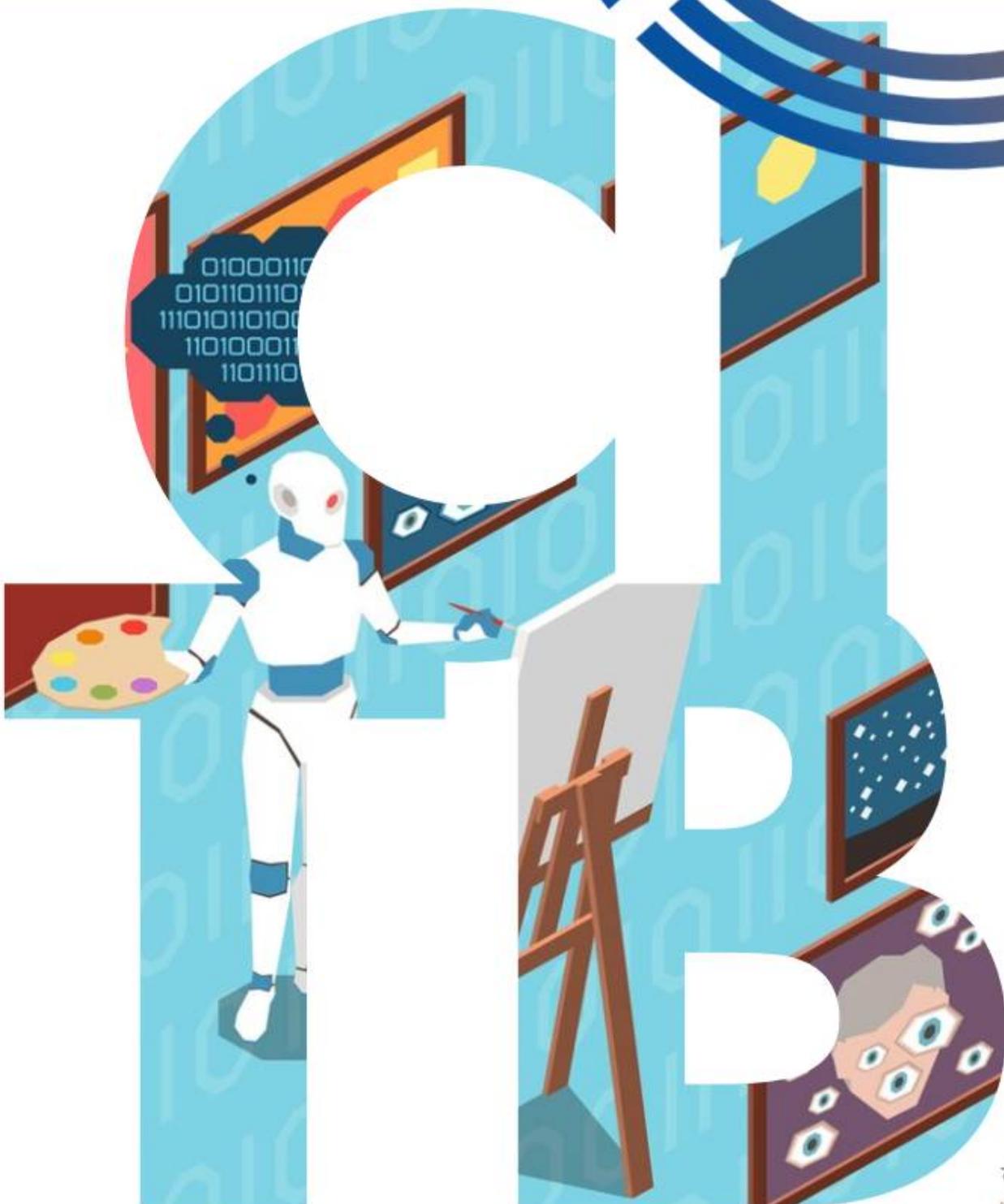


콘텐츠 제작 생성형 AI 서비스 등장: 콘텐츠산업 영향과 이슈

Content Industry Trend Brief 24-4호



목 차

- I. 서 론
- II. 콘텐츠 제작 생성형 AI 서비스 동향
- III. 콘텐츠산업에 미칠 영향과 이슈

요 약

- (목적) 생성형 AI의 최신 동향을 영상콘텐츠 제작 분야를 중심으로 분석하여 향후 발생하게 될 변화 양상과 주요 이슈들을 검토
- (서비스 동향) Text-to-Video에서부터 사운드, 배경 음악, 더빙, 번역 등 영상제작에 직접적으로 활용할 수 있는 다양한 서비스들이 등장하여 진화 중
 - 특히, 이번에 사전 공개된 Text-to-Video 서비스인 <Sora>는 기존 유사 서비스에서 한계로 지적된 맥락 일관성 문제 등을 극복하며, 발전 모멘텀을 예고
 - 그 외에도 사운드(일레븐랩스 등), 배경음악(사운드로우 등), 더빙(크로바 더빙 등), 번역(플리토 등)과 같이 영상제작에 직접적으로 도움을 주는 생성형 AI 서비스 등장
- (시사점) 콘텐츠산업에 미칠 영향과 이슈
 - 생성형 AI 활용 활성화는 콘텐츠산업의 IP 가치 증대, 글로벌화, 제작현장 변화를 가속화에 기여시킬 것임
 - IP 활용 및 확장 가능성의 확대에 따른 원천 IP의 중요성 증대
 - 언어문화적 장벽 완화로 글로벌 One-market 가속화
 - 제작비용 감소에 따른 1인/소수 창작의 활성화, 일자리 및 제작역량의 변화
 - 한편, 생성형 AI 활용 활성화로 가짜뉴스, 딥페이크, 저작권 문제, 할루시네이션, 데이터 오염, 탄소배출과 환경오염 등 부정적 외부효과에 대한 이슈가 부각될 것으로 보임
 - 한국 콘텐츠의 글로벌 위상을 보장하기 위해서는 생성형 AI 서비스 활용 확대에 나타날 수 있는 콘텐츠산업계 변화와 국제사회 이슈에 대한 정책적 대응이 필요한 시점임

작성자

양지훈

콘텐츠연구본부 연구원

yangjh@kcti.re.kr / Tel : 02-2669-6990

I. 서 론

1. 목적 및 범위

- (목적) 콘텐츠산업에 파급력이 높을 것으로 예상되는 생성형 AI의 최신 동향을 영상 제작 분야를 중심으로 분석하여 향후 발생하게 될 변화 양상과 주요 이슈들을 검토
 - 생성형 AI 기술의 영상콘텐츠 적용 시도가 활발해지면서 런웨이, 스테이블 비디오 디퓨전 등 관련 서비스들이 다양하게 출시되었으며, 최근에는 Open AI의 Text-to-Video 서비스 <Sora>가 사전 공개되면서 생성형 AI가 영상콘텐츠산업에 얼마나 큰 변화를 가져올지 이목이 집중됨
 - 생성형 AI 기술이 영상콘텐츠 제작의 전문영역까지 활용될 것이 예견되는 상황에서 발전 현황을 구체적으로 살펴봄으로써 향후 발생하게 될 변화들을 전망하고 정책적 이슈들을 선제적으로 논의
 - 생성형 AI 기술은 영상 콘텐츠산업의 제작 방식, 인프라, 일자리 등 산업 구조를 바꿀 수 있을 정도의 파급력이 예상되고 있기 때문에 급격한 변화 과정에서의 시행착오와 다양한 정책적 수요가 발생할 것으로 전망됨
 - 이에 본 동향 분석 보고서를 통해, 생성형 AI가 유발할 위협 요인들에 대해서는 대응 방향을 사전적으로 모색하고 기회 요인들에 대해서는 혁신을 위한 촉매제로 잘 활용하기 위한 정책적 대응 방안 마련에 기여하고자 함
- (범위) 영상콘텐츠 제작 분야의 국내 및 글로벌 생성형 AI 동향을 포괄적으로 분석
 - (산업적 범위) 영상콘텐츠 분야 중에서도 제작 분야에 집중하여 생성형 AI 기술이 적용된 다양한 서비스들의 동향을 파악
 - Text-to-Video 서비스뿐 아니라 영상콘텐츠 제작에 활용될 수 있는 사운드, 배경음악, 더빙, 통번역 등 다양한 서비스 동향 분석을 포괄
 - (시간적/공간적 범위) 시간적 범위는 생성형 AI가 본격적으로 발생한 2022년 11월에서 현재까지이며 공간적 범위으로는 국내와 글로벌 동향 모두를 포괄하여 조사

II. 콘텐츠 제작 생성형 AI 서비스 동향

1. AI 기반 Text-to-Video 서비스 동향

□ Open AI의 Text-to-Video 서비스 <Sora> 사전 공개

○ ChatGPT의 개발사 Open AI는 '24년 2월 15일에 Text-to-Video 서비스*인 <Sora>를 사전 공개**

* 텍스트 명령(프롬프트)을 입력하여 원하는 동영상 콘텐츠를 제작할 수 있는 서비스

** Open AI의 CTO 미라 무라티(Mira Murati)에 따르면 '24년 연말에 공식 출시 예정(WSJ, '24.3.13)

○ (사전 공개 의도) 정식 런칭이 아니라 현재 영상·디자인 등 관련 분야 전문가에게만 테스트를 진행 중이며, 사전 공개에서는 <Sora>가 생성한 48개의 비디오 콘텐츠 예시를 공유하여 <Sora>가 무엇을 할 수 있는지를 보여주고자 함(The Washington Post, '24.2.22)

- OpenAI는 공식적으로 서비스를 출시하기 전에 잘못된 정보, 증오, 편견, 성적 내용, 딥페이크 등 사회적 문제를 제기할 수 있는 부분을 사전에 방지하기 위해 전문가를 대상으로 베타 테스트를 진행하고 있다고 밝힘

- 언제 정식 서비스를 제공할지, 어느 정도의 서비스 이용 비용을 요구할지, 동영상 콘텐츠에 얼마나 많은 시간과 컴퓨팅이 소비되는지 등에 대한 구체적 정보는 발표하지 않음

[그림 1] Open AI의 Sora 공개 페이지 및 CI



자료: Open AI 홈페이지(<https://openai.com/sora>)

- (기술 보고서) Open AI에서는 서비스 공개와 더불어 ‘Video generation models as world simulators’라는 보고서를 발표하며 기술에 대해 구체적으로 설명
 - 다만, AI 모델이 어떠한 과정을 거쳐 텍스트를 구현하는지에 대한 자세한 내용들은 보고서에 포함되지 않음
- (핵심 기능: 일관성 있는 고품질 동영상 제공) 이전 서비스들과 차별화되는 특이점은 프롬프트에 대해 공간 모델링을 통해 3D로 생각하고 2D로 시각화하여 피사체를 조각적으로 이해, 기존 서비스와 차별화된 품질의 콘텐츠 제공하는 것
 - Open AI는 이를 ‘새로운 시뮬레이션 기능(Emerging simulation capabilities)’이라고 명명
 - 동영상의 사물과 피사체를 퍼즐과 같이 조각적으로 각각 ‘맥락’**에 따라 정리하여 시뮬레이션
 - * 예를 들어, 중력이 다양한 피사체에 미치는 영향 등을 각각 고려하여 지금까지의 다른 어떤 생성형 AI 서비스보다 훨씬 더 정교하게 시뮬레이션 하여 생생함을 극대화
 - ** 프롬프트에 중력이 어떻게 작용하는지, 물이 어떻게 튀고, 눈이 어떻게 쌓이는지 설명하지 않아도 맥락적으로 이해하여 텍스트에 따른 가장 적절한 시뮬레이션 결과를 제공
 - 이러한 과정을 통해 놀랍도록 일관성 있고 실제와 같은 공간과 피사체 표현
 - 고정된 사진을 붙이는 방식이라기보다는 게임엔진을 학습하여 3D 모델링을 구현하는 것으로 전문가들은 추측하고 있기도 함(Gizmodo, 2024.2.16.)

[그림 2] <Sora>가 제작한 동영상 콘텐츠에서의 3D 공간 및 털 표현



자료: Open AI 홈페이지(<https://openai.com/sora>)

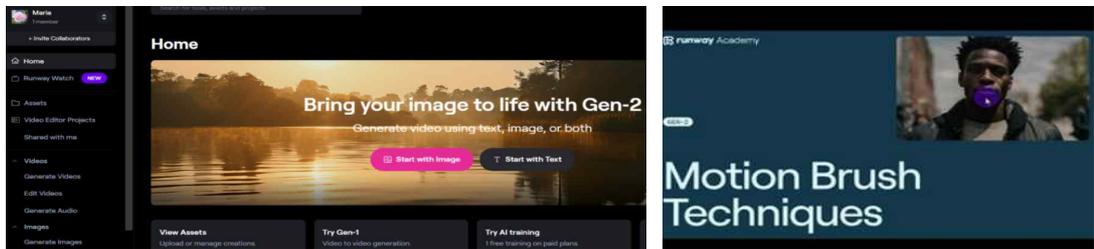
- Dall-E 3 모델의 연구를 기반으로 프롬프트의 논리를 언어적으로 이해*하고 생성
 - * 물체들이 실제 세계에서 어떻게 행동하는지 이해하고 그에 따라 자연스럽게 생성
- (추가 기능) 텍스트로 동영상 생성 뿐 아니라, 이미지의 애니메이션화*, 동영상 시점 변환**, 비디오 연결***, 이미지 생성****도 가능
 - * 정적 이미지를 입력하면 그 이미지를 기반으로 동영상 생성
 - ** 다양한 시작점을 가진 비디오를 생성하고 원하는 결말 도출
 - *** 두 개의 입력 동영상을 자연스럽게 연결
 - **** Diffusion Model을 통해 2048 X 2048 해상도의 다양한 크기의 이미지 생성

□ 기존 Text-to-Video 서비스

○ 런웨이(Runway)

- '18년에 온라인 영상 편집 및 제작 사이트로 시작한 런웨이는 생성형 AI 기술을 접목하여 동영상 생성 서비스를 선도적으로 시도
- 현재 텍스트를 동영상으로 만드는 기능, 이미지를 동영상으로 만드는 기능, 이미지와 텍스트를 동영상으로 만드는 기능 등을 제공 중
- '23년 12월에는 게이이미지와 제휴하는 등 다양한 이미지 데이터를 확보하여 결과물에 대한 품질 개선 노력을 진행
- 모션 브러시를 활용하여 브러시로 지정된 피사체를 마음대로 움직일 수 있는 모션 브러시 기능이 탑재

[그림 3] 런웨이 구동화면 및 모션 브러시 기능 소개 화면

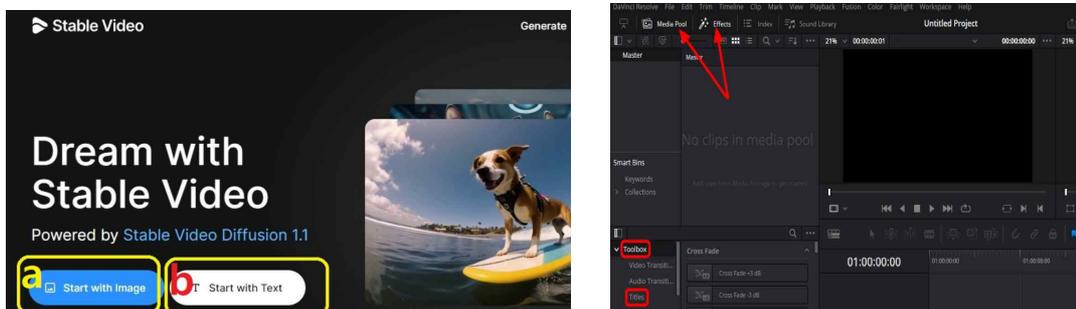


자료: 런웨이 홈페이지(<https://runwayml.com/>)

○ 스테이블 비디오 디퓨전(Stable Video Diffusion)

- '스테이블 디퓨전'이라는 생성형 AI 기반 Text-to-image 서비스로 유명세를 탄 스테빌리티 AI는 Text-to-video 서비스인 스테이블 비디오 디퓨전을 '23년 11월에 출시
- 초당 24프레임의 영상을 제공하며, 25개의 생성 프레임과 24개의 FILM 보간 프레임으로 구성된 2초 분량의 비디오를 평균 41초 내의 분량으로 생성
- 여러 개의 비디오 클립을 연결, 긴 동영상을 만들 수도 있음

[그림 4] 스테이블 비디오 디퓨전 구동 화면

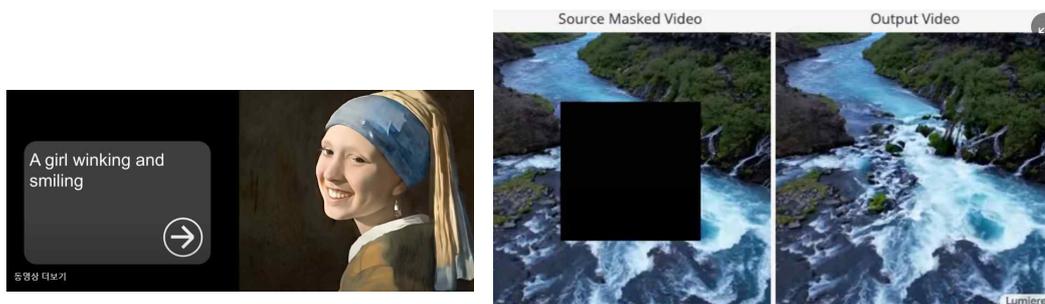


자료: 스테이블디퓨전 홈페이지(<https://stability.ai/>)

○ 구글, 루미에르(Lumiere)

- 구글은 비디오 생성용 시공간 확산 모델(Space-Time Diffusion Model for Video Generation)을 적용한 동영상 생성 모델 <루미에르(Lumiere)>를 '24년 1월에 발표, 서비스는 아직 런칭하지 않음
- 텍스트나 이미지 입력으로부터 생성되는 비디오의 공간적 사실성과 시간적 일관성 개선 시도
 - * 비디오 속 사물이 있는 위치에 대한 공간적 측면과 동영상 전체에서 사물이 어떻게 움직이고 변화하는지에 대한 시간적 측면을 동시에 처리하도록 설계
- 캡션이 달린 3,000만개의 비디오 데이터 셋(data set)으로 훈련했으며, 1024×1024픽셀의 초당 16~80프레임으로 최대 5초 길이의 비디오를 생성
- 특정 개체를 삽입하는 인페인팅 등의 다양한 기능도 제공할 예정

[그림 5] 구글 <루미에르> 설명 화면 및 인페인팅 구현 화면



자료: 구글 <루미에르> 홈페이지(<https://lumiere-video.github.io/>)

□ 기존 서비스와 <Sora>와의 차이점

- 기존 서비스의 결과물들은 화질이나 디테일 무너지는 건 당연하고 프레임이 충분하지 않아 뚝뚝 끊기는 현상이 보였음
 - 여러 이미지를 붙인 것 같이 구현되어 결과물 화면이 튀거나 일렁이는 현상 발생
- 반면, <Sora>의 경우 프레임이 굉장히 자연스럽게 맥락에 맞는 일관성이 60초라는 상대적으로 긴 시간동안 지켜져 이전과 비교해서 엄청난 품질의 향상이 느껴짐
 - 편집과 컷이 세련되게 자동으로 이루어지고, 빛에 대한 이해와 안경이나 선글라스와 같은 소품으로 굴절되는 현상은 물론, 슬로우 모션이나 구도, 앵글과 같은 연출이 매우 자연스러움

[그림 6] 같은 명령어(프롬프트)에 따른 동영상 콘텐츠 결과물 비교



자료: 유튜브 채널<AI Community> 갈무리(<https://www.youtube.com/watch?v=rSnsYfC8QHQ>)

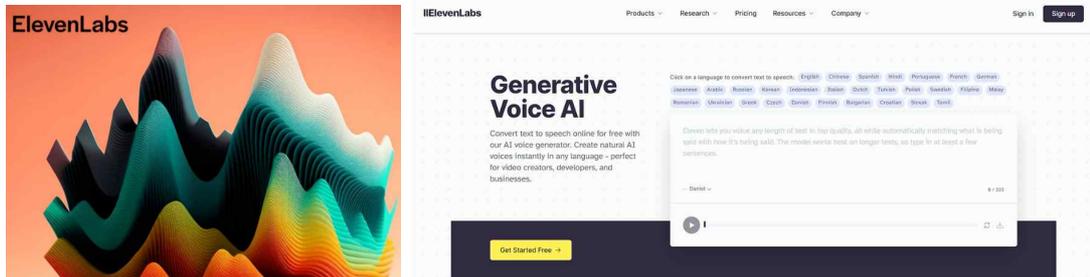
2. 기타 영상콘텐츠 제작 관련 생성형 AI 서비스

□ 사운드, 배경 음악

○ 일레븐랩스(Eleven Labs)

- '22년에 설립한 업체로, 텍스트를 음성이나 사운드로 생성하는 서비스를 제공
- 최근 X(트위터)에 오픈시의 동영상 생성 AI <소라>가 만들어낸 동영상에 자신들의 기술로 만들어낸 효과음을 생성
- 자동차 달리는 소리와 지하철 소리, 거리를 걷는 여자의 구두발 소리 등 효과음 등
- 사운드 생성은 음성이나 음악보다 활용도가 제한적이고 다양한 데이터 학습이 필요한 탓에 관심이 덜했지만, 동영상 제작과 결합되며 큰 관심을 받고 있음
- 일레븐랩스의 기본 서비스는 사용자의 목소리를 학습하여 여러 언어로 말할 수 있게 도와주는 음성 서비스에서 시작해, 다양한 사운드 분야로 진출 시도 중

[그림 7] 일레븐랩스 소개 화면과 서비스 구동 화면



자료: 일레븐랩스 홈페이지(<https://elevenlabs.io/>)

○ 가우디오랩(Audio Lab)

- 국내의 인공지능(AI) 오디오 전문 기업으로
- 실시간 소음 제거, 목소리 강화 등으로 콘텐츠 몰입도를 높이는 '저스트 보이스', 이미지 입력에 대응하는 효과음을 AI가 자동으로 생성하는 '폴리', 실시간으로 원곡 음악을 노래방 음원으로 바꿔주는 '가우디오 씽' 등을 제공
- 동영상 제작에서는 별도 텍스트 작업을 거치지 않아 작업시간을 획기적으로 줄일 수 있는 게 장점
- 실제로 <Sora> 제공 동영상에 이 기술을 적용하여 동영상 화면에 따라 소리가 나오는 콘텐츠를 제시하기도 함

[그림 8] 가우디오랩 Ci와 <Sora> 동영상에 사운드를 삽입한 콘텐츠

GAUDIO

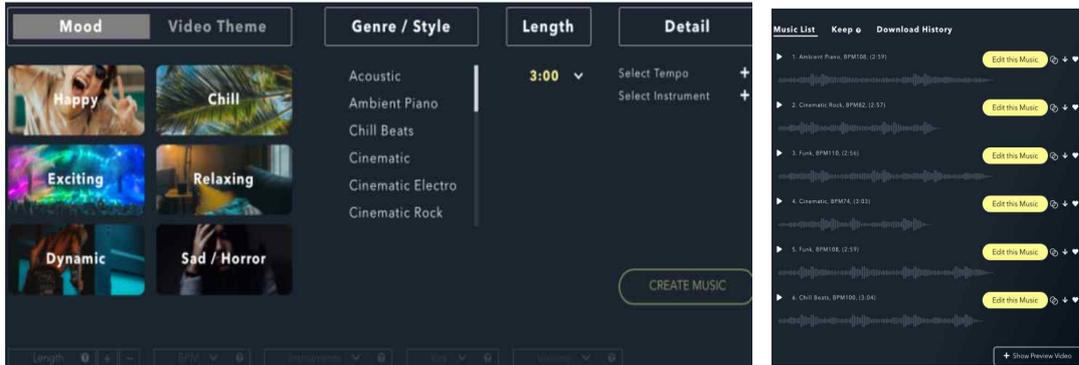


자료: 유튜브 채널<GAUDIO> 갈무리(<https://www.youtube.com/watch?v=B3qb6Hr03Y>)

○ 사운드로우(Soundaw)

- AI가 작곡을 하는 서비스를 제공하는 서비스
- 분위기, 장르/스타일, 길이, 빠르기(템포), 악기를 골라 작곡을 진행
- 패턴 에디터가 있어 이를 이용하면 시간대별로 음악의 흐름을 편집
- 다만, 사운드로우를 통해 생성된 모든 곡의 저작권과 소유권은 사운드로우에 있으며 사용자에게는 사용권만 주어지며, 수익 창출을 목적으로 음악을 동영상 등에 활용하는 것은 괜찮지만, 음악 자체를 판매하는 행위는 금지

[그림 9] 사운드로우 구동 화면



자료: 사운드로우 홈페이지(<https://soundraw.io/>)

□ 더빙, 번역

○ 크로바 더빙(Clova Dubbing)

- 클로바 더빙은 CLOVA Voice의 음성합성 기술로 제작한 AI 보이스로 동영상에 쉽게 더빙할 수 있는 서비스
- 어른, 아이, 여자, 남자, 외국인 등 다양한 보이스와 여러 종류의 언어와 감정 표현이 제공되어 목적에 맞는 자연스러운 콘텐츠를 제작 가능

[그림 10] 크로바 더빙 구동 화면

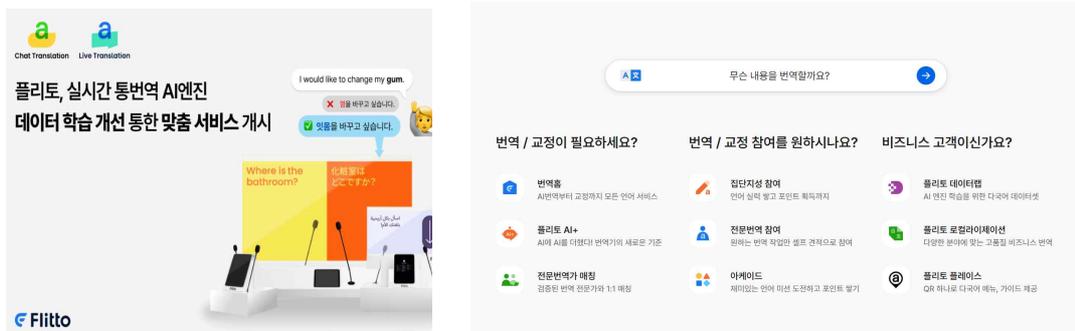


자료: 크로바 더빙 홈페이지(<https://clovadubbing.naver.com/>)

○ 플리토(Flitto)

- 국내의 생성형 AI 기반 번역 전문 서비스
- 플리토는 인공지능(AI) 기반 실시간 통번역 솔루션 '챗 트랜슬레이션(대화 번역)' '라이브 트랜슬레이션(라이브 번역)' 서비스를 제공
- CT 엔진을 활용하는데, STT(음성-텍스트 변환) 엔진과 AI 번역 엔진을 결합한 것으로, 고품질의 다국어 병렬 말뭉치 데이터와 음성 데이터 학습을 통해 음성 인식률과 번역 정확성을 높인 기술

[그림 11] 플리토 구동 화면



자료: 플리토 홈페이지(<https://www.flitto.com/portal/ko>)

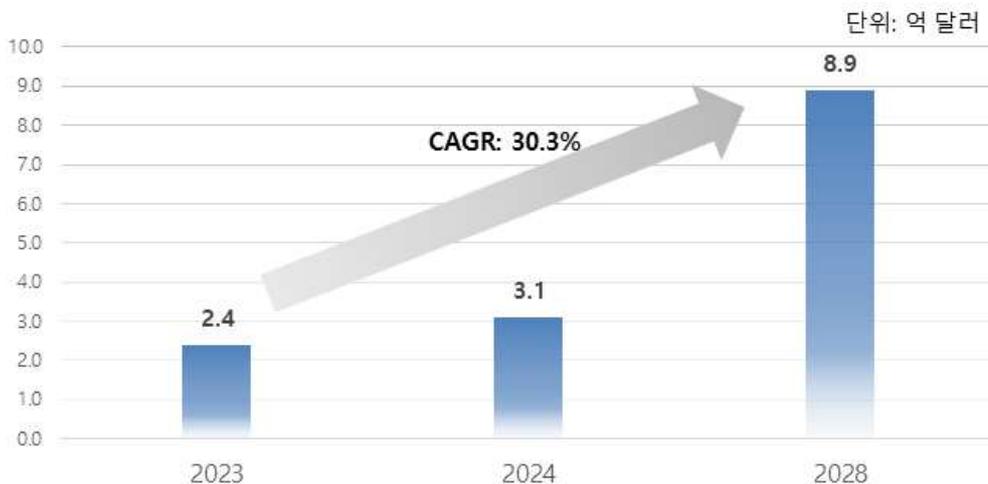
Ⅲ. 콘텐츠산업에 미칠 영향과 이슈

1. 콘텐츠산업의 영향

□ 관련 시장 규모 전망

- Text-to-Video 시장은 '23년 2억 4천만 달러 규모에서 '28년 8억 9천만 달러 규모로 연평균 성장률(CAGR) 30.7%를 기록할 것으로 예상됨(Research & Markets, 2024)
 - 시장 조사업체 Research & Markets에 따르면 시각적 콘텐츠에 대한 수요 증가, 소셜 미디어 사용량 급증, 접근성 기능 통합, 미디어콘텐츠 산업의 수요 증가 등에 기인하는 것으로 분석
 - 여기에 다국어 기능 통합, 온라인 학습 확장, 실시간 비디오 생성, 향상된 얼굴 표정 인식, 시각적 현실감 및 품질의 제고와 같은 기술적 발전도 성장에 큰 요인으로 작용할 것으로 전망

[그림 12] 글로벌 Text-to-Video 예상 시장규모



자료: Research & Markets, 2024

□ 예상 시장 변화

○ IP 중요성 증대

- 잘 개발된 캐릭터와 이미지 한 장이 쉽게 웹툰이 되고, 애니메이션이 될 수 있을 뿐 아니라 영화/방송 콘텐츠로 확장이 가능
- 생성형 AI를 통해 영상화가 웹툰과 웹소설 콘텐츠가 애니메이션이나 영화/방송 콘텐츠로 제작되는 OSMU가 더욱 활성화되어 IP 비즈니스는 더 고도화 될 것으로 예상됨
- 배우나 아이돌 IP의 활용도도 높아져 이들이 직접 촬영장에서 제작에 참여하지 않아도 초상권을 제공하기만 해도 콘텐츠를 만들 수 있는 환경이 열릴 수 있음
- 완성품뿐만 아니라 촬영현장에서 내용과 연출의 이해도를 높이기 위해 제공하는 콘티와 같은 자료도 생성형 AI를 통해 애니메이션 형태로 활용이 용이

[그림 13] 다른 장르의 영상화 예시



- 생성형 AI를 통해 새로운 캐릭터를 만들거나 자신이 연기하는 화면 위에 덧입혀 활동하는 버튜버 등이 더 각광 받을 수 있음
- 가상현실을 활용한 부캐나 원하는 외형으로 활동이 가능해져 개인미디어에서도 가상 캐릭터가 더욱 활성화 될 것으로 예상됨
- 현재도 단순한 수준의 그림이나 그래픽을 입혀 활동하고 있는 버튜버는 새로운 트렌드로 자리잡으며 일본, 미국, 영국을 비롯해 여러 국가에서도 구독자 100만명 이상을 모은 버튜버가 활동 중

[그림 14] 버추얼 아이돌 <Plave>와 버튜버 <미츄>

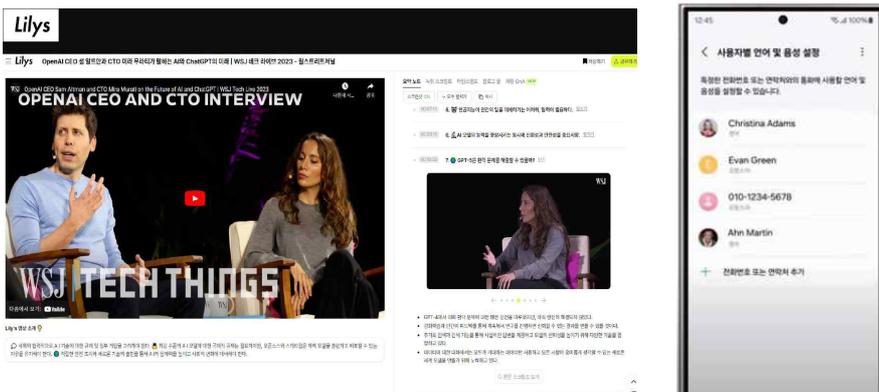


자료: Valst 홈페이지, 미츄 홈페이지

○ 글로벌 One-Market의 가속화, 글로벌 수요의 중요성 증대

- 넷플릭스, 유튜브 등 글로벌 플랫폼 등장으로 인해 전 세계 소비자들이 하나의 플랫폼에 집중되는 현상이 시를 통해 언어장벽을 허물며 더 집중되는 현상이 발생할 것
- 현재도 실시간 번역과 자막이 시를 통해 지원되고 있지만, 향후 AI 통번역이 더 고도화되고 입술과 표정의 AI 편집이 가능해지면 전세계 콘텐츠가 모두 모국어로 들리는 환경이 열릴
- 이에 따라, 콘텐츠의 기획과 제작에 있어서 전 세계의 소비자의 보편성과 문화적 코드를 고려한 전략이 요구되어짐

[그림 15] 실시간 ai 유튜브 스크립트 번역 서비스 <Lilys>(좌), 갤럭시 S24의 실시간 통역 통화 기능(우)



자료: Lilys 홈페이지, 유튜브 삼성전자 채널

○ 콘텐츠 산업 일자리 변화

- 직접 촬영 없이 특수한 상황까지 시를 통해 동영상 제작이 가능해짐에 따라 직접 촬영 시 발생하는 비용과 인력의 감축이 이루어질 수 있음
- * <Sora> 예시 동영상에서 나타난 콘텐츠를 예로 들자면, 하늘에서 촬영하는 드론 기사나, 카메라 촬영에 필요한 레일이나 그립팀이 필요 없으며, 군중처리를 위한 CG처리, 엑스트라 인원, 로케이션 매니저 관련 비용이 불필요

- 기존에는 동영상으로 구현할 수 없는 상황까지 적은 비용으로 상상력을 영상으로 표현할 수 있는 요건이 마련

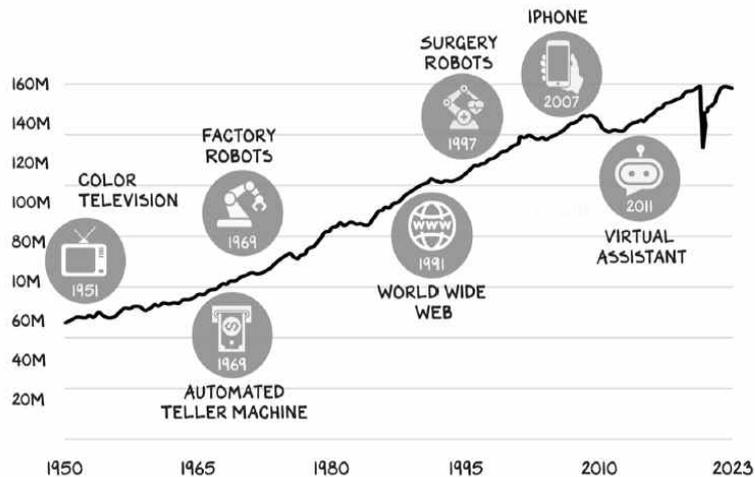
〈표 1〉 생성형 AI 동영상 제작을 통해 불필요해질 역할 및 직종 예시

역할	AI 대체 위기 직종
촬영	촬영 감독, 촬영팀
배경/세트 디자인	미술 감독, 미술팀, 그래픽디자이너
군중 처리	엑스트라, 그래픽 디자이너
후가공	그래픽 디자이너, 편집자
조명	조명 감독/조명팀
드론 촬영	드론 기사
주연/조연/기타 연기	배우

- 이미 미국 기업 4곳 중 1곳은 생성 AI로 직원을 대체하고 있다는 조사 결과도 있으며(레주메 빌더닷컴, 2023)
- 골드만삭스는 AI 도입의 생산성 향상 효과가 신흥국보다 선진국에서 클 것이며 미국 등 선진국 업무의 25%, 신흥국 업무의 20%를 자동화 할 것으로 예측(골드만삭스, 2023)
- 실제로 2023년 할리우드의 작가조합과 메이저 스튜디오들은 주요 쟁점이었던 인공지능(AI) 대본 활용 이슈로 148일간 파업이 진행됨
- 반면, 미국의 직업군별 고용 분포를 연구한 보고서에서는 지난 80년 동안 미국 뉴욕시 고용의 85%가 1940년 전에 없던 직종이새로 생겼다는 분석을 토대로 혁신기술이 새로운 직종을 창출할 것으로 주장(NATIONAL BUREAU OF ECONOMIC RESEARCH)이 있기도 함

[그림 16] 혁신기술 발생에 따라 새로운 일자리가 나타날 것이라는 통계 근거

NUMBER OF EMPLOYED AMERICANS



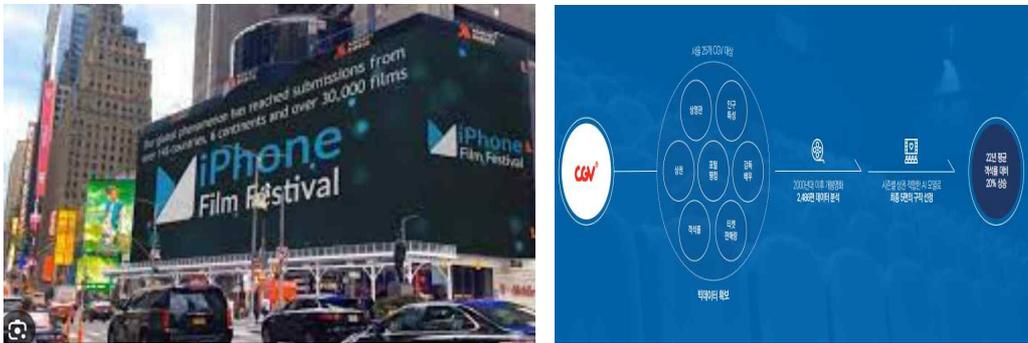
SOURCES: BUREAU OF LABOR STATISTICS, TRADING ECONOMICS (Scott Galloway, 미국 노동 통계국)

- 촬영, 후가공, 디자인 등 영상 콘텐츠 제작 관련 고급인력 전문성에 강점을 지닌 우리나라의 입장에서는 일시적으로 큰 위협으로 작용할 가능성이 높기 때문에, 정책적 대응마련이 요구됨

○ 제작 비용의 감소에 따른 1인/소수 창작의 활성화

- 영상콘텐츠의 제작비가 감소하고 제작 방식이 쉬워짐에 따라 그동안 진입장벽이 높았던 영화 콘텐츠나 방송 콘텐츠의 제작이 개인 수준에서도 가능해짐
- 유튜브를 중심으로 형성된 1인 방송시장도 스마트폰을 통한 고화질 촬영이 가능해지고, 마이크 등의 시설이 대중화되어 제작 진입 장벽이 낮아지면서 활성화됨
- 1인 독립영화를 비롯해서 시를 통해 블록버스터 콘텐츠까지 제작이 가능해지면 장편영화까지 소수나 개인이 콘텐츠를 제작할 수 있는 시대가 기대됨
- 스마트폰 영화제처럼 1인 제작 영화제나 생성형 AI 영화제 등이 등장할 여지가 있을 정도로 주된 제작방식으로 자리할 전망
- 인기 콘텐츠가 시청률 50% 이상을 차지하기도 했던 과거의 방송시장과 달리 다양한 콘텐츠가 발생하고 취향이 파편화되어 초개인화를 촉진
- 이에 따라 소비자 데이터분석과 알고리즘의 역할과 영향력은 더 커질 전망

[그림 17] 아이폰 영화제(좌), 데이터분석을 통해 극장 개방사례_CGV 우리동네 명작(우)



자료: iPhone Film Festival 홈페이지, CJ 뉴스룸(시가 추천하는 지역 맞춤 영화... CGV, '우리동네 명작 기획전' 진행)

2. 주요 이슈

○ 가짜뉴스(Fake News), 딥페이크(Deep Fake)

- 잘못된 정보를 유포하거나 허위 콘텐츠를 조작하는 데 잠재적으로 오용될 수 있다는 우려
- 생성된 비디오는 부정확한 정보를 제공하거나 시청자를 속이기 위해 쉽게 조작될 수 있기 때문에 생성된 비디오의 진위성과 신뢰성이 면밀한 조사 필요
- 전기전자공학협회는 잘못된 정보 유포, 공격적이거나 유해한 콘텐츠 생성, 조작이나 협박을 위한 딥페이크 비디오 생성 등 악의적인 목적으로 AI 콘텐츠 생성기를 오용할 가능성을 있음을 강조(Institute of Electrical and Electronics Engineers, 2023)
- AI 개발자 Havemeyer는 워싱턴포스트지와 인터뷰에서 2024년 대선을 앞둔 상황에서 AI를 통한 Fake 뉴스는 더욱 주의해할 것으로 지적(Washington Post, 2024)

[그림 18] 가짜 뉴스 사례



자료: techchurch(2023.03.29, From Balenciaga Pope to the Great Cascadia Earthquake, AI images are creating a new reality)

- 러시아-우크라이나 전쟁이 한창이던 2022년 3월, 블로디미르 젤렌스키 우크라이나 대통령이 러시아에 항복을 선언하는 내용의 가짜 뉴스 영상이 유튜브와 페이스북 등에 유포된 사례도 있었음
- 딥페이크 기술이 단순 개인의 인권침해를 넘어 국가·사회적 혼란까지도 초래할 수 있음
- 유명인이나 지인의 초상권을 성적인 용도로 딥페이크 작업을 하여 심각한 인권 침해를 일으키는 문제도 발생 중

○ 저작권 문제

- 생성형 AI의 경쟁력은 얼마나 많은 데이터를 학습했는지에 따라 좌우되는데, 이 데이터가 누군가의 저작물일 경우 심각한 저작권 침해 유발
- 구매한 데이터를 가져다 학습하지 않는 한 저작권 침해 논란이 빚어질 수밖에 없지만, 결과물을 보고 어떤 데이터를 가지고 학습을 시켰는지 알기가 어려움

[그림 19] 네이버 웹툰 ‘도전 만화’ 작가들의 ai 웹툰 보이콧 화면(좌) 음저협에서 저작권료 지급을 인정하지 않은 ai 작곡가 이봄(우)



자료: 연합뉴스(2023.06.04., 거센 반발 직면한 시웹툰…네이버웹툰 도전만화서 보이콧 운동도), 유튜브 세종테크노파크 채널

- 한편, 생성형 AI의 활용이 급속도로 확대되자 저작권 관련 논의가 이미 활발하게 진행되고 있으며 향후 문제와 혼란을 가중시킬 것에 대한 우려가 커짐
- 생성형 AI를 통해 창작물을 만들었을 때, 현행법상 저작권물로 인정이 어렵지만 제작자, 이용자, 경우에 따라서는 공동저작자라는 견해 등이 제시되며 아직까지 법리적으로 명확하게 정리되지 못함

미국 저작권청의 AI 저작물 등록 가이드라인

- AI 저작물 관련 논란이 심화되자, 2023년 3월 16일에 미국 저작권청은 AI로 생성된 콘텐츠에 대한 저작권 심사 및 등록기준 가이드라인을 발표
- 해당 작품이 단지 보조 도구로 컴퓨터 등을 이용하여 인간이 저작한 것인지, 인간이 아닌 기계에 의해 실제 고안되고 실행된 것인지 검토(판단은 AI 도구가 어떻게 작동하였는지, 최종 결과물을 창작하는데 어떻게 이용하였는지에 따라 달라짐)
- 프롬프트는 마치 작품 창작을 촉탁하면서 아티스트에게 내리는 지침과 유사하다고 해석
- 다만 인간이 AI에서 생성된 자료를 충분히 창의적인 방법으로 선택하거나 배열할 경우 결과물은 전체적으로 인간의 저작물이 될 수 있는 것으로 제시

○ 할루시네이션(환각현상), 데이터 오염

- 현재까지 가장 완성된 형태로 예측되는 생성형 AI 동영상 서비스인 <Sora>의 예시 영상에서도 아직까지도 잘못된 부분이 발견
 - * 체스의 왕이 3개 존재하고 체스 판이 8X8이 아니라 7X7이거나, 촛불을 붙였지만 촛불이 그대로 있기도 하고, 의자를 옮기는 중에 모래가 겹치는 등 오류 발생

[그림 20] <Sora> 예시 동영상에서 발견된 오류들



자료: Open AI 홈페이지(<https://openai.com/sora>)

- 문제는 ai는 학습 데이터를 가지고 강화학습을 통해 새로운 동영상 결과물을 산출하는데 자료로 활용하게 되는데, 잘못된 정보가 반복되면 잘못된 정보가 사실처럼 이해될 수가 있음
- 특히, 최근 들어 쇼츠 동영상을 ai 텍스트 생성 서비스와 ai 동영상 생성 서비스를 활용하여 제작하고 있는 콘텐츠가 넘쳐나고 있는데, 이러한 ai 콘텐츠가 누적되면 이들이 학습대상이 됨

- AI 데이터가 다수를 차지하게 되면 결국 오염된 정보를 지속적으로 강화학습하게 되어 개선의 여지가 없어지게 됨

[그림 21] ai를 통해 제작되는 유사 콘텐츠가 많아진 유튜브 화면



자료: 유튜브 '우주의 끝' 검색 화면

○ 탄소배출과 환경 오염

- 생성형 AI가 학습하고 동영상이라는 결과물을 도출하는 과정 가운데 매우 많은 컴퓨팅이 필요하여 많은 양의 전기를 필요
- 즉, AI는 데이터를 학습하고 결과를 산출하는 과정에서 엄청난 양의 에너지를 사용하고 탄소를 배출
- AI 모델이 발전할수록 학습에 쓰이는 데이터 세트가 늘어 에너지 소비도 증가
- 'AI 인덱스 리포트 2023'에 따르면 언어 생성형 AI 모델인 GPT-3가 학습으로 배출하는 이산화탄소량은 502톤으로, 한 사람이 1년간 배출하는 양(5.51톤)의 91배에 달함(Stanford University, 2023)
- 실제로 Open AI가 <Sora>를 발표하면서 제공한 기술 리포트에 따르면, 동일한 프롬프트에 대해 컴퓨팅 파워에 따라 품질이 어떻게 달라지는지 구체적으로 보여주기도 함
- 즉, 높은 품질의 동영상 결과물을 산출하기 위해서는 그 만큼 높은 수준의 전기와 컴퓨팅 파워를 요구

[그림 21] 컴퓨팅 파워에 따른 결과물의 품질 차이



자료: 더코어(2024.02.18., OpenAI 소라(Sora), 숨겨진 혁명 이야기)