

생성형 AI 기반의 콘텐츠 제작 기술 동향과 안전성 이슈

Trends and Safety Issues in Content-Creation Technology Development
Based on Generative AI

유정재 (J.J. Yu, jungjae@etri.re.kr)

남도원 (D.-W. Nam, dwnam@etri.re.kr)

이정수 (J.S. Lee, jslee2365@etri.re.kr)

이지원 (J.W. Lee, ez1005@etri.re.kr)

문성원 (S.-W. Moon, moonstarry@etri.re.kr)

송대영 (D.Y. Song, eadyoung@etri.re.kr)

지능형콘텐츠인식연구실 책임연구원

지능형콘텐츠인식연구실 책임연구원

지능형콘텐츠인식연구실 책임연구원

지능형콘텐츠인식연구실 책임연구원

지능형콘텐츠인식연구실 책임연구원

지능형콘텐츠인식연구실 연구원

ABSTRACT

Recently, the generative AI results announced by big tech companies have been surprising in terms of their quality and completeness. Moreover, attempts are being made to utilize generative-AI technology in various existing content industries, such as novels, webtoons, and movies. However, when the released technologies are used directly, unlike the results selectively disclosed by developers, users have found room for improvement in terms of the accuracy and convenience of generative control, as well as problems with content consistency between individual generative results. Therefore, long-form content production automation is still limited in many cases. In this paper, we introduce the development trends of production-support technologies that can practically utilize the current level of generative AI in the production of novels, webtoons, and short video content, and briefly discuss the safety issues from a sociocultural perspective.

KEYWORDS AI, 동영상, 생성, 안전성, 영상, 웹소설, 웹툰, 이미지, 인공지능, 제작, 콘텐츠, 텍스트, 편견

I. 서론

2020년 6월, OpenAI가 개발한 생성형 인공지능(AI) GPT-3가 발표되며 전 세계적으로 큰 주목을

받았고, 이어서 생성형 AI를 현실 업무에 적용하려는 다양한 시도가 이어졌다. 미국 채프먼대학교의 방송영상 전공 학생들은 자신들이 작성한 단편 이야기를 GPT-3에 입력해서 각색된 영화 시나리오

* DOI: <https://doi.org/10.22648/ETRI.2025.J.400204>

* 본 연구는 문화체육관광부의 재원으로 한국콘텐츠진흥원의 지원을 받아 수행되었음[과제명: 사용자들의 다양한 관심사 맞춤형 문화유산 스토리텔링 AI 기술 개발, 과제번호: RS-2023-00220195, 기여율: 100%].

를 생성하고, 이를 기반으로 단편영화 “Solicitors”[1]를 제작했다. 또한, 코카콜라는 30년 전 광고였던 “Holidays are Coming”을 생성형 AI를 활용하여 재해석한 광고 영상을 공개하여 주목받기도 했다.

콘텐츠 제작 관점에서 생성형 AI 기술의 가장 큰 장점은 대규모 리소스와 전문성을 지니지 않은 창작자도, 아이디어만으로 해당 도메인의 콘텐츠를 제작 가능하다는 점이다. 게임, 영화, 애니메이션 등 생성형 AI가 활용될 수 있는 분야는 광범위하지만, 본고에서는 개인 또는 소규모 팀 단위로 제작 가능한 소설류, 웹툰 및 단편 영상 콘텐츠 분야에 집중하여 생성형 AI가 창작활동을 실질적으로 도울 수 있는 제작 지원 기술의 현황을 살펴보고 사회문화적 관점에서의 안전성 이슈까지도 간략히 소개하고자 한다.

II. 생성형 AI 기술 동향

현재 개발되는 생성형 AI 기술은 생성 결과물의 형태 관점에서 텍스트, 이미지, 동영상, 3D 모델 생성형 AI 기술로 분류할 수 있다. 앞서 출판되었던 전자통신동향분석 기고문[2]과의 중복성을 피하고, 앞서 설명한 바와 같이 소설, 웹툰, 단편 영상 콘텐츠 분야에서의 제작 지원 기술에 집중하기 위해서 본고에서는 텍스트, 이미지, 동영상 분야에서의 생성형 AI 최신 기술 동향만을 간략히 정리하고자 한다.

1. LLM 기술 동향

AI에 기반하여 전문가 수준의 텍스트를 생성하고 분석하는 대규모 언어 모델(LLM: Large Language Model)은 산업계를 중심으로 빠르게 발전하고 있다. 이는 인간의 능력에 필적하거나 이를 넘어서는 것을 목표로 하는 범용인공지능(AGI: Artificial General

Intelligence)에 가장 가까이 다가간 분야라 할 수 있다.

현재 시점에서 출시된 LLM 중 기술적으로 선두에 있는 모델은 OpenAI의 GPT-4o, Google의 Gemini, 그리고 Meta의 LLaMa, 이 3가지로 볼 수 있다.

GPT-4o는 OpenAI에서 2024년 5월에 공개한 모델로 대표적인 LLM 벤치마크 6개 중 5개에서 1위를 석권할 만큼 우수한 텍스트 생성/분석 성능을 가진 것으로 알려져 있다[31]. 전 세계 언어의 97%에 해당하는 50개 이상의 언어를 지원하며 대화에서 맥락과 뉘앙스를 파악하는 등 강력한 자연어 이해 능력으로 유명하다.

Google에서 공개한 Gemini의 경우 현재 2.0버전이 출시되었으며, Google의 강력한 검색 기능을 활용하여 생성한 텍스트에 대한 사실 확인뿐만 아니라 검색에 기반한 최신 지식 답변이 가능하다. 이는 다른 LLM에는 존재하는 지식 컷오프(지식을 언제까지 학습했는지)가 존재하지 않는 것이며, 따라서 실용적으로 사용하기에 더 적합하다.

Meta에서 공개한 LLaMa는 현재 오픈소스 모델 중에서 유일하게 상업용 폐쇄형 모델들과 경쟁할 수 있는 수준의 모델이다. 이 모델은 2024년 12월 기준 3.3 70B 버전이 공개되었고, 현재까지 3억 건 이상 다운로드되었다[3]. LLaMa는 또한 8B, 70B, 405B 등 다양한 크기의 모델을 공개하여 용도에 맞게 파인튜닝 및 커스터마이징할 수 있어서 다양한 연구와 응용으로 쉽게 확장 가능하다.

2. 이미지 생성 기술 동향

가. 이미지 생성 관련 연구동향

잠재 확산 모델(Latent Diffusion Model)[4] 연구에 기반한 대표적인 Text-to-Image 생성형 AI 모델인 스테이블 디퓨전(Stable Diffusion)의 공개 이후 다양한 이미지 생성형 AI 연구 방향이 제시되었다. 특히

이미지를 원하는 형태로 생성하는 방법은 다양한 접근 방식을 통해 끊임없이 연구되고 있는 현실이다. 대표적인 초기 연구로는 적은 수의 이미지를 활용하여 사용자가 원하는 개념을 임베딩 공간의 새로운 단어를 통해 재현하는 Textual Inversion[5]이나 사전 학습된 초기대 모델의 Spatial Conditioning 제어를 통해 원하는 형태로 생성 이미지를 제어하는 ControlNet[6]과 같은 기술이 있다.

저장대학교와 알리바바가 2025년 공개한 EliGen[7]은 생성 이미지를 개체(Entity) 수준으로 제어하는 기술이다. 이 연구는 디퓨전 트랜스포머(Diffusion Transformer)를 위한 지역적 어텐션(Regional Attention) 메커니즘을 제안하여 기존 글로벌 프롬프트를 통한 이미지 생성형 AI 제어를 대신해 로컬 프롬프트와 마스크 활용 기능을 제공한다. 이러한 접근은 기존의 직사각형 형태의 바운딩 박스 방식 제어를 넘어 다양한 형태의 마스크에 대해 자유로운 개체 수준 제어를 가능하게 하였다.

또 다른 최신 이미지 생성 및 제어 기술로는 엔비디아(NVIDIA)가 2024년 공개한 Edify Image[8]가 있다. 이 기술은 라플라시안 확산(Laplacian Diffusion) 방식을 통해 Text-to-Image 생성, ControlNet을 통한 이미지 제어, 4K 업샘플링과 같은 다양한 기능을 제공하면서 높은 품질의 이미지를 생성한다.

나. 이미지 생성 관련 상용 서비스

생성형 AI를 활용하여 이미지를 생성하는 대표적인 상용 서비스는 DALL · E3[9]와 Midjourney[10]를 얘기할 수 있다. OpenAI의 DALL · E3는 HD급 ($1,792 \times 1,024$ 픽셀) 고해상도 이미지 생성을 지원하며, GPT-4를 이용하여 사용자가 입력한 프롬프트 입력을 생성모델에 전달하기 전에 최적화하는 Prompt Rewriting 기능을 제공한다. 또한 DALL · E3는 현재 ChatGPT에서의 멀티-턴 대



출처 DALL-E3, OpenAI, Accessed: Feb. 1, 2025. <https://openai.com/index/dall-e-3/>

그림 1 연속적인 이미지 생성 사례(ChatGPT-DALL-E3)

화 형식을 통하여 일반인들도 쉽게 다양한 이미지를 생성하는 기능을 사용해 볼 수 있다. 그림 1은 ChatGPT(DALL · E3)에서 남녀 무사의 이미지를 각각 생성하고, 이 이미지들 속의 동일 인물들이 결투하는 모습을 생성하도록 대화창에서 요청했을 때의 결과이다.

2022년 7월 처음 공개되었던 Midjourney는 이 기술을 활용하여 제작되었던 ‘스페이스 오페라 극장’이라는 작품이 콜로라도 주 박람회 미술대회, 디지털 아트 부문에서 1위를 차지하며 주목받았고, 이후 많은 아티스트, 콘텐츠 창작자에게 활발히 이용되어 왔다. Midjourney는 기본적으로 $1,024 \times 1,024$ 픽셀 해상도의 이미지를 생성하고, 업스케일 기능을 통해 $2,048 \times 2,048$ 까지 확대 가능하다. 현재까지는 프롬프트를 간략하게 입력했을 때의 생성 결과(그림 2(a, c))보다는 권장하는 형식으로 상세히 입력했을



출처 Midjourney, Midjourney, Accessed: Feb. 1, 2025. <https://www.midjourney.com/home>

그림 2 프롬프트 정밀도에 따른 생성 결과 차이(Midjourney):
(a, c) 간략한 프롬프트 입력, (b, d) 동일 주제로
Midjourney가 권장하는 형식의 상세한 프롬프트 입력

때의 결과(그림 2(b, d)) 품질이 월등히 우수하다.

3. 동영상 생성 기술 동향

가. 비디오 잠재 확산 모델

초창기에 이미지들을 시간 축으로 연결하여 동영상을 구성하는 생성방식은 연속되는 프레임들 간의 관계를 신경망이 고려하기 어렵기 때문에 시간이 흐르면서 물체의 형상이 변하거나, Flickering 등의 아티팩트(Artifact)를 생성하는 취약점이 있었다. 엔비디아에서 발표한 VideoLDM[11]이나 Stability AI사의 Stable Video Diffusion(SVD)[12] 모델은 인코더-디코더로 이루어진 사전학습된 신경망 사이 사이에 Temporal Layer를 삽입하여 키프레임을 생성하고, 시간 축에 대한 Frame Interpolation을 수행하여 아티팩트를 완화하였다. 텐센트에서 공개한 VideoCrafter[13]는 아키텍처를 구성할 때 합성곱 레이어에 공간(Spatial) 트랜스포머와 시간(Temporal) 트랜스포머를 결합하여 고해상도 동영상 생성의 새로운 베이스라인을 제시하였다. AnimateDiff[14] 역시 사전훈련된 이미지 생성용 Stable Diffusion 모델에 추가 모듈을 부착하여 비디오로 확장하였다.

나. 물리적 난제

RunwayML의 Gen-3[15], OpenAI의 Sora[16]와 같은 최신 공개 서비스들은 Flickering 등의 아티팩트를 상당 부분 개선하였고, 엔비디아는 물리적 직관을 학습시키는 것에 중점을 둔 Cosmos[17]라는 새로운 파운데이션 모델을 공개하였다.

그림 3은 상용 서비스(Gen-3, Sora)와 오픈소스(Cosmos)를 이용하여 토끼가 한 바퀴 구르는 동작(Tumble)을 표현하도록 동일한 프롬프트를 입력했을 때의 생성 결과이다. Gen-3(RunwayML)에 비하여 Sora, Cosmos는 구르는 동작을 제대로 표현하려

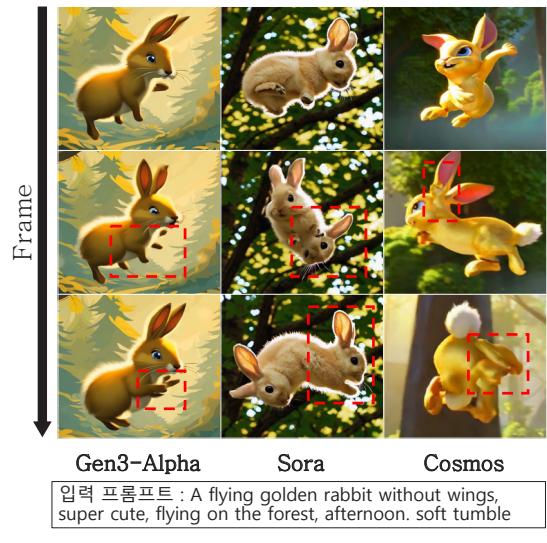


그림 3 공개된 서비스/모델의 동영상 생성 결과 오류 사례

고 시도하지만, 붉은 점선 영역에서 여전히 오류가 발생하고 있는 것을 확인하였다. Google DeepMind에서 개발한 Veo2[18] 모델은 이러한 동영상 생성의 물리적 난제에 초점을 맞춰 동물의 자연스러운 움직임과 물 표현과 같은 복잡한 물리적 현상에 강점을 갖는 것으로 알려져 있지만, 아직 일반인들을 대상으로 서비스가 공개되지는 않았다.

III. 콘텐츠 제작 기술 동향

앞서 설명한 LLM, 이미지/영상 생성형 AI 기술의 결과물들은 그 자체로서 활용될 수도 있겠지만, 이미 콘텐츠 제작 현업에 종사하고 있는 창작자들에게는 자동으로 결과물을 생성하기보다는 아이디어를 작품화하는 과정에서 오류를 검증하고 작업 효율성을 향상시키는 제작 지원 기술로서 더 실

용적일 수 있다. 그리고 기존 웹소설 IP(Intellectual Property)를 적은 비용으로 웹툰, 영상 콘텐츠로 변환 제작하여 추가 수익을 달성하는 방법도 좋은 활용 모델이다. ‘나 혼자만 레벨업’, ‘유미의 세포들’ 등은 웹소설 → 웹툰 → 애니메이션/드라마로 IP 전환이 이루어진 대표적인 사례이다.

1. 스토리 창작 지원 기술

가. 검색 증강 생성 기술(RAG) 활용

대규모 언어 모델(LLM) 분야에서 최신 정보 반영의 어려움과 사실관계에 대한 오류(환각, Hallucination)와 같은 한계를 극복하기 위해 등장한 기술이 검색 증강 생성(RAG: Retrieval Augmented Generation) 기술이다. RAG는 텍스트 생성 과정에서 외부 데이터를 검색하고 이를 사전 학습된 내용과 통합하여 더욱 정확하고 풍부한 콘텐츠를 생성함으로써 LLM의 단점을 보완하고 창작활동에 새로운 가능성을 제시하고 있다.

특정 주제에 대한 소설을 쓰고자 하는 창작자는 RAG를 활용하여 관련된 위키피디아 문서를 분석하고, 이를 바탕으로 LLM과 대화형으로 스토리 생성 과정을 진행할 수 있다. 예를 들어, 중세 유럽을 배경으로 한 소설을 작성할 때 RAG가 시대적 배경과 주요 사건을 검색하고 이를 스토리에 반영해 스토리의 맥락과 사실성을 높이는 것이다.

게임의 스토리 제작에서도 RAG는 중요한 도구로 사용되고 있다. RAG를 통해 특정 장소나 시대적 배경을 가진 게임의 스토리를 작성할 때 기존의 문헌과 자료를 검색하고 이를 바탕으로 세계관을 창작하는데, 창작자의 아이디어가 새로운 세계관에 부합하는지, 역사적/과학적 고증이나 오류검증에 도움을 줄 수 있다.

몇몇 작가와 개발자는 OpenAI의 API를 활용해

대화형 소설 창작 프로젝트를 시도하고 있다. 또한, Ubisoft와 같은 대형 게임사는 AI 기반 기술을 게임의 대사, 세계관, 스토리라인을 생성하는 작업에서 활용하고 있으며, Sudowrite[19]와 같은 AI 기반의 창작도구 플랫폼은 작가가 초안을 작성하고 부족한 부분을 외부 자료로 보완하도록 지원하는 과정에서 RAG 기술을 활용하고 있다.

미래에는 독자의 선호도에 맞춘 맞춤형 소설이나 대본이 보편화될 가능성이 높고, AI 기반 기술들은 창작자의 동반자로서 협업형 스토리 제작을 주도할 것으로 보인다. 그리고 이러한 기술이 영화, 게임, 교육 콘텐츠 등 다양한 분야에 활용되면서 창작 과정의 혁신이 가속화될 것으로 예상된다.

나. 스토리 구성 지원 기술

스토리 구성 기술은 현재 창작 중이거나 혹은 전환 중인 콘텐츠의 전체 타임라인과 공간구성을 파악하여 캐릭터 관계나 세계관 설정의 일관성을 잃지 않도록, 스토리의 설정이 충돌하는 부분을 자동으로 분석하고 대안을 제시하는 기술을 의미한다.

이를 위해 장편 스토리 맥락 이해를 위한 문학 데이터셋 구축, 이벤트 중심 분석, 감정 흐름 분석, 캐릭터 간 관계 추론 등의 연구가 학계에서 진행되었다. LitBank[20]는 문학 텍스트에서 구조적 정보를 추출하기 위해 설계된 데이터셋 구축 프로젝트로서, 고전 문학 100여 편에서 인물, 사건, 장소 등의 정보를 태깅하여 소설의 주요 이벤트와 캐릭터 간의 상호작용 분석에 활용이 가능한 최초의 데이터셋이다. 이 연구를 시작으로 AI에 기반하여 소설 텍스트로부터 스토리 맥락을 이해하려는 시도가 지속적으로 진행되고 있으며, 그 중 BookNLP[21]는 입력된 소설로부터 스토리 분석과 캐릭터 관계 도출에 중점을 둔 자연어 처리 툴로 2020년 GitHub에 공개되어 많은 연구자가 활용하고 있다. 이 외

에도 GPT에 기반한 문학 분석이나 소설 서사 이해(Fiction Narrative Understanding)와 관련한 개별 연구들이 학계에서 진행되고 있다. 다만, 이러한 연구들은 AI에 기반하여 소설과 인물의 맥락을 이해해보고자 하는 학계 수준의 연구이며, 제작자를 지원할 수 있을 만큼의 기술적 성숙도를 보이지 못하고 있다. 즉, 맥락을 이해하는데 중심을 두고 있어 스토리 내 설정 충돌을 인식하거나 이에 대한 해결책을 제시하기에는 한계가 있으며, 전문적인 제작자 지원을 위한 하나의 통합된 솔루션은 추가로 개발이 필요한 부분이다.

2. 웹툰 제작 지원 기술

가. 최적 컷 구성 추천 기술

기존 웹소설이나 소설을 웹툰으로 전환할 때, 또는 새로운 웹툰을 제작할 때 제작자가 가장 관심을 갖는 부분은 컷 구성이다. 그 이유는 웹툰의 장면 구성과 시각적 흐름이 몰입감에 큰 역할을 하기 때문이다. 하지만 웹툰 창작에 익숙한 제작자가 아닌 이상 컷 구성에 대한 통찰을 얻기가 어려우며, 이는 아무리 스토리라인이 흥미롭다고 하더라도 웹소설에서 웹툰으로 IP가 전환되면서 상업적 가치가 저하되는 문제가 되기도 한다.

이미 시장에는 AI에 기반하여 웹툰을 조금 더 쉽게 제작할 수 있도록 지원하는 다양한 플랫폼이 존재한다. AI Comic Factory[22], ComicsMaker.ai[23], AI Comic Generator[24]와 같은 웹툰 제작 지원 플랫폼들은 텍스트에 기반하여 웹툰을 제작하는 과정에서 컷 구성을 추천하며, 또한 포즈와 캐릭터 외형의 수정까지 지원하기도 한다.

하지만 이러한 플랫폼들은 전문 웹툰 제작자가 컷 구성이나 작화를 좀 더 쉽게 작업하도록 도움을 줄 수는 있으나, 스토리의 뉘앙스와 복합적인 맥락

을 이해하여 몰입감을 최대화하도록 어떻게 컷을 구성해야 하는지에 있어서는 한계가 있다. 이를 개선하기 위한 추가 기술 개발이 필요한 상황이다.

나. 웹툰 이미지 관련 연구 동향

웹툰 제작에서 생성형 AI를 활용하기 위해서는 단편적인 이미지 생성형 AI 기술만으로는 부족하다. AI가 웹툰 창작자에게 유의미하게 도움이 되려면 캐릭터와 화풍의 일관성, 스토리를 고려한 맥락 유지, 창작자의 의도에 따른 결과물 편집 등 다양한 기능을 제공할 수 있어야 한다.

바이트댄스(ByteDance)사가 2024년 NeurIPS에 발표한 스토리디퓨전(StoryDiffusion)[25]과 같은 확산 모델이 웹툰 제작 지원에 활용 가능한 대표적인 연구라고 할 수 있다. 이 기술은 피사체와 다양한 디테일이 포함된 이미지에 대해 일관성을 유지하는 것에 집중하고 있다. 바이트댄스사는 논문을 통해 셀프 어텐션을 계산하는 새로운 방식을 제안함으로써 텍스트 기반 스토리를 일관된 이미지 시퀀스인 만화(Comic) 형태로 표현 가능함을 실험을 통해 보여주었다.

알리바바와 화동사범대학교가 2024년 공개한 ArtAug[26]와 같은 기술 또한 웹툰 지원에 활용 가능하다. 해당 기술은 인간의 선호도를 학습하여 생성 이미지의 카메라 노출, 촬영 각도 변경, 분위기 효과 추가와 같은 다양한 미적 효과 수정을 지원한다. 이 기술의 장점으로는 기존 생성 모델 대비 추가적인 계산량 증가를 필요로 하지 않아 창작활동 비용에 추가 부담을 주지 않는 점을 들 수 있다.

앤틴그룹(Ant Group)이 발표한 MagicQuill과 같은 이미지 편집 기술도 웹툰 제작에 도움을 줄 수 있다. 이 기술은 멀티 모달 LLM을 활용한 상호작용을 통해 이미지의 자유로운 편집을 제공한다. 자연어 입력을 통해 원하는 형태로 이미지를 제어하는 기술

은 AI 활용이 낯선 웹툰 창작자도 손쉽게 활용 가능할 것으로 기대된다.

다. 웹툰 이미지 생성 관련 서비스

완성된 스토리를 실제 웹툰으로 작품화하기 위해 이미지를 제작하는 과정에서, 현재 창작들이 가장 많이 사용하는 툴은 일본 CELSYS사의 '클립 스튜디오 페인트'[27]이다. 세부적인 드로잉과 페인팅, 캐릭터 포즈 편집 등의 다양한 기능을 제공하지만, 이미지 생성형 AI기능은 작가들의 반발을 의식하여 탑재하려던 계획을 철회한 사례가 있다(2022년 12월).

국내 웹툰 제작에서 AI 기술 활용에 가장 적극적인 기업은 네이버 웹툰이다. 네이버 웹툰은 전체 직원 중 절반에 가까운 인원을 개발자로 구성하고, 별도의 AI 연구조직인 '웹툰 AI'를 운영하고 있다. 2021년 10월 베타 출시한 '웹툰 AI 페인터'는 채색 작업의 자동화율을 향상시켰고, 현재 개발 중인 '웹툰 크리에이티브 에디터'는 전경/배경을 분리하는 '누끼따기' 작업, 불필요한 물체를 지우는 작업을 자동화할 것으로 기대된다. 아직 웹툰 이미지 생성형 AI 기능은 지원하지 않지만, 특정 작가가 보유한 이미지만을 학습 대상으로 삼아 저작권 침해 없이 창작의 생산성, 다양성 측면에서 혁신을 달성하려는 계획[28]을 밝힌 바 있다.

그 밖에 투닝(Tooning), 망고보드 등의 국내 서비스는 비록 인기 웹툰 수준의 콘텐츠 표현력 수준은 아니지만, 일부 이미지 생성/편집 기능을 제공한다. 해외 사이트인 StoryDiffusion[32]은 장문의 텍스트를 입력으로 스토리를 따르는 연속적인 이미지, 동영상 등을 생성하는 기능을 제공한다. 하지만 직접 실험해 본 결과 인물의 얼굴, 복장의 일관성과 입력한 스토리 흐름과의 부합성 측면에서 여전히 개선 필요성이 있음을 확인했다(그림 4).

앞서 설명했던 범용적인 이미지 생성형 AI 서비스



입력 프롬프트 : 참조 이미지의 주인공이 숲 속 길을 혼자 걷고 있었다. 갑자기 나타나는 3명의 도적들이 금품을 요구하자, 주인공은 굴복하지 않고 칼을 뽑아 그들과 싸운다. 주인공이 승리하고 도적들은 멀어져서 용서를 빈다. (같은 내용을 영문으로 번역하여 입력)

출처 StoryDiffusion, ByteDance, Accessed: Feb. 1, 2025. <https://storydiffusion.com/>

그림 4 스토리→연속적인 이미지 생성(StoryDiffusion)

스를 웹툰 제작에 활용하는 방법도 가능하다. Midjourney[10]는 개인맞춤화(Personalize) 기능을 최근에 추가하여 사용자가 선호하는 이미지들을 선별하여 고유의 스타일을 구성하고, 생성 시 일관된 스타일을 유지하는 기능을 제공한다. 또한 Midjourney[10]는 참조 이미지의 스타일뿐만 아니라 콘텐츠 내용(인물, 객체 등)을 참조하여 새로운 이미지를 생성하는 기능을 제공한다. 하지만 인물의 얼굴, 복장 일관성을 유지하면서 새로운 상황으로 이미지를 생성하려고 시도할 때, 일관성 유지 측면에서 한계가 있음을 확인하였다(그림 5).

3. 단편 동영상 제작 지원 기술

동영상 콘텐츠 제작에 활용 가능한 상용 서비스



(1) 원본 이미지 프롬프트 : A Korean man in his early 40s, wearing glasses, a neatly-groomed face, and no beard. A relaxed, not sharp impression. He is wearing a white T-shirt, a black cardigan, and a beige ring. sitting at a desk, looking at a monitor and typing on a keyboard. Japanese anime style.

(2) 작업→식사, 프롬프트 최소변화 : A Korean man in his early 40s, wearing glasses, a neatly-groomed face, and no beard. A relaxed, not sharp impression. He is wearing a white T-shirt, a black cardigan, and a beige ring. sitting at the table eating breakfast. Japanese anime style.

(3) 작업→식사, 프롬프트 완전교체 : same man, same clothe, sitting at a round table, eating a grand banquet

출처 Midjourney, Midjourney, Accessed: Feb. 1, 2025. <https://www.midjourney.com/home>

그림 5 생성 이미지 일관성 유지 이슈(Midjourney)

는 RunwayML의 Gen-3(RunwayML)[15], OpenAI의 Sora[16], KLING AI[29], Hailuo AI[30] 등이 있으며, 특히 Sora는 “Washed Out”이라는 음원의 공식 뮤직비디오 제작에도 사용된 바 있다. 이미지 생성 서비스와 마찬가지로, 텍스트 입력으로부터 고품질의 생성 결과를 얻기 위해서는 상세한 프롬프트 입력이 필요로 하기 때문에, GPT-4 등의 LLM을 사용하여 해당 툴에서 요구하는 상세한 프롬프트를 획득하는 방법이 많이 사용된다.

고품질의 생성 영상을 얻는 또 다른 방법은 Midjourney 등을 활용하여 이미지를 먼저 생성하고, 동영상 생성 서비스에서 애니메이팅을 적용하는 방식이다. 이와 관련하여 RunwayML은 Camera Control, Act-One 기능을 제공한다. Camera Control 기능은 사용자가 입력한 참조 이미지에 회전, 줌인/줌아웃 등의 카메라 제어와 텍스트 프롬프트를 입력하여 동영상 시점을 제어하고 움직임을 부여하여 5초 또는 10초 영상을 생성한다(그림 6). Act-One은 입력 이미지의 인물이 또 다른 참조 동영상의 인물이 말하는 입 모양이나 표정을 모방하도록 합성하는 얼굴 애니메이팅 기술이다.

이러한 상용 서비스들은 1K 이상의 해상도를 지원하며, Flickering과 같은 문제는 대부분 해결한 추세이다. 하지만 반사되는 모습, 손가락과 같은 말단 부위 표현 등에서 아직 종종 오류가 발생한다. 그리고 얼굴뿐만 아니라 신체 전체를 반영하는 움직임



출처 Gen-3, RunwayML, Accessed: Feb. 1, 2025. <https://runwayml.com/research/introducing-gen-3-alpha>

그림 6 카메라 컨트롤 동영상 생성 결과(RunwayML)

합성에서는 생성 품질 및 인터페이스 편의성 측면에서 아직 개선이 필요하다. 또한, 이미지 생성 서비스의 경우와 마찬가지로, 생성 결과들 간의 일관성 유지 역시 이슈가 되고 있다. 따라서 아직까지는 기존 작품의 타이틀 영상, 홍보용 단편 콘텐츠 제작 등에서 활용 가능성이 높은 상황이다.

IV. 생성형 AI 기술의 안전성 이슈

근래에는 생성형 AI 기술의 실용성뿐만 아니라 안정성 이슈도 주목을 받고 있다. 생성형 AI가 자동으로 생성하는 텍스트, 이미지, 영상 콘텐츠 내에 사회문화적으로 바람직하지 않은 영향을 초래할 요소들이 포함될 수 있다는 우려이다. 우선 생성 결과물에 포함될 수 있는 선정성, 폭력성 요소가 문제될 수 있고, 이미 많은 생성형 AI 서비스들이 프롬프트 입력 단계에서 키워드 차단기능을 제공하고 있다. 하지만 이렇게 명확한 위해성 요소 외에도 성별, 인종에 대한 편견 역시 문제가 되는 것으로 보고되고 있다[33]. 예를 들어 빅테크 기업에서 일하는 근로자의 이미지를 생성할 경우, 백인 남성의 이미지를 자주 생성하는 편향성이 발생할 수 있고, 이러한 편향성은 채용 등의 과정에서 불공정 요소로 작용하며 불평등을 심화시킬 수 있다.

이러한 생성형 AI의 위해성 요소를 방지하기 위해 우선 생성모델 자체를 개선하려는 연구들이 진행되었다. MIT는 생성모델을 Fine-Tunning하면서 제외하고자 하는 특정 콘텐츠 요소(Nudity, 특정 Artistic Style 등)를 망각(Unlearning)하는 기술을 개발하였고[34], 테크니온 공과대학교, 노스이스턴대학교 등은 확산형 생성모델의 Cross-attention 조작만으로, 불필요한 암시적 가정(예를 들어, 간호사는 항상 여성으로 표현하는 편향성 등)을 수정하고 유해성을 완화하는 기술[35-38]을 개발하였다.

하지만 이처럼 생성모델 자체를 개선하는 방식은 해외 빅테크 기업의 생성형 AI 기능을 활용하여 다양한 응용 서비스를 제공하려는 국내 기업의 입장에서는 제약이 발생할 수 있다. 위해성, 편향성에 대한 기준은 해당 지역의 사회문화적 특성에 따라 차이가 발생할 수 있고, 그렇기에 생성모델 외부에서 보완할 수 있는 기술적인 해결책이 필요하기 때문이다. 이러한 취지에서 노스텍사스대학교는 프롬프트 입력 단계에서의 수정을 통해서, 생성모델의 편향성(성별, 인종)을 보완하는 연구[37]를 수행하였다. 또한, Meta는 텍스트를 출력하는 멀티모달 입력 프롬프트(이미지+텍스트)에서 위해성 요소를 판별하는 Llama-Guard3[38] 기술을 공개하였다.

하지만 이러한 기술적 접근과 별개로, 바람직하지 않은 사회문화적 편견의 기준을 어떻게 정립하는가에 대한 인문학적 연구 또한 필요하다. 예를 들어 특정 직군의 인종, 성별 분포를 현재 사회의 통계적인 데이터와 일치하도록 표현할지, 또는 이와 별개로 사회 윤리적인 차원에서 지향해야 할 상태로 표현할지와 같은 선택의 문제가 발생한다. 향후 생성형 AI 기술의 사회, 문화적 파급력이 증가할 것으로 예상되기에, 이러한 과제를 해결하기 위한 다학제적인 논의와 함께 사회적 합의를 도출하기 위한 노력이 필요하다.

V. 결론

근래 빅테크 기업들이 발표하는 생성형 AI 기술의 결과물들은 그 품질과 완성도 면에서 많은 이에게 경탄을 자아내고 있다. 이러한 생성형 AI 기술은 콘텐츠 제작에서의 스토리 구성, 이미지 제작 및 영상화 작업 등에서 다양하게 활용될 수 있다. 하지만 앞서 살펴본 바와 같이, 스토리 맥락을 따라가는 장편 콘텐츠 제작을 자동화하기에는 아직 제약이 존

재하며, 작업의 각 단계에서 개발된 기술을 선별적으로 활용하는 방식이 더 실용적인 상황이다. 또한 선정성, 편향성과 같은 위해적 요소를 최소화하기 위한 안전성 확보 노력 역시 필요하다.

그리고 웹툰과 같은 특정 콘텐츠 장르에서 생성형 AI 기술이 제대로 활용되려면, 해당 도메인의 학습데이터 구축이 필수적이다. 하지만 현업 작가들과 IP 보유기관들은 저작권 보호와 현업 종사자의 경쟁력 약화를 우려하여 학습데이터 제공에 비협조적인 경향이 있다. 이를 해결하려면 인공지능의 기반이 되는 학습데이터 작품을 만들기 위해서, 기존 작가들이 전문성을 습득하고 콘텐츠를 제작하는 과정에서 들여야 했던 노고에 대해서 정당하게 보상하는 기술적, 제도적 지원이 필요하다. 마치 특허 제도를 통한 기술 공개와 합리적인 수익 배분이 이루어지듯이, 생성형 AI가 창출하는 경제적 가치에 대해서 학습데이터를 제공한 이들이 정당한 수익 배분을 지속적으로 제공받는 방법이 한 가지 대안이 될 수 있을 것이다.

장기적인 관점에서 기술 발전에 따른 사회 구성의 변화는 불가피하다. 하지만 급속도로 진행 중인 인공지능의 발전 속도는, 기존의 사회 체계 안에서 성실하게 일해왔던 많은 이들에게 생업이 위협받을 수도 있다는 불안감을 일으키거나 새로운 사회문화적 문제를 유발할 수도 있다. 인공지능이 인간을 대체하기보다는 인간의 삶을 더 풍요롭게 하는 일에 기여하도록, 인공지능 생성 기술이 기존의 질서와 상생하며 균형 있게 발전하도록 이끌어가는 혁명한 대처가 필요한 시점이다.

약어 정리

| | |
|-----|------------------------------------|
| AGI | Artificial General Intelligence |
| GPT | Generative Pre-trained Transformer |
| LDM | Latent Diffusion Model |

| | |
|------|---------------------------------|
| LLM | Large Language Model |
| MLLM | Multimodal Large Language Model |
| RAG | Retrieval Augmented Generation |

참고문헌

- [1] 장준하, “인공지능 GPT가 시나리오를 썼다. 단편영화 개봉박두!” AI타임스, 2020. 10. 21. <https://www.aitimes.com/news/articleView.html?idxno=132976>
- [2] 이기석 외, “디지털 에셋 창작을 위한 생성형 AI 기술 동향 및 발전 전망,” 전자통신동향분석 제39권 제2호, 2024. 4. pp. 32-42.
- [3] <https://www.llama.com/>
- [4] R. Rombach et al., “High-Resolution Image Synthesis with Latent Diffusion Models,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (New Orleans, LA, USA), Jun. 2022, pp. 10684–10695.
- [5] R. Gal et al., “An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion,” arXiv preprint, 2022. doi: 10.48550/arXiv.2208.01618
- [6] L. Zhang et al., “Adding Conditional Control to Text-to-Image Diffusion Models,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (Vancouver, Canada), Jun. 2023, pp. 3836–3847.
- [7] Z. Duan et al., “ArtAug: Enhancing Text-to-Image Generation through Synthesis–Understanding Interaction,” arXiv preprint, 2024. doi: 10.48550/arXiv.2412.12888
- [8] H. Zhang et al., “EliGen: Entity-Level Controlled Image Generation with Regional Attention,” arXiv preprint, 2025. doi: 10.48550/arXiv.2501.01097
- [9] <https://openai.com/index/dall-e-3/>
- [10] <https://www.midjourney.com/>
- [11] A. Blattmann et al., “Align your Latents: High-Resolution Video Synthesis with Latent Diffusion Models,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (Vancouver, Canada), Jun. 2023, pp. 22563–22575.
- [12] A. Blattmann et al., “Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets,” arXiv preprint, 2023. doi: 10.48550/arXiv.2311.15127
- [13] H. Chen et al., “VideoCrafter1: Open Diffusion Models for High-Quality Video Generation,” arXiv preprint, 2023. doi: 10.48550/arXiv.2310.19512
- [14] Y. Guo et al., “AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning,” in Proc. Int. Conf. Learn. Representations., (Vienna, Austria), May. 2024.
- [15] <https://app.runwayml.com/>
- [16] <https://openai.com/sora/>
- [17] <https://github.com/NVIDIA/Cosmos/>
- [18] <https://deepmind.google/technologies/veo/veo-2/>
- [19] <https://www.sudowrite.com/>
- [20] D. Bamman et al., “An Annotated Dataset of Literary Entities,” in Proc. North Amer. Chapter Assoc. Comput. Linguist., (Minneapolis, MN, USA), Jun. 2019, pp. 2138–2144.
- [21] <https://github.com/booknlp/booknlp>
- [22] <https://aicomicfactory.com/>
- [23] <https://www.comicsmaker.ai/>
- [24] <https://comic-ai.ai/>
- [25] Y. Atzmon et al., “Edify Image: High-Quality Image Generation with Pixel Space Laplacian Diffusion Models,” arXiv preprint, 2024. doi: 10.48550/arXiv.2411.07126
- [26] Y. Zhou et al., “StoryDiffusion: Consistent Self-Attention for Long-Range Image and Video Generation,” arXiv preprint, 2024. doi: 10.48550/arXiv.2405.01434
- [27] <https://www.clipstudio.net>
- [28] 한광범, “[단독] ‘내 그림만 학습자료로’…네이버웹툰, 작가별 AI툴 구축한다,” 이데일리, 2023. 12. 10. <https://www.edaily.co.kr/News/Read?newsId=01521926635837208&mediaCodeNo=257>
- [29] <https://klingai.com/>
- [30] <https://hailuoi.video/>
- [31] <https://openai.com/index/hello-gpt-4o/>
- [32] <https://storydiffusion.com>
- [33] 김현진, “생성형AI와 편향성,” IP & Data 法, 제3권 제1호, 2023. pp. 75–105.
- [34] R. Gandikota et al., “Erasing Concepts from Diffusion Models,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., (Paris, France), Oct. 2023, pp. 2426–2436.
- [35] H. Orgad et al., “Editing Implicit Assumptions in Text-to-Image Diffusion Models,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., (Paris, France), Oct. 2023, pp. 7030–7038.
- [36] R. Gandikota et al., “Unified Concept Editing in Diffusion Models,” in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis., (Waikoloa, HI, USA), Jan. 2024, pp. 5099–5108.
- [37] C. Clemmer et al., “PreciseDebias: An Automatic Prompt Engineering Approach for Generative AI to Mitigate Image Demographic Biases,” in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis., (Waikoloa, HI, USA), Jan. 2024, pp. 8581–8590.
- [38] <https://huggingface.co/meta-llama/Llama-Guard-3-11B-Vision>