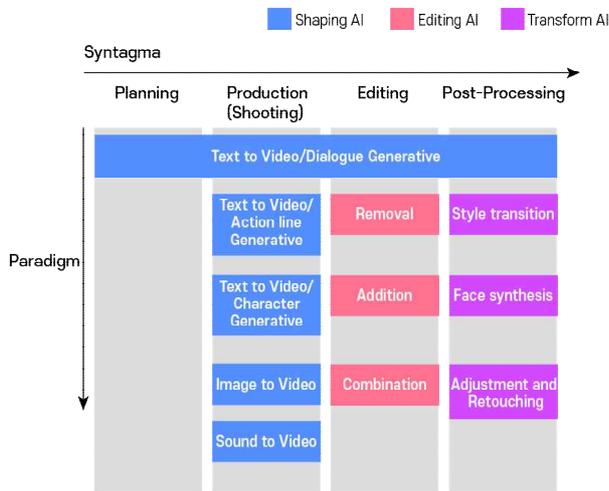


그림 9. 통합체, 계열체 구조에 따른 AI 영상 저작도구 유형
 Fig. 9. Types of AI video authoring tools based on syntagma and paradigm structure

영상 제작 단계를 기획, 제작(촬영), 편집, 후가공의 4단계로 분류하고 작업 순서대로 수평적으로 나열하여 그림 9처럼 통합체 구조를 정의했다. 계열체는 각각의 단계에서 AI 유형에 따라 수직적으로 구분하였다. 전통적으로 편집과 후가공 위주로 기능을 수행하던 영상 저작도구가 형상화 AI로 인해 제작과 기획 단계까지도 기능을 확장한 모습이 드러난다. 따라서 역사적으로 전례가 없는 변화를 일으키는 그 무엇인 메타미디어의 정의에 가까운 것은 형상화 AI 중심의 영상 저작도구라고 볼 수 있다.

본 연구에서는 인공지능이 영상 저작도구에 가져온 변화를 메타미디어 측면에서 바라보았다. 영상 제작 단계에 따른 입

출력 개념을 적용하여 2024년 3월까지 공개된 AI 기반 영상 저작도구의 유형화를 시도하고 동향을 분석했다. 그 결과 AI 영상 저작도구는 크게 통합체 구조와 계열체 구조로 진화하고 있었다. 통합체 측면에서의 변화는 영상 제작 효율 상승뿐만 아니라, 영상 창작 주체의 확장과 기존 영상 콘텐츠 제작 방식의 다양화를 촉진한다. 계열체 측면에서의 변화는 영상 제작 시간과 비용을 절감하여 영상 제작 효율의 상승으로 이루어진다. 이 같은 변화는 지금까지 통용되었던 영상 제작 리더러시의 변화가 이루어질 수 있음을 시사하며, 향후 AI를 활용한 새로운 방식의 영상 콘텐츠 장르 등장과 소비 시장이 나타날 수 있음을 암시한다.

IV. 결 론

본 연구는 AI 기반 영상 콘텐츠 저작도구를 영상 제작 단계 및 입력(input)과 출력(output) 개념을 통해 유형화하여 동향을 파악하였다. 이를 통해 인공지능이 메타미디어으로써 영상 저작도구에 어떤 변화를 가져왔는지 통합적으로 바라볼 수 있는 관점을 제시하였다. 더불어 영상 제작 단계에 따라 형상화 AI, 변형 AI, 편집 AI로 영상 관련 인공지능을 분류한 뒤, 기호학적으로 분석함으로써 각각의 소프트웨어를 연계적으로 파악할 수 있는 기준을 제시한 점에서 학술적 의미를 가진다. 본 논문에서 분석한 저작도구의 유형화는 AI 영상 저작도구뿐만 아니라, 향후 더욱 복잡해지는 인공지능 관련 소프트웨어를 분류하고 개념화하는 데 유용하게 활용될 것이다.

AI 기반 영상 콘텐츠 저작도구는 영상 제작(촬영)과 편집(가공) 단계에 속하는 다양한 AI를 활용한다. 그중에서도 텍스트 to 영상의 형상화 AI 저작도구가 통합체 측면에서 전례 없는 변화를 주도한다. Adobe Premiere Pro 같은 전통적 방식의 영상편집 소프트웨어는 계열체 측면에서 작업 효율을 높일 수 있는 편집 AI 기능을 꾸준히 추가하고 있다. 이는 영상 제작의 편의성을 높일 뿐만 아니라 제작, 편집의 자동화로 이어지고 있다. 영상 저작도구에 AI 기능이 추가되어 복잡해질수록 더 좋은 성능의 하드웨어가 필요하다. 따라서 모바일 시대에 PC용 저작도구가 모바일 버전으로 출시되었던 것과 달리, CapCut, Vrew의 사례처럼 AI 시대에는 모바일 어플리케이션으로 출시되었던 저작도구가 다시 PC 버전으로 출시되는 양상을 보인다.

전통적 방식의 영상 콘텐츠 저작도구 유저 인터페이스(UI)에도 의미 있는 변화가 보였다. Adobe Premiere Pro, Vrew, Invideo 등에서 보이는 변화로, 아날로그 필름에서 따온 편집 인터페이스였던 타임라인 방식 이외에, 텍스트를 기준으로 영상을 편집하는 인터페이스가 새로 나타났다. 이는 영상 기획에서 텍스트 결과물을 사용하는 AI가 늘어났고 예전보다 영상의 텍스트 정보가 중요해졌음을 시사한다. 더불어 인공지능 시대의 영상편집 인터페이스 변화에 관한 후속 연구 또한 필요하다.

본 논문에서는 영상이 출력물로 나오는 경우를 한정하여 AI 기반 영상 저작도구의 동향을 살펴보았다. 하지만 Vrew 등 이미 여러 저작도구에 영상 기획 단계의 텍스트 출력물을 만들 수 있는 기능이 존재한다. 이 같은 변화를 통합체 구조의 진화라고 보았다. 통합체 구조의 저작도구를 통해 창작자가 영상 리터러시 없이 아이디어나 키워드만 있어도 AI 영상 제작이 가능해졌다. 영상 콘텐츠의 대중성 측면에서 볼 때, 이는 영상 콘텐츠 저작도구가 워드프로세서처럼 차세대 저작도구로 활용될 가능성이 있다는 것을 의미한다. 더불어 기존에 통용되던 영상 제작 리터러시에 관한 재정의가 이루어질 수 있음을 시사한다. AI 영상 저작도구를 통한 영상 제작 자동화는 창작자 기반을 넓히고 표현의 기회를 확대하여 독특한 발상의 영상 제작이 가능해지므로, AI를 활용한 새로운 장르의 영상 콘텐츠가 등장할 가능성이 있다. 또한 영상 제작 인력의 간소화를 촉진하여 영상 제작 산업의 구조를 단순화할 수 있으며, AI 기술을 활용하여 타 분야와 영상 콘텐츠의 융합을 시도하여 기존에 없던 새로운 영상 경험을 창조하기에 유용할 것이다.

마지막으로 본 논문이 작성된 시점에서의 AI 영상은 인간이 정교하게 제작한 영상보다는 품질이 떨어졌다. 특히 영상 스토리의 맥락상 공감과 재미를 유발하는 영역을 인공지능이 구현하기에는 부족한 점이 많았다. 개발 측면에서 향후 출시되는 AI 영상 저작도구는 이를 보완할 알고리즘 개발을 연구해야 할 것이다. 더불어 본 논문은 2024년 3월을 기준으로 작성되었다. 따라서 향후 각 저작도구 별로 활용된 생성형 AI 기술에 대한 설명이 보완될 수 있으며 기능이 추가될 수 있다. 다양한 기능의 AI 저작도구가 출시되고 있는 시점에서, 영상 창작자는 영상 제작 단계별로 효용성이 뛰어난 AI 영상 저작도구를 인지하고 작업 효율성을 높이는 동시에, 인공지능으로 대량 생산되는 영상 콘텐츠보다 경쟁력 있는 다른 형태의 콘텐츠를 만들기 위한 깊은 고민이 요구된다. 컴퓨터와 모바일이 메타미디어으로써 큰 변화를 이끌고 새로운 산업을 만들었듯이, 인공지능도 다양한 분야에서 큰 변화를 이끌고 새로운 산업을 만들어간다. 다만 산업적으로 확대되고 성장하려면, 창작자 입장에서의 AI 저작도구 활용 이외에도 소비자 입장에서 AI를 활용한 창작물에 대해 논의할 필요가 있다. 컴퓨터와 모바일은 소비자 입장에서 편의를 제공하였기에 비용을 지불할 가치가 있었으나, 소비자가 AI가 만든 콘텐츠에 비용을 지불할 수 있을지는 아직 미지수다. 이 같은 불확실성을 해소하기 위해서는 AI 창작물 인식에 관한 연구가 필요하며, AI 창작과 상업성에 관한 깊은 논의가 수반되어야 한다.

감사의 글

이 논문은 한양대학교 교내연구지원 사업으로 연구됨 (HY-2023-3446).

참고문헌

- [1] L. Manovich, *Software Takes Command*, J. H. Lee, trans. Seoul: Communication Books, 2014.
- [2] A. Kay and A. Goldberg, "Personal Dynamic Media," *Computer*, Vol. 10, No. 3, pp. 31-41, March 1977. <https://doi.org/10.1109/C-M.1977.217672>
- [3] S. T. Kwon, "Visual Identity and Narrative Identity in a Video Editing Software: The Semiotic Analysis of the Final Cut Pro X," *Visual Culture*, No. 32, pp. 137-159, June 2018.
- [4] D. Jung and S. Kim, "The Development Direction and Analysis on Graphic Authoring Tool according to Design Work Frame -Focused on Interface Design-," *Journal of the Korea Convergence Society*, Vol. 8, No. 11, pp. 233-240, November 2017. <https://doi.org/10.15207/JKCS.2017.8.11.233>
- [5] MADTimes. TikTok Will Be the Leading Entertainment Genre App in 2022, with More than Twice as Many Downloads as YouTube over the Last Decade [Internet]. Available: <https://www.madtimes.org/news/articleView.htm?idxno=16384>.
- [6] S. Park, "A study on the Another Competition Induced by Changes in User Usage Behavior due to the Digitalization of Media: Focusing on Niche Analysis of Naver, Google, and YouTube as Information Search Platforms," *Korean Journal of Broadcasting & Telecommunications Research*, No. 38, pp. 38-74, April 2022. <http://doi.org/10.22876/kjbr.2022..118.002>
- [7] B. Jeon, "Artificial Intelligence and Film & Video Production : Through the Cases of the Use of Artificial Intelligence as a Film & Video Production Tool," *PREVIEW: The Korean Journal of Digital Moving Image*, Vol. 20, No. 1, pp. 133-162, April 2023. <https://doi.org/10.23120/kadmi.2023.20.1.006>
- [8] S. Y. Kim and H. T. Kim, "Artificial intelligence(AI) Composition Technology Trends & Creation Platform," *Culture and Convergence*, Vol. 44, No. 6, pp. 207-228, June 2022. <https://doi.org/10.33645/cnc.2022.6.44.6.207>
- [9] H. B. McLuhan, *Understanding Media: The Extensions of Man*, S. H. Kim, trans. Seoul: Communication Books, 2011.
- [10] J.-W. Jeong, "Development and Perspective of the Image Editing System," *Journal of East-West Communication Council*, Vol. 9, pp. 75-107, December 2005.
- [11] S. Y. Ra and S. K. Hong, "A Study on the Efficiency of Digital Video Contents Integrated Management Program," *Korea Science & Art Forum*, Vol. 12, pp. 53-64, April 2013. <https://doi.org/10.17548/ksaf.2013.04.12.53>
- [12] K.-B. Kim, "A Case Study on Artificial Intelligence Based

- Multimedia Video Production,” *Journal of Digital Art Engineering & Multimedia*, Vol. 8, No. 4, pp. 463-472, December 2021. <https://doi.org/10.29056/jdaem.2021.12.10>
- [13] S. H. Jang, *TV News Video Editing Theory*, Seoul: Chasong, p. 15, 2001.
- [14] OpenAI. Sora Overview [Internet]. Available: <https://openai.com/index/video-generation-models-as-world-simulators/>.
- [15] Samsung Securities Research Center. Generative AI, the Start of a Video Revolution [Internet]. Available: https://www.samsungpop.com/common.do?cmd=down&contentType=application/pdf&inlineYn=Y&saveKey=research.pdf&fileName=3020/2024032513551696K_02_04.pdf.
- [16] Vrew. Official Website [Internet]. Available: <https://vrew.voyagerx.com/ko/>.
- [17] Sound2Scene. Official Website [Internet]. Available: <https://sound2scene.github.io/>.
- [18] S.-B. Kim, A. Senocak, H. Ha, A. Owens, and T.-H. Oh, “Sound to Visual Scene Generation by Audio-to-Visual Latent Alignment,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, pp. 6430-6440, June 2023. <https://doi.org/10.1109/CVPR52729.2023.00622>
- [19] Korea Communications Agency. The Light and Shadow of Deepfake Technology Korea Broadcasting and Telecommunications Organization [Internet]. Available: <https://welcon.kocca.kr/ko/info/trend/1939513>.
- [20] 2eConsulting. Deepfake Isn't All Bad! Turtleman, Kim Kwang-Seok, Yoo Gwan-Soon, Na Yeon Yi, and Others Use Deepfake Technology for Good [Internet]. Available: <https://www.2e.co.kr/news/articleView.html?idxno=301831>.
- [21] A. Edgar and P. Sedgwick, *Key Concepts in Cultural Theory*, M. J. Park, Y. W. Lee, C. N. Kim, J. Y. Jung, B. W. Sohn, E. W. Joo, ... and Y. R. Kim, trans. Seoul: Hannarae, 2003.
- [22] F. de Saussure, *Curso de Linguística*, H. K. Kim, trans. Seoul: Greenbee, 2022.
- [23] A. A. Berger, *Media and Communication Research Methods: An Introduction to Qualitative and Quantitative Approaches*, C. Y. Ryu, D. H. Kim, and E. M. Kim, trans. Seoul: Communication Books, 2001.



이세희(SeHui Yi)

2009년 : 국민대학교
시각디자인학과(미술학사)
2014년 : KAIST 문화기술대학원
(공학석사)

2023년~현 재: 한양대학교 문화콘텐츠학과 박사과정
※ 관심분야: 인공지능 창작, 생성 AI, 캐릭터,
디지털콘텐츠 등



이진(Jin Lee)

2006년 : 이화여자대학교
국어국문학과(문학학사)
2009년 : 이화여자대학교
디지털미디어학부(문학석사)
2018년 : 이화여자대학교
융합콘텐츠학과(문학박사)

2020년~현 재: 한양대학교 ERICA 문화콘텐츠학과 조교수
※ 관심분야: 게임 스토리텔링, 트랜스미디어, 메타버스 등