디지털 에셋 창작을 위한 생성형 AI 기술 동향 및 발전 전망

Generative AI Technology Trends and Development Prospects for Digital Asset Creation

이기석 (K.S. Lee, mvr_lks@etri.re.kr) XR콘텐츠연구실 책임연구원/실장
이승욱 (S.W. Lee, tajinet@etri.re.kr) 공간콘텐츠연구실 책임연구원
윤민성 (M.S. Yoon, msyoon@etri.re.kr) XR콘텐츠연구실 책임연구원
유정재 (J.J. Yu, jungjae@etri.re.kr) 지능형콘텐츠인식연구실 책임연구원
오아름 (A.R. Oh, aro1116@etri.re.kr) 콘텐츠융합연구실 선임연구원
최인문 (I.M. Choi, inmchoi@etri.re.kr) 감성디지털휴먼연구실 연구원
김대욱 (D.W. Kim, dooroomie@etri.re.kr) 실감상호작용연구실 연구원

ABSTRACT

With the recent rapid development of artificial intelligence (AI) technology, its use is gradually expanding to include creative areas and building new content using generative AI solutions, reaching beyond existing data analysis and reasoning applications. Content creation using generative AI faces challenges owing to technical limitations and other aspects such as copyright compliance. Nevertheless, generative AI may increase the productivity of experts and overcome barriers to creative work by allowing users to easily express their ideas as digital content. Thus, various types of applications will continue to emerge. As images and videos can be created using text input on a prompt, generative AI allows to create and edit digital assets quickly. We present trends in generative AI technology for images, videos, three-dimensional (3D) assets and scenes, digital humans, interactive content, and interfaces. In addition, the prospects for future technological development in this field are discussed.

KEYWORDS 경량 딥러닝, 디지털 에셋 창작, 메타버스, 모델 압축, 생성형 AI

Ⅰ. 서론

우리는 글쓰기와 말하기, 그림 그리기, 프로그래밍 등의 창작은 인간만이 가능한 영역이라 생각해 왔으나, 기술 발전으로 생성형 AI가 콘텐츠 창작 영역을 넘보는 시대가 도래하게 되었다. 생성형 인공

* DOI: https://doi.org/10.22648/ETRI.2024.J.390204
* 본 연구는 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임[No. RS-2023-00225441, 디지털 에셋의 다형성 변형을 위한 지식, 정보 구조화 기술].

본 저작물은 공공누리 제4유형
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.
©2024 한국전자통신연구원

지능 모델은 텍스트, 이미지, 비디오, 3차원 모델, 코딩, 음향 등에 빠르게 적용되기 시작하였으며, 사회적·산업적 지각 변동을 일으키고 있음을 알 수 있다. 보편적인 수단인 텍스트, 음성 등에 기반하는 대화형 상호작용을 활용하는 에셋 창작 기술은 일반인도 수준 높은 품질의 결과물을 만들 수 있게 빠르게 발전하고 있다. 콘텐츠 개발의 기본 재료인 디지털 에셋 창작에 있어 생성형 AI를 활용한 생산성 증가는 게임, 영화 등의 기존 산업과 더불어 메타버스 산업 성장에 기폭제가 될 것으로 예상된다. 이에 본고에서는 생성형 AI를 활용한 디지털 에셋 창작에 대한 동향과 앞으로의 발전 전망을 소개하고자 한다.

II. AI 기반 디지털 에셋 창작 기술 동향

대표적인 생성형 AI 서비스인 GPT와 함께 DALL-E, Midjourney, 국내 칼로(카카오브레인) 등의 이미지 생성AI 서비스도 이제 일반인들에게조차 익숙한 용어가 되었다. 이 장에서는 다양한 서비스 소개와 공개된 자료를 바탕으로 기술적인 발전 동향에 대해서 설명하고자 한다.

1. 이미지, 동영상 생성

이미지, 동영상에 대한 다양한 생성 모델이 경쟁적으로 공개되고 있으며, 이미지의 품질과 속도를 높이고 편집을 용이하게 하는 등의 지속적인 기술적 발전과 더불어 기존 포토샵처럼 이미지 편집이 되는 도구도 상용화되고 있다.

가. 이미지 생성기술 동향

GAN(Generative Adversarial Network)을 중심으로 발전하는 이미지 생성AI 연구는 근래 확산모델(Diffusion Model) 중심으로 전환되기 시작했다. 확산모델은 개념적으로 입력 이미지의 노이즈를 줄이는 화질개선 작업을 반복하는 방식으로, 완전 잡음형태의 입력으로부터 선명한 화질의 이미지를 생성하려는 접근 방법이다. LDM(Latent Diffusion Model, 뮌헨대학교, 2022)[1]은 이러한 방식을 이미지 공간이 아닌 압축된 잠재적 공간(Latent Space)에서 U-Net구조 네트워크의 잡음제거(Denoising Step)를 반복하는 방식을 사용한다. 스테이블 디퓨젼(SD: Stable Diffusion)으로도 알려진 이 기술은 DALL-E, Midjourney 등이 서비스만 공개한 것과 대조적으로, 코드와 학습 모델을 공개했고 상업적인 활용이 가능했기에 이후 이미지 생성AI 기술이 빠르게 발전하는 기폭제가 되었다. ControlNet(스탠포드, 2023)은 사람의 포즈 등 다양한 추가 정보를 입력하여 생성 이미지를 제어하는 방법을 제공한다. 이 기술은 대용량의 이미지를 사전 학습한 원본 SD 모델은 고정(Freeze)한 상태에서, 복사된 별도의 SD 학습을 거치면서 디코더의 중간 매개값을 조정(합산)하면서 출력을 만들어 가는 기법인 하이퍼네트워크 방식을 활용한다.

나. 동영상 생성기술 동향

동영상 생성은 이미지 생성에 비해서는 한 걸음 늦게 발전되고 있는 상황이다. 2022년, Imagen Video(구글), Make a Video(메타) 등 빅테크 기업들이 텍스트를 입력하여 5초 길이의 동영상을 출력한 결과와 서비스를 공개하였다. 하지만 선별된 공개 결과와 달리 실제 서비스 테스트 결과에서는 Flickering 등의 문제를 발견할 수 있다. 2023년 3월 Runway(미국)가 공개한 Gen-2는 텍스트와 이미지를 입력하여 4초 분량의 동영상을 생성하였고, 같은 해 7월에 공개된 업데이트 버전은 출력 결과가 제한적인 움직임이지만 실사 영상에 가까운 느낌으로 발전하였다. 동영상 생성을 선도하는 기관은 Runway(미국)이지만 이후로 라이벌들이 계속 등장하고 있다. 논

문 발표 형식으로 공개된 기술 중 현재 가장 우수한 품질을 보여주는 것은 VideoLDM(뮌헨대, NVIDIA 등, 2023)[2]이다. 기존 이미지 LDM 모델을 활용하면서 반복적인 Frame interpolation, Temporal layer를 활용하고 1280×2048 해상도의 동영상을 출력한다. 저자는 작년 11월, SVD(Stable Video Diffusion, Stability AI)라는 이름으로 코드와 모델 파라미터를 공개하였다. 576×1024 해상도, 25프레임을 생성하는 이 생성모델은 자체 실시한 품질평가에서 Runway Gen-2를 능가하였다. SVD는 현재 상업적인 활용은 제한되고, 학술적인 목적으로 자유롭게 활용 가능하다.

2. 3D 모델, 장면 생성

이미지 생성 모델의 폭발적 증가에 영감을 받아 텍스트나 이미지로부터 수초 이내에 3D 모델을 생성하는 기술이 등장하고 있다.

가. 3D 모델 생성기술 동향

Nvidia는 가상 3D 세계 및 메타버스 구축에 사용될 3D 객체를 생성하기 위한 인공지능 기술을 적용한 도구로써 캐릭터, 건물, 차량 및 기타 유형의 3D 개체를 생성하는 GET3D를 공개한 바 있으며, Generative AI(GAN) 기술을 이용하여 2차원 이미지를 이용하여 3차원 객체를 생성하는 것으로, 단일 2D 이미지에 GAN을 적용하여 해당 물체 주변을 찍은 듯한 여러 이미지를 생성하여 이를 바탕으로 3D 객체 모델을 재현하는 GANverse3D 기술을 공개하였다[3].

OpenAI는 2023년에 3D 모델을 생성하는 오픈소스인 쉐이프-E(Shap-E)를 공개하였다. 이 AI 도구는 텍스트 또는 이미지를 입력하면 자동으로 신속하게 3D 모델을 생성하는 기능을 제공한다. 이 도구를 활용하면 단일 GPU에서 텍스트 또는 이미지 입력에 대해 1분 이내로 3D 모델을 생성할 수 있기 때문에, 신제품 개발이나 제조 공정 개선을 위한 부품 및 조립품의 3D 모델을 효율적으로 생성하는 데 활용될 수 있다. 이 쉐이프-E는 3D 개체의 질감에서 세밀한 표현을 효과적으로 잡아내지 못하는 부분, 그리고 이 개체의 여러 속성 결합 및 개체들의 정확한 개수를 생성하는 부분 등에서 기술적으로 아직 부족한 측면이 나타나고 있다. 그러나, 향후 대규모 3D 데이터 세트로 훈련되면서 이러한 제약이 극복되고 다른 3D 생성 기술들과 조합하게 되면, 이 쉐이프-E 도구를 통해 고정밀도의 3D 모델을 신속하게 생성할 수 있을 것으로 예상된다.

나. 3D 장면 생성기술 동향

3D 장면을 360도 돌려가면서 다양한 각도에서 관측하려면, 2D 이미지 대신 3D 렌더링이 필요하다. 최근에 엔터테인먼트, VR/AR, 광고, 게임 등과 같은 다양한 분야에서 몰입형 3D 장면에 대한 수요가 급증하고 있다. 종래의 잘 알려진 3D 장면 생성 기술들 중의 하나로서, Nvidia는 여러 각도에서 촬영된 2D 사진들로부터 사실적인 디지털 3D 장면으로 바꾸어 주는 초고속 신경망 및 렌더링을 이용한 Instant NeRF(Neural Radiance Fields) 기술을 2022년에 공개한 바 있다. 이 신경망 기반의 컴퓨터 비전 기술은 불완전한 이미지 데이터로부터 실제 물체의 크기와 모양을 이해하기 위한 공간 탐사 및 로봇학·자율 차량의 훈련 분야에 응용될 수 있을 뿐만 아니라 장애물에 의해 개체가 가려지는 폐색(Occlusion) 문제도 해결할 수 있다. CVPR 2023에서 인텔(Intel)은 텍스트 프롬프트에서 깊이 정보를 갖는 360도 3D 이미지 제작용 생성형 AI 모델 LDM3D(Latent Diffusion Model for 3D)를 발표했다[4]. 블로케이드 랩스와의 협업을 통해 개발된 이 모델은 기존의 잠재안정확산모델(LDM)과 동일한 수의 파라미터들로도

텍스트 입력 기반 이미지 및 깊이맵의 생성이 가능하며, 사실감 및 몰입감을 제공하는 360도의 3D 공간 파노라마 콘텐츠로 전환할 수 있다. 또한, 2023년 SIGGRAPH에서 Nvidia는 생성형 3D 장면 아티스트 도구용 클라우드 기반의 Nvidia 피카소(Picasso) 생성형 AI 서비스를 발표했다. 즉, 셔터스톡(Shutterstock)은 Nvidia와의 긴밀한 협력을 통해 3D 장면용 생성형 AI를 도입하여 몰입형 3D 장면 배경을 빠르게 생성 및 커스터마이징할 수 있는 360도 HDRi 서비스를 구축하였다. 이 서비스는 프롬프트에 간단한 텍스트 또는 참조 이미지 등을 입력하여 아티스트가 Nvidia 피카소 툴에서 파노라마 이미지를 빠르게 생성할 수 있게 한다. 특히, 이 고성능 3D 콘텐츠 제작 기술은 디지털 트윈이나 엔터테인먼트 및 게임을 위한 3D 장면의 배경 설정, 그림자 조절, 밝기 조절 등을 할 수 있는 8K급 해상도의 사실적인 360도 환경맵 생성을 지원해 준다.

3. 디지털 휴먼 창작

타인과의 교류에 익숙한 우리에게 가장 선호하는 형태인 디지털 휴먼을 활용하려는 수요는 다양하게 존재하며 보다 사실적인 표현과 상호작용이 가능하도록 발전하고 있다.

가. 디지털 휴먼 언어생성 모델

최근에 등장한 언어 모델들은 인간이 구별하기 어려울 정도로 고도화되었으며, 특히 ChatGPT4.0은 뛰어난 추론 능력을 선보이고 있다. 구글은 트랜스포머 알고리즘으로 언어 모델의 새로운 패러다임을 제시하였다. 트랜스포머는 자기 집중(Self-attention)을 통해 입력 시퀀스의 모든 원소 간 상호작용을 반영하고, 위치정보 인코딩을 위해 positional encoding을 사용하며, 여러 층으로 적층한다[5]. 멀티-헤드 어텐션을 도입하여 다양한 특징을 동시에 학습한다. ChatGPT는 기본적으로 이러한 트랜스포머 구조와 RLHF(Reinforcement Learning from Human Feedback)를 활용하여 자연스러운 문장 생성이 가능하게 설계되었다. 이로써 디지털 휴먼이 인간과 같은 대화능력을 갖게 되었다. 더욱이 최근에는 파인튜닝 혹은 프롬프트 확장과 같은 방법으로 인격을 주입하게 되어, 사용자는 자신만의 디지털 휴먼을 만들 수도 있게 되었다.

나. 디지털 휴먼 동작 생성 모델

MotionGPT[6]는 프롬프트 입력을 근거로 인간의 동작 정보를 도출하는 기술을 지칭한다. 이 기술은 다양한 방식으로 구현되며, 그중 하나가 Text-to-Motion(Motion Generation)으로, 이는 문장 입력으로 인간의 동작을 창출하는 기법이다. 다른 접근 방식으로서 Motion-to-Text(Motion Translation)는 동작 영상을 분석하여 그에 해당하는 문자 정보로 전환하는 과정을 말한다. 마지막으로, Motion Complete는 주어진 동작 영상의 후반 부분을 창출하고 완성하는 기술로 미완성된 동작 데이터를 바탕으로 동작의 나머지 부분을 생성하는 데 사용된다. 이렇게 MotionGPT는 인간의 동작을 이해하고 생성하는 데 다양한 접근 방식을 제공함으로써 동작의 인식 및 생성 분야에서 발전을 도모하고 있다.

4. 인터랙티브 콘텐츠 생성

사용자가 콘텐츠 체험에 있어 생성형 AI와 서로 상호작용하며 소통하는 기술은 특히 게임 콘텐츠를 중심으로 다양한 형태로 접목되고 있다. 크게 사용자가 텍스트 기반으로 직접 생성형 AI와 대화하면서 콘텐츠를 생성하는 인터랙티브 스토리텔링 분야와 콘텐츠를 무작위적이고 무한히 생성하는 PCG(Pro-

cedural Content Generation) 분야 두 가지가 활발하게 연구되고 있다. PCG 분야는 생성된 콘텐츠가 에러나 버그 없이 정상 작동하는 검증 작업이 중요한데, 딥러닝과 같은 기계학습과 생성형 AI가 결합하여 기존의 단점을 극복하는 기술들이 연구되고 있다.

가. 인터랙티브 스토리텔링

스탠포드대학은 NPC들이 작은 가상의 마을 안에서 스스로 다른 NPC들과 상호작용하며 새로운 스토리를 만들어나갈 수 있도록 시뮬레이션하였다[7]. 가상 마을 속 NPC들은 LLM과 기억 모듈을 통해 스스로 판단/계획/행동할 수 있었으며, 인간 사회와 유사하게 각자 일하러 가고, 점심을 먹고, 다른 에이전트들과 만나 대화를 수행했다.

AI Dungeon[8]은 사용자의 선택, 행동, 또는 다양한 외부 요인들에 따라 환경, NPC 행동과 반응, 스토리의 흐름과 결말 등이 실시간으로 변화하는 콘텐츠를 제공하고 있다.

중국 게임사 넷이즈[9]는 상용 MMORPG '역수한(逆水寒)'이라는 게임에 생성형 AI를 도입하여, 플레이어가 채팅창에 입력한 말에 따라 게임 캐릭터가 반응하며 행동이 바뀌는 기술을 개발하였다.

나. 게임 콘텐츠 생성

Bitmagic Game[10]은 실시간으로 오픈 월드에서 유저가 채팅창에 명령을 입력하면 그것에 맞게 게임 오브젝트, 몬스터, 환경, 플레이 메커니즘 등을 추가하여 주는 기술을 공개하였다.

기계학습이 적용된 PCG의 초기에는 LSTM과 같은 방법[11]이 사용되었다. 어떤 게임 맵의 상태로부터 자연스럽게 다음 맵을 생성할 수 있도록 유도하는 방식이었다. 이후 생성형 AI로 대표되는 GAN 알고리즘의 등장으로 이를 접목하여 게임 맵을 생성[12]하기 시작하였고, 콘텐츠의 생성 속도도 빨라지고 더 다양한 맵을 생성할 수 있게 되었다. 이후 Conditional VAE와 같은 알고리즘을 통해 실제로 게임 개발자가 게임 콘텐츠 생성에서 상호작용을 할 수 있게 되었다[13].

생성형 AI를 학습시키려면 반드시 인간이 사전에 만들어준 정답 데이터가 필요한데, 게임 콘텐츠의 특성상 다량의 인간 정답 데이터를 구하기가 쉽지 않고, 추후 저작권 문제가 발생할 수 있다. 따라서 최근에는 게임 콘텐츠를 생성하는 데 필요한 인간 정답 데이터 수를 줄이는 방향으로 PCGRL[14]과 같은 기술들이 연구되고 있다. 강화학습 알고리즘을 통해 인공지능은 스스로 게임 맵을 탐색하면서 새로운 콘텐츠를 만들어보며, 인간이 정해준 보상함수를 따라 다양한 콘텐츠를 오류 없이 만들 수 있도록 학습된다. 마찬가지로 PCGRL 또한 생성될 콘텐츠의 방향을 인터랙티브하게 조절할 수 있도록 Controllable PCGRL[15]이 연구되었으며, 3차원 게임에 확장/적용하는 연구[16]가 진행되고 있다.

5. 콘텐츠 인터페이스 개발

생성형 AI와 결합된 인터페이스는 사용자의 상황을 이해하고 자연어를 활용한 직관적인 방식의 소통과 서비스 시나리오를 이해하는 중재자로서 사용자의 편리성을 극대화할 수 있다.

가. 멀티모달 XR 인터페이스 기술 발전 전망

멀티모달 XR 인터페이스는 음성, 텍스트, 이미지 등 다양한 입력 방식을 통합하여 상호작용을 제공한다. 생성형 AI의 발전은 기존 하드웨어를 통해 더 효율적인 결과물을 만들 수 있게 하며, 새로운 인터페이스와 기술을 통한 결과물은 더욱 효과적으로 제작될 수 있다. 다양한 입력 방식을 복합적으로 활용하는 인터페이스의 진화는 생성형 AI를 활용하고

새로운 결과물을 창출하기 위해 두 개 이상의 입력을 결합하여 보다 편리하게 콘텐츠를 체험하고 새로운 창작물을 만들어내는 데 도움을 줄 수 있다.

나. 사용자 친화형 인터페이스 기술 동향

사용자 콘텐츠의 소비방식이 변화함에 따라 개인의 행동 패턴과 관심사를 분석하는 생성형 AI 기술의 도입으로 개인화된 사용자 인터페이스를 제공하는 서비스가 시도되고 있다.

네이버는 생성형 인공지능 검색 환경에 최적화될 수 있는 UX(User Experience)와 UI(User Interface)를 개편하여 전 사용자를 대상으로 확대 적용한 바 있다. 콘텐츠의 특성에 따라 제공하는 정보의 종류와 표시할 정보량, 순서 같은 데이터 구성의 최적화를 실시하여 사용자 친화형 인터페이스를 제공할 수 있다. 예를 들면 '패션', '맛집'과 같은 주제의 결과는 이미지를 강조하는 구조로 개선하고 '경제', '비즈니스' 분야의 정보는 텍스트를 우선하는 미리보기를 제공하여 핵심 내용 파악 후 문서를 선택할 수 있도록 개선하여 서비스 최적화를 제공한다.

이처럼 사용자의 개인화된 특정 사용 패턴을 분석하고 적합한 최적의 정보를 제시하고 이해하는 방식의 인터페이스에 생성형 AI가 다각도로 접근하고 있음을 알 수 있다.

Ⅲ. 에셋 창작을 위한 AI 발전 전망

이 장에서는 각 활용 분야에 생성형 AI를 활용하는 기술적 한계와 이를 해결하기 위한 발전 전망을 제시하고자 한다.

1. 이미지, 동영상 생성

이미지 생성AI 기술은 이제 생성 결과의 품질 측면에서는 실제 사진, 사람이 그린 그림과 흡사한 수준에 이르렀다. 이제는 오히려 생성AI를 통하여 만들어진 가짜 이미지를 정확하게 구별할 수 있는 기술이 연구되는 상황이다.

하지만 기술적인 측면에서 생성되는 이미지가 사용자의 의도를 정확히 반영하고 있는가를 생각하면, 여전히 개선이 필요한 부분이 있다. 현재 확산 모델에 기반한 생성기술은 학습 데이터 분포에 의존하기 때문에 사용자가 입력하는 요구정보(프롬프트, 각종 Condition)가 학습된 확률분포를 벗어날 경우, 의도하지 않은 이미지가 생성되는 현상이 발생한다. 또한 생성과정에서의 무작위성으로 인하여, 인물이나 장면 구성의 일관성을 유지하면서 스토리 흐름에 맞는 연속적인 이미지를 생성하는 것은 여전히 진행 중인 연구 주제이다. 그 밖에 이미지 생성 모델을 학습하기 위해 사용된 대용량 이미지 데이터의 저작권 문제가 이슈화되고 있으며, 생성AI를 이용하여 발생하는 경제적 이익을 학습 데이터의 저작권을 소유한 이들에게 분배하는 기술적 해결책 등이 연구되고 있다.

이미지에 비하여 동영상 분야에서의 생성AI는 해결해야 할 기술적 난제들이 많은 상황이다. 기존의 Flickering 문제는 비교적 해결되어 가는 추세이지만, 움직임이 큰 동작을 요구할 경우 의도하지 않은 비현실적인 변화가 발생하는 문제는 여전히 해결해야 할 과제이다. 현재 가장 앞서가는 기술 중 하나인 Runway의 Gen-2 역시, 걸어가는 사람을 생성하면 앞모습이 뒷모습으로 바뀌는 등의 변화가 발생한다.

이론적으로 고차원 데이터를 생성하는 모델을 학습하려면 더 많은 개수의 학습 데이터가 필요하다. 이미지를 생성하는 DALL-E가 4억 장의 이미지-텍스트 쌍 데이터를 학습한 것을 생각하면, 100프레임 이상의 고해상도 동영상을 생성하는 모델을 완전 시작부터 학습하려면 수백억 개 이상의 동영상-텍스트 쌍 데이터 학습이 필요하다. 그렇기에 기술

이 공개된 SVD(Stable Video Diffusion)의 생성모델 학습 과정을 살펴보면, 이미지 생성 기학습 모델(Pre-trained Model)을 기반으로, 상대적으로 소량의 동영상 학습데이터를 추가로 학습하는 방식을 사용하였다. 그렇기에 약간 움직이는 이미지 느낌의 동영상은 비교적 양호하게 생성하지만, 모션의 강도를 높게 설정하고 사용자의 요구(참조 이미지, 프롬프트 입력)를 입력하면 앞에서 설명한 문제가 발생하는 경우가 빈번하다. 이러한 문제가 해결되려면, SD 모델이 이미지 생성 분야의 비약적인 발전을 가져왔듯이, 동영상 생성 분야에 적합한 새로운 혁신적인 방법(Breakthrough)이 등장해야 할 것으로 예상된다.

2. 3D 모델, 장면 생성

대규모 디지털 3D 에셋의 학습을 통해 새로운 3D 모델을 생성하는 연구가 다양하게 진행되고 있다. 디지털 3D 에셋의 지식화를 통해 재활용도를 높이기 위한 생성형 AI 기반의 3D 모델 생성 기술의 향후 지향점은 텍스트, 이미지, 그리고 오디오 데이터 등을 단일한 모델로 통합하는 멀티모달(Multimodal) 타입의 AI 서비스를 구현하는 것이다. 이 멀티모달 AI 기술은 다양한 종류의 입력 데이터들을 동시에 처리할 수 있기 때문에, 3D 모델의 디자인 관련 다양한 상황 인식 애플리케이션에서 효율적인 작업 수행을 가능케 한다. 이 멀티모달 타입의 차세대 AI 기술은 3D 개체 및 배경, 공간 컴퓨팅 영상 등의 콘텐츠를 생성하는 데 사용될 뿐만 아니라, 디지털 트윈, MR/XR, 메타버스와 같은 360도 초실감 복잡계의 시뮬레이션에도 적용될 수 있을 것으로 전망된다. 예를 들어 현재 카메라로 사물을 360도로 촬영하면, 이 촬영된 사진이나 동영상 데이터를 입력으로 사용하여 사물을 자동으로 3D 모델로 변환할 수 있게 발전되었다. 이러한 3D 모델 생성을 위한 혁신적인 생성형 AI 기술은 디자이너와 개발자들이 고가의 스캔 장비와 복잡한 작업을 필요로 했던 종래의 과정들을 효율적으로 자동화하거나 대체하고 있으며, 나아가 360도의 3D 콘텐츠 제작을 편리하게 수행할 수 있게 해주고 있다. 그러나, 현재까지 개발된 이러한 기술들은 금속과 유리 또는 금속 재질에서 광의 투과나 반사, 재질의 표면 특성과 같이 개체의 다양한 물성 조건들에서도 실제감 표현이 가능하도록 3D 모델링의 성능을 고도화할 필요가 있다.

이와 더불어, 생성형 AI를 활용하여 3D 장면을 생성하고 렌더링하기 위해서 각 공간 지점에서 색상 및 밝기 정보를 학습하여, 고해상도의 시각적 실감 효과를 생성할 필요가 있다. 이를 통하여 현실 세계 또는 가상 세계의 3D 장면을 정확하게 모델링하고, 사용자가 새로운 시점에서 해당 장면을 관찰할 수 있게 하여 준다. 이 기술은 다양한 시뮬레이션 및 렌더링 응용 분야에서 혁신적인 역할을 하며, 실제 세계의 장면을 더 현실적으로 표현하고 모델링하는 데에 기여함으로써 큰 주목을 받고 있다.

현재 NeRF와 관련된 연구 분야에서는 이전보다 현격한 진보가 이뤄지고 있다. 초기에는 느린 학습 과정과 렌더링 속도, 그리고 다양한 카메라 파라미터 요구사항 등의 여러 가지 제약 사항으로 인해 한계가 있었지만, 최근 연구에서는 이러한 제한을 극복하기 위한 혁신적인 모델들이 나와 있다. Instant NeRF와 같은 연구들과 특히 다양한 환경과 움직이는 물체에 적용 가능한 구글의 'NeRF in the Wild' 논문은 실시간 학습 및 렌더링 기능을 향상시키는 기술을 제시하고 있으며, 나아가 데이터 요구량을 줄이고 객체 일반화를 높이기 위한 방법들도 적용되고 있다[17].

또한, 다양한 딥러닝 기법들과 영상학적 기술을 통합하여 모델의 성능을 향상시키는 연구들도 활발히 진행되고 있다. 이러한 발전은 NeRF의 활용 가

능성을 확장하며, 현재의 3D 장면 생성용 AI 연구 및 관련 응용 분야에서 더 다양하고 효과적인 활용을 가능케 할 것으로 기대된다.

Nvidia 피카소는 텍스트 설명을 기반으로 이미지, 비디오, 그리고 3D 장면의 생성을 지원하는 클라우드 기반 생성형 AI 서비스를 제공하고 있다. 이 서비스는 Nvidia의 클라우드 환경을 활용하여 구축되었으며, 방대한 데이터로 학습된 AI 모델을 활용하여 사용자의 요청을 신속하고 정확하게 처리하도록 개선되고 있다. 피카소 서비스는 크리에이티브 디자인, 가상 현실 설계, 실감 콘텐츠 제작, 게임 개발 등 다양한 용도로 활용되고 있는 중이다.

나아가 Nvidia의 협력사인 셔터스톡은 3D 장면용 생성형 AI를 구축하여 8K 해상도의 360도 3D 장면에 대한 HDRi 서비스를 제공하고 있는 상황이다. 따라서, 3D 장면용 생성형 AI 기술 분야의 지속적인 향상은 초실감 공간 영상 재현과 관련된 향후 생성형 AI의 발전을 가속화 기여에 대한 기대도 더욱 커지고 있다.

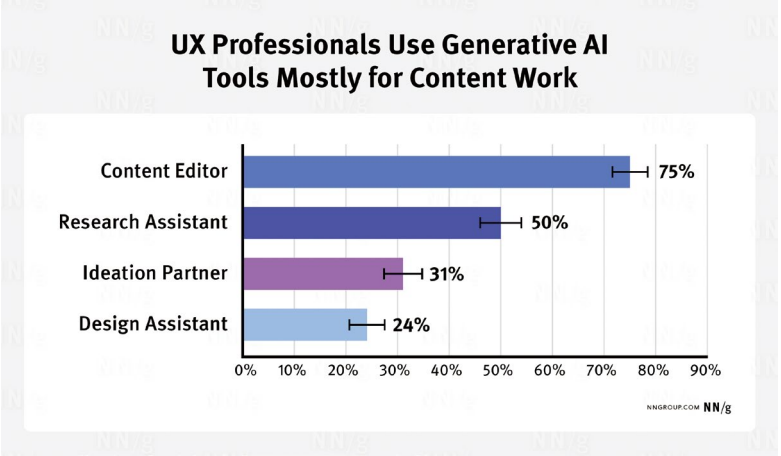
3. 디지털 휴먼 창작

디지털 휴먼은 외형, 음성, 움직임, 인터랙션 등 모든 영역이 통합되어 서비스되는 영역이다. 따라서 생성형 AI를 이용하여 디지털 휴먼을 만들려면, 다양한 멀티모달 데이터를 일괄 학습하도록 수행해야 한다. 최근의 개발 이슈는 LLM(Large Language Model)을 넘어서 멀티모달을 추구하는 LMM(Large Multimodal Model)의 연구가 주를 이룬다. 발화하는 디지털 인간을 창출하고자 한다면, 현재의 기준으로는 음성 네트워크 모델과 디지털 인간 생성 모델 두 종류를 활용하여 추론해야 하는 상황이다. 그러나 텍스트-영상-오디오가 일체로 학습된다면, 단일 모델을 통하여 발화하는 디지털 인간을 창출할 수 있을 전망이다. 이러한 접근을 통한 성능의 증대가 기대된다. 기존의 트랜스포머의 한계점들을 극복하고자 하는 연구들이 진행된다. 트랜스포머에 사용하는 토큰의 크기를 극단적으로 키워 디지털 휴먼의 성능을 증가하는 방법으로 발전될 전망이다. 이러한 기술 전진이 가속화되면, 단일 모델을 통해 대량의 토큰을 처리하여 오디오, 영상, 동작 등을 동시에 창출하는 완벽한 디지털 인간의 출현이 가능할 것으로 예상된다.

4. 인터랙티브 콘텐츠 생성

인터랙티브 스토리텔링을 실현하기 위해 사용자와 AI NPC가 인터랙션을 하면 사용자는 비윤리적인 말과 행동을 NPC에게 할 수 있다. 이는 관리자가 미리 설정해둔 검열 방법을 통해 어느 정도 극복은 할 수 있으나 검열이 과하면 콘텐츠를 즐길 수 없으므로 사용자에게 불쾌한 경험을 줄 수 있으며, 반대로 검열이 없으면 부적절한 콘텐츠를 과하게 생성하여 잘못된 방향으로 콘텐츠가 소비될 수 있다. 이러한 관점에서 안전하고 믿을 수 있는 Believable AI가 중요하게 연구될 것으로 전망된다. 또한, NPC들은 가족 구성원에게도 매우 공식적인 어투로 말하거나 남자 NPC와 여자 NPC가 같은 화장실을 사용하고, 점심을 식당이 아닌 바에서 먹는 등의 다소 비현실적인 행동을 하기도 한다. NPC에게 사회적 인격을 부여하고 그것에 맞게 행동할 수 있도록 유도하는 연구도 필요하다.

특히, 게임 콘텐츠를 생성하는 분야는 점점 학습 데이터 없이 콘텐츠를 생성할 수 있는 쪽으로 연구가 진행되고 있다. 그러나 무결성 문제는 여전히 해결되지 못하고 있다. 사용자가 콘텐츠를 다양하게 생성하고 조작하는 경우, 이러한 콘텐츠들 사이에서 사용자가 의도한 방향으로 문제없이 작동하여



출처 Reprinted with permission from [18], © Nielsen Norman Group.

그림 1 UX 전문가의 생성형 AI 활용 분야

야 한다. 따라서 콘텐츠를 만들면서 콘텐츠에 문제가 없는지 스스로 테스트까지 가능한 지능형 QA 및 콘텐츠 최적화 분야가 중요하게 연구될 것으로 전망된다. 또한, 게임 개발 속도를 높이고 체험 경험을 확장하기 위해서 게임 개발자나 사용자는 생성될 콘텐츠에 현재보다 더 많은 방식과 다양한 방법으로 개입할 것이며, 이를 뒷받침할 기술이 필요하다.

5. 콘텐츠 인터페이스 개발

멀티모달 XR 인터페이스는 텍스트, 이미지, 음성 등 다양한 입력을 복합적으로 활용하여 콘텐츠 체험 및 다양한 창작 결과물을 만들어 낼 수 있다. 그러나 제스처나 시선 처리의 입력을 기반으로 하는 인터페이스는 아직 부족한 편이며, 이에 대한 추가적인 연구 개발이 필요하다.

사람이 생성형 AI를 사용하기 위해 영상이나 음성에 국한되지 않고, AI에게 입력을 주는 궁극적인 인터페이스는 뇌의 전기신호를 활용하는 것이다. BCI(Brain Computing Interface)와 같은 기술로 뇌에 직접적으로 전극을 적용하여 텍스트나 음성을 거치지 않고 상상하는 것을 자극으로 직접 제시할 수 있는 새로운 영역에 대한 연구 개발이 필요하다.

생성형 AI를 활용한 사용자 친화적 인터페이스는 사용자 경험을 향상시키고 사용 문턱을 낮추기 위해 입출력을 좀 더 사용 친화적으로 만드는 방법을 말한다. 이를 위해 사용자와 서비스 플랫폼 간의 경험을 공유하고 보조 기능을 디자인하는 것이 중요해질 것으로 보인다. 하지만 그림 1[18]과 같이 아직까지 생성형 AI는 콘텐츠 편집과 리서치에 편중되어 있고 디자인 영역에서 활발하게 적용되는 것은 아니기 때문에 생성형 AI를 활용한 에이전트의 도입 등 좀 더 다양한 시도가 필요할 것으로 보인다.

네이버는 Cue: 서비스를 통해 사용자의 의도에 따른 맞춤형 답변을 요약, 정리한 형태로 제공하고 쇼핑, 플레이스 등 다양한 서비스와 연계할 수 있는 UX를 적용할 예정이다. 삼성전자의 경우 차세대 노트북 갤럭시4 시리즈에 생성형 인공지능 기능을 탑재한 프로세서를 이용한 온디바이스 AI 노트북과 스마트폰을 선보일 예정이다. 이는 다양한 생성형 AI 모델을 탑재하여 좀 더 사용자가 원하는 기능을 개인 맞춤형 정보 기반으로 대화형 인터페이스를 통해 선

택할 수 있게 도와주는 것이 가능할 것으로 보인다.

이런 시도들은 기존의 UX/UI 디자인 단계에서 사용자의 의도에 맞는 결과물을 만들기 위해 단계를 줄이는 과정이다. 현재 생성형 AI를 활용하기 위해서 활용되는 프롬프트를 사용자의 의도대로 이끌어낼 수 있도록 학습의 방향을 설정하는 것이 중요하며, 소비자가 원하는 UX에 부합하는 서비스를 빠르게 제공하기 위한 미래 시장 선점에 키포인트가 될 것으로 기대된다.

Ⅳ. 결론

생성형 AI를 활용한 디지털 에셋 창작은 게임, 영화 등의 콘텐츠 산업 생산성 향상과 메타버스를 대중화하기 위해 문턱을 낮추는 데 기술적 대안을 제시할 수 있는 핵심 기술이다. 또한, 생성형 AI 기반 콘텐츠 상호작용은 멀티모달을 연계하여 다양한 형태의 체험과 창작 방식을 확장할 수 있는 발전 가능성이 매우 높은 연구분야이다. 앞서 제시한 다양한 분야의 콘텐츠 생성 기술들의 상호 연결과 변환이 가능해지면 창작 한계를 극복하고 활용성을 극대화하여 메타버스에서 활용되는 3차원 사용자 창작 콘텐츠(3D UGC)의 궁극적인 원형으로서 발전이 기대된다.

약어 정리

LLM	Large Language Model
NPC	Non-Player Character
PCG	Procedural Content Generation
PCGRL	PCG via Reinforcement Learning
VAE	Variational AutoEncoder

참고문헌

[1] R. Rombach et al., "High-resolution image synthesis with latent diffusion models," in Proc. IEEE/CVF CVPR, (New Orleans, LA, USA), June 2022, pp. 10684-10695.

[2] A. Blattmann et al., "Align your latents: High-resolution video synthesis with latent diffusion models," in Proc. IEEE/CVF CVPR, (Vancouver, Canada), June 2023, pp. 22563-22575.

[3] GANverse3D: A neural network from NVIDIA reconstructs a 3D Model from a single photo, 2022, https://neurohive.io/en/news/ganverse3d-a-neuralnetwork-from-nvidia-reconstructs-a-3d-model-from-a-single-photo/

[4] G.B.M. Stan et al., "LDM3D: Latent diffusion model for 3D," arXiv preprint, CoRR, 2023, arXiv: 2305.10853.

[5] A. Vaswani et al., "Attention is all you need," in Proc. NIPS, (Long Beach, CA, USA), Dec. 2017, pp. 5998-6008.

[6] B. Jiang et al., "MotionGPT: Human motion as a foreign language," arXiv preprint, CoRR, 2023, arXiv: 2306.14795.

[7] J.S. Park et al., "Generative agents: Interactive simulacra of human behavior," in Proc. ACM UIST, (San Francisco, CA, USA), Oct. 2023, pp. 1-22.

[8] AI Dungeon, https://play.aidungeon.com/(Retrieved Date Accessed 2024. 1. 1.)

[9] 미라클아이, ""잘 지내?" "그렇게 좋진 않아"…같은 게임 샀는데 내용 다르다고?," 2024. 1. 1.

[10] Bitmagic Games, https://bitmagic.games/(Retrieved Date Accessed 2024. 1. 2.)

[11] A. Summerville and M. Mateas, "Super mario as a string: Platformer level generation via lstms," arXiv preprint, CoRR, 2016, arXiv: 1603.00930.

[12] V. Volz et al., "Evolving mario levels in the latent space of a deep convolutional generative adversarial network," in Proc. GECCO, (Kyoto, Japan), Jul. 2018, pp. 221-228.

[13] A. Sarkar, Z. Yang, and S. Cooper, "Conditional level generation and game blending," arXiv preprint, CoRR, 2020, arXiv: 2010.07735.

[14] A. Khalifa et al., "Pcgrl: Procedural content generation via reinforcement learning," AAAI Conf. Art. Intell. Interact. Digit. Entertain., vol. 16, no. 1, 2020, pp. 95-101.

[15] S. Earle et al., "Learning controllable content generators," in Proc. IEEE CoG, (Copenhagen, Denmark), Aug. 2021, pp. 1-9.

[16] Z. Jiang et al., "Learning controllable 3D level generators," in Proc. FDG, (Athens, Greece), Sept. 2022, pp. 1-9.

[17] R. Martin-Brualla et al., "Nerf in the wild: Neural radiance fields for unconstrained photo collections," in Proc. IEEE/CVF CVPR, (Virtual), June 2021.

[18] F. Liu, M. Zhang, and R. Budiu, AI as a UX Assistant, Oct. 2023, https://www.nngroup.com/articles/ai-roles-ux/#Design_Recommendations