생성형 AI를 활용한 메타버스 콘텐츠 개발 동향

ㅁ 방준성, 이상민* / ㈜와이매틱스, *경희대학교

요 약

메타버스는 XR, AI, 블록체인 등 융합 기술을 기반으로 한 몰입형 디지털 경험을 제공하며, 최근 생성형 AI 기술의 발전이 메타버스 콘텐츠 제작의 효율성과 창의성을 증대시켰다. 3D 모델 생성, 대화형 에이전트, 멀티-모달 인식 기술 등을 통해 사실적 그래픽 제작, 동적 환경 조성, 개인화된 사용자 경험을 제공하며, 콘텐츠 제작 비용 절감과 효율성을 높이고 있다. 생성형 AI는 몰입형 콘텐츠 제작 비용을 절감하고, 사용자의 상황과 선호를 기반으로 맞춤형 서비스를 제공하며, 메타버스의 상업적 가능성을 확대한다. 그러나, 데이터 편향성, 개인정보 보호, 딥페이크 등 윤리적 문제와 신뢰 구축이 중요한 과제로 남아있다. 본 기고에서는 생성형 AI 기술을 활용한 메타버스 콘텐츠 개발의 동향과 전망을 다룬다.

I. 서 론

'메타버스(Metaverse)'는 가상·현실의 융합 공간에서 다양한 형식의 인터랙션을 통해 사회·문화·경제 활동에 대한 디지털 경험을 가능하게 하는 세계 또는 플랫폼이다[1][2]. 메타버스 개념의 등장 초기에 코로나19 상황과 디지털 전환에 대한 사회적 요구에 따라 메타버스에 대한 시장 전반의 기대감이 높았다. 그러나, 메타버스에 대한 기대감과 함께, 당시의 메타버스 서비스 개발을 위한 기술적 한계와 서비스 다양성의 부족에 따른 회의론도 함께 제기되었다. 소프트웨어정책연구소(SPRi)의 2023년도 연구보고서에서는 이를 '메타버스 성장통'이라고 정의하고 성장과 하락을 겪고 있는 메타버스 생태계 현황에 따른 효과적 대응 전략과 정책 지원 방향에 대해 다루었다[3]. 콘텐츠 개발에 있어서 가상현실(VR: Virtual Reality)·증강현실(AR: Augmented Reality) 기술의 미성숙, 메타버스 서비스 구현의 높은 복잡도, 사용자 편의성의 한계 등은 메타버스 시장 확대에 장애물이 되었지만, 글로벌 기업들의 메타버스 투자 기조가 유지되며 확장현실(XR: eXtended Reality) 기술의 지속적 발전, 디스플레이의 경량화와 고

해상도화, 생성형 인공지능(AI: Artificial Intelligence) 기술의 활용 등이 이루어지며 실용적인 측면에서 새로운 메타버스 서비스들이 출시되고 있다.

인공지능(AI)은 메타버스의 다양한 콘텐츠 개발에 효율성을 제공할 수 있는 핵심 기술이다[1][4]. ChatGPT, DALL-E, MidJourney 등에 대한 시장의 관심만큼 생성형 AI(Generative AI)의 활용 가능성에 대해 기대감이 높아졌다. 생성형 AI는 사전에 데이터를 학습한 모델에서 입력 정보를 바탕으로 텍스트(Text), 코드(Code), 이미지(Image), 비디오(Video), 오디오(Audio) 등을 출력으로 생성하는 기술이다[5][6]. 생성형 AI는 사용자의 다양한 형태의 입력에 따라 새로운 창작물을 생성하는 기술로도 활용된다. 메타버스라는 디지털 세계를 구성하는 데 있어서 생성형 AI는 가상공간과 그 공간 내 가상객체들을 생산적으로 제작하는 데 활용될 수 있으며 반응형 출력을 이용한 서비스나 콘텐츠 개발에도 활용될 수 있다.

본 기고에서는 생성형 AI를 활용한 메타버스 콘텐츠 개발 동향에 대해 다룬다. II장에서는 메타버스 구성에 활용될 수 있는 생성형 AI 기술을 알아본다. III장에서는 생성형 AI 기술에 의해 메타버스 콘텐츠 개발 환경이 어떻게 변화하게 되었는지 살펴보고, 생성형 AI 기술이 활용됨에 따라 메타버스 서비스나 콘텐츠가 어떤 방향으로 개발될 것인지 예상해 본다. IV장에서는 메타버스 서비스나 콘텐츠 개발에 있어서 생성형 AI 기술을 활용함에 따라 고려해야 하는 이슈들에 대해 논의한다. V장에서 몇 가지 시사점을 제시하며 본 기고를 마무리한다.

II. 메타버스에 활용 가능한 생성형 AI 기술

메타버스 R&D 추진을 위한 대한민국의 '메타버스 신산업 선도전략'에서는 메타버스의 실감성과 사용자 편의성을 향상시키기 위해 XR 기술과 D(Data)·N(Network)·A(AI)와 S(Security, Blockchain) 기술 요소[1]가 결합된 융합 기술 중심의 5대 핵심기술로 '도시 규모 광역 메타버스 공간 구성', '사실적 감성 표현, 지능형 디지털 휴먼', '입체·공간·객체 기반 초실감 디지털 미디어', '대규모 참여자의 사실적 상호작용을 위한 실시간 UI/UX', '상호운용성을 갖는 분산개방형 플랫폼'을 선정했다[7]. 특히, 메타버스 신산업 선도전략에서 AI는 현실과 가상세계 간 인터랙션을 촉진하는 역할을 할 것으로 기대되고 있다. 광역 메타공간은 가상·현실 융합 공간에서 사회·문화·경제 활동을 지속할 수 있는 메타 정보를 내재화한 디지털 공간 구성 및 운용에 대한 융합 기술 개발을 목표로 하며, 실세계에 연결된 가상 공간 및 객체별 속성 정보를 갖는 3D 모델 데이터를 구축한다. 디지털 휴먼 분야에서는 정교한 시각적 표현과 감성 표현을 구현하고, 자율적으로 반응하는 지능의 내재화를 지향하며, 경제 활동의 신뢰성 확보를 위한 사용자/아바타 신원인증 기술 개발도 함께 추진된다. 초실감 미디어는 미디어를 3차원의 디지털 객체로 구현하여 사용자 시점과 공간 이동에 따라 동적으로 반응하는 환경을 조성할 수 있는 기술 개발을 목표로 한다. 마지막으로, 실시간 UI/UX는 다중 사용자 접속 지연 없이 자유로운 상호작용과 오감으로 확장된 인터페이스 구현을 추구한다.

1) 메타버스 신산업 선도전략 중 기술별 역할(이준우, 2022):
(1) XR: 현실과 가상(디지털) 세계를 연결하는 인터페이스로 현실·가상세계의 공존 촉진과 몰입감 높은 가상융합 공간과 디지털 휴먼 등 구현; (2) 디지털 트윈: 현실 세계를 가상세계에 3D로 복제하고 동기화한 뒤 시뮬레이션·가상훈련 등을 통해 지식의 확장과 효과적 의사결정 지원; (3) 블록체인: 메타버스 창작물에 대한 저작권 관리, 사용자 신원확인 및 데이터 프라이버시 보호, 콘텐츠 이용내역 모니터링 및 저작권료 정산 등 지원; (4) 인공지능: 메타버스 내 데이터 및 사용자 경험 학습, 실시간 통·번역, 사용자 감성 인지 및 표현 등을 통해 현실-가상세계 간 상호작용 촉진; (5) 데이터: 실세계 데이터 취득 및 유효성 검증, 데이터 저장·처리·관리 등 수행; (6) 네트워크: 초고속·초저지연 5G/6G 네트워크, 지능형 분산 컴퓨팅(MEC) 등을 통해 대규모 이용자 동시 참여, 실시간 3D·대용량 콘텐츠 서비스 제공; (7) 클라우드: 이용자 요구나 수요 변화에 따라 컴퓨팅 자원을 유연하게 배분

이와 함께 분산·개방형 플랫폼은 사용자가 다양한 형태의 메타버스를 넘나들 수 있도록 서비스의 개방성과 상호운용성 기술 개발을 목표로 하고 있다.

생성형 AI는 2014년에 생성적 적대 신경망(GAN: Generative Adversarial Network)이라는 인공신경망 기술의 잠재적 응용 가능성 때문에 관심을 받기 시작하였다[8]. 생성형 AI는 일반적으로 생성 모델(Generative Model)과 판별 모델(Discriminative Model)로 구성되는데, 생성 모델은 데이터셋을 학습하여 이와 유사하면서도 새로운 출력을 생성하며 판별 모델은 입력 데이터를 특정 기준에 따라 분류한다. 예를 들어, 생성 모델이 고양이 이미지의 특성을 학습하여 새로운 고양이 이미지를 생성한다면, 분류 경계를 찾는 것을 목적으로 학습한 판별 모델은 새롭게 생성한 고양이 이미지가 고양이인지 아닌지를 판별하여 분류한다.

생성형 AI에 다양한 기술이 있지만, 3차원 모델 생성(3D Model Generation) 기술, 대화형 에이전트(Conversational Agent) 기술, 멀티-모달 인식(Multi-Modal Recognition) 기술, 데이터 스페이스(Data Space) 기술 등이 주목된다.

1. 3차원 모델 생성 기술

DALL-E, MidJourney 등에서 보듯이 생성형 AI 기술을 활용하여 스타일 변환, 캐릭터 생성, 이미지 증강 등이 가능하게 되었다. 사용자의 프롬프트 입력에 따른 선택적 옵션에 따라 여러가지 사실적 표현이 가능하지만, 여전히 세부적인 통제 하에서 이미지를 생성하기에 한계가 있어서 AI 모델의 투명성, 설명가능성 등에 대한 연구가 필요하다. 대표적인 이미지 생성 모델로는 AE(Auto-Encoder)[2], VAE[3][9], GAN[8] 등이 있다. GAN은 생성 모델[4]과 판별 모델[5]이 경쟁하면서 실제와 가까운 이미지, 동영상, 음성 등을 만들어 내는 인공신경망이다. GAN은 학습 데이터가 적어도 생성 모델과 판별 모델을 활용하여 성능 향상이 가능한 장점이 있으나, 데이터 분포를 바탕으로 출력을 생성하기 때문에 생성기(Generator)와 판별기(Discriminator)가 균형 있게 학습되지 않으면 편향적 결과가 나올 수 있다는 단점이 있다. GAN의 파생 모델로 CycleGAN이나 DCGAN(Deep Convolutional Generative Adversarial Network) 등이 있다. CycleGAN은 이미지를 다른 스타일의 이미지로 변환할 수 있는 GAN 기반의 이미지 변환 모델이다. DCGAN은 컨볼루션 신경망 구조에서의 생성기와 판별기를 사용하는데, 저해상도 이미지를 고해상도로 변환하는 어플리케이션에 활용된다. 한편, 디퓨전 모델(Diffusion Model)은 이미지 생성 초기에 노이즈 상태에서 시작하여 점진적으로 노이즈를 제거해 가며 이미지를 복원하는 방식으로 작동하는데[10][11], 이러한 생성 모델들의 성능을 검증하기 위해서 CIFAR-10, CelebA, STL-10, LSUN, FFHQ 등의 표준화된 데이터셋이 널리 활용되고 있다[6].

이미지 생성 모델이 발달함에 따라 다양한 3D 이미지/콘텐츠 생성 모델들이 개발되어 메타버스 환경에 필요한

2) AE는 인코더(Encoder)와 디코더(Decoder)의 수가 같은 인공신경망 모델로 입력 데이터의 가장 중요한 특징을 학습한다. AE는 비지도-학습(Unsupervised Learning) 방식 기반으로 레이블(Label)이 없는 입력 데이터 활용이 가능하다는 장점이 있으나, 훈련 데이터셋의 것과 비슷하게만 만드는 과적합(Overfitting) 문제가 있을 수 있다. AE는 손상된 이미지 복구나 워터마크 제거 등에 활용된다.

3) VAE에서 인코더는 주어진 데이터를 디코더가 원래의 데이터로 잘 복원할 수 있도록 하는 확률분포를 찾는 역할을 하고 디코더는 확률분포와 입력 데이터를 바탕으로 유사하게 재생성하는 역할을 한다. GAN에 비해 출력 결과의 품질이 낮다.

4) 생성 모델인 Generator는 노이즈 벡터를 입력에 추가하여 출력을 생성하는데, 이때 생성된 출력은 가짜(Fake)이다.

5) 판별 모델인 Discriminator는 Generator가 생성한 출력을 입력 받아 그것이 가짜(Fake)인지 실제(Real)인지를 판단한다.

6) CIFAR-10은 airplane, automobile, bird, cat, dog 등 8,000만 개의 작은 이미지 데이터셋, CelebA는 약 200,000개 정도의 얼굴 데이터셋, LSUN는 침실, 식당, 교회 등 10개 범주의 약 300만장 규모의 장면 데이터셋, FFHQ는 연령, 민족 등으로 구분된 70,000개의 고품질 얼굴 데이터셋으로 구성되어 있다.

3D 객체 제작을 효율적으로 지원하고 있다. OpenAI의 Point-E는 프롬프트 입력을 바탕으로 포인트 클라우드 방식의 3D 이미지를 생성한다[12]. Adobe에서는 3D 디지털 콘텐츠 생성을 위해 3D 모델링 소프트웨어인 Substance 3D 모델러[7]를 개발했다. 특히 3D 모델 편집 및 조각에 중점을 두고 있다[13]. NVIDIA의 GET3D(Generate Explicit Textured 3D)는 2D 이미지 입력으로 3D 폴리곤 메쉬를 합성할 수 있는 3D 생성 모델로, 메타버스 공간에서 사용할 건물, 차량, 캐릭터 등 다양한 객체의 3D 이미지를 생성한다[14]. 3D 생성 모델에서 생성기는 토폴로지 폴리곤 메쉬와 그 폴리곤 메쉬의 표면 지점의 색상이나 재질을 나타내는 텍스처 필드를 생성하고 판별기는 3D 모델의 합성 결과를 기반으로 출력에 대한 품질을 평가하고 그 결과를 최적화한다. 이외에도 다양한 3D 모델러들이 개발되고 있다.

2. 대화형 에이전트 기술

ChatGPT, Claude, Perplexity 등 대규모 언어 모델(LLM: Large Language Model) 기반의 서비스들이 다수 등장하고 있다. 자연어 처리(NLP: Natural Language Processing) 분야에서 과거에 번역, 요약, 질문-응답 생성 등에 언어 생성 모델로 트랜스포머(Transformer)[8][15], BERT(Bidirectional Encoder Representations from Transformers)[16] 등이 사용되었다. 트랜스포머는 순환신경망 계열의 순차적인 연산 방식에서 벗어나 대량의 병렬 처리를 통한 대규모 언어 학습이 가능하다. GPT(Generative Pre-trained Transformer)[9]는 OpenAI가 개발한 LLM이다. 대규모 말뭉치 데이터셋을 통해 모델을 사전학습시킨(Pre-trained) 다음에 특정 태스크에 맞게 상대적으로 적은 양의 데이터를 사용하여 파인 튜닝을 하여 사용할 수 있다[17]. 이러한 사전학습 모델은 전이 학습 능력을 통해 다양한 언어 처리 태스크에 유용하게 적용될 수 있으며, 특히 순방향 어텐션 메커니즘을 활용하여 자연스러운 문장 생성에도 직접적으로 활용할 수 있다는 장점이 있다. 대규모 언어 모델은 학습 및 운용을 위해 고성능 컴퓨팅 및 메모리 자원이 필요하지만, 번역, 요약, 질의응답 등이 가능한 챗봇이나 가상 비서 형태의 인터페이스에서 자연스러운 대화를 생성하는 데 유용하다. GPT-3부터는 자연스러운 글 작성, 코딩 등이 가능해졌는데, GPT-3.5에서 인간 피드백을 통한 강화 학습(RLHF: Reinforcement Learning with Human Feedback) 적용으로 응답의 정확도가 더 상승했다. GPT-4는 멀티-모달(Multi-modal) 입력이 가능한 LLM으로, 텍스트와 이미지 입력을 바탕으로 질의응답이 가능하다. GPT-3 이후에 개발된 대표적인 초거대 언어 모델(Hyper-scale Language Model)로는 Meta의 LLaMA 시리즈, Google의 LaMDA와 PaLM, NVIDIA의 Megatron, Huawei의 Pan-GU 등이 있다[18].

NVIDIA의 Omniverse Avatar Cloud Engine은 쇼핑몰, 서비스 센터, 전자 상거래 플랫폼 등에서 사용될 수 있는 대화형 에이전트를 제공한다[19]. 이 에이전트는 실시간 음성 인식, 자연어 처리 등의 기술들과 결합되어 사용자와 대화하며 맞춤형 상품 추천, 재고 확인, 결제 지원

7) 2002년에 설립된 프랑스 회사인 Allegorithmic에서 Substance 3D Modeler를 개발했고, Adobe가 2019년에 Substance 도구 모음을 포함하여 Allegorithmic을 인수하였다. Substance 3D Modeler는 Substance Painter, Substance 3D Sampler, Substance 3D Stager, Substance Designer 등의 3D 도구와 다른 3D 모델링/렌더링 소프트웨어와 통합이 가능하다.

8) 트랜스포머는 문장 내에서 단어들 사이의 관계를 추적하여 맥락과 의미를 학습하는 모델이다. Seq2Seq(Sequence to Sequence) 모델 구조인 인코더-디코더를 따르면서 어텐션(Attention) 메커니즘을 이용하여 구현한 모델로, 순환신경망(RNN: Recurrent Neural Network)에서 병렬처리가 어려워서 학습 시간이 오래 걸리는 문제 등을 해결할 수 있다. 어텐션 메커니즘은 입력 시퀀스에서 의미 있는 요소들 위주로 정보를 추출하는 기법으로 단어들 사이의 연관도를 계산하여 문장 입력 시퀀스 출력 과정에서 어디에 더 중점을 둬야 할지를 판단한다.

9) GPT 모델은 GPT-1(2018년, 1,700만 개 파라미터), GPT-2(2019년, 15억 개), GPT-3(2020년, 1,750억 개), InstructGPT(2022년, 1,750억 개), GPT-3.5(2022년, 1,750억 개), GPT-4(2023년) 시리즈로 발전해 왔다.

등을 수행할 수 있다. 예를 들어, 사용자가 가상 상점에서 특정 아이템을 찾고자 할 때 대화형 에이전트가 실시간으로 정보를 제공하고 구매 과정을 지원할 수 있게 된다. Epic Games의 MetaHuman Creator는 메타버스 환경에서 사용할 수 있는 고품질의 가상 캐릭터를 제작할 수 있는 도구를 제공한다[20]. 대화형 에이전트는 게임 내 NPC(Non-Playable Character)로 활용되거나 게임 내 스토리 진행, 사용자 가이드 제공, 캐릭터 간 협력 활동 등을 지원할 수 있다. Roblox는 AI 기반 NPC를 활용해 커뮤니티 활동을 지원한다[21]. AI NPC 형태의 이 대화형 에이전트는 가상 이벤트를 조직하거나 사용자 간의 연결을 촉진하는 역할을 할 수 있다. Soul Machines는 감정 기반 상호작용을 가능하게 하는 디지털 휴먼을 개발했는데[22] 이 디지털 휴먼은 메타버스에서 고객 서비스, 심리 상담, 건강 관리 등을 위해 사용자와 실시간 대화를 통해 문제를 해결하거나 정서적 지원을 제공한다. 예를 들어, 의료 분야에서는 가상 상담사로서 환자들에게 질병 관리 및 심리 상담 서비스를 제공할 수 있다. 이외에도 여러 도메인에서 AI 기술의 발전으로 다양한 인터랙션이 가능한 대화형 에이전트들이 개발되고 있다.

3. 멀티-모달 인식 기술

AI 기술의 발전으로 멀티-모달(Multi-Modal) 입력을 바탕으로 사용자와 인터랙션이 가능하게 되었다. 멀티-모달이란 다양한 모달(데이터 입력의 방식)로 입력되는 데이터를 동시에 처리하거나 결합하는 것을 의미한다. 예를 들어, AI 시스템이 텍스트와 이미지를 동시에 처리하여 사용자 의도를 인식할 수 있다면 이 시스템은 멀티-모달을 지원한다고 말한다[23].

메타버스 환경에서 사용자 경험을 극대화하기 위해 감정 인식 기술이 중요한 역할을 할 수 있다. 멀티-모달 감정 인식은 음성, 표정, 제스처, 생체 신호(심박수, 피부 전도도 등)와 같은 다양한 데이터를 종합적으로 분석하여 사용자의 감정 상태를 추론하는 방식일 수 있다. 예를 들어, Meta의 AI 연구팀은 음성 톤, 안면 표정 분석, 그리고 사용자의 상호작용 패턴을 결합하여 사용자의 정서적 상태를 실시간으로 파악하는 기술을 개발 중이다. 이러한 기술은 메타버스에서 감정적으로 민감한 환경을 조성하거나, 교육 및 상담 분야에서 맞춤형 정서 지원을 제공하는 데 활용될 수 있다. 예를 들어, 가상 교실에서 학생이 스트레스를 받는 징후를 보이면, 시스템이 이를 감지해 적절한 도움을 제안할 수 있다.

VR 기반 메타버스에서는 사용자의 행동, 음성, 움직임 데이터를 결합하여 상황을 실시간으로 인식하는 멀티-모달 기술이 유용할 수 있다. 이러한 기술은 사용자의 몸짓이나 시선 움직임을 추적하여 의도를 파악하고, 이를 가상 환경에 반영함으로써 자연스럽고 몰입감 있는 상호작용을 제공할 수 있다. 예를 들어, NVIDIA의 Omniverse는 사용자의 손 동작, 음성 명령, 공간 내 이동 경로를 통합적으로 분석하여 가상 협업 환경에서 적응형 작업 지원을 제공한다. 이러한 기술은 VR 게임, 가상 회의, 디자인 협업 등 다양한 분야에서 사용자 경험을 향상시키는 데 기여한다. AR 기반 메타버스에서는 현실 세계의 물리적 환경과 디지털 콘텐츠 간의 통합적 인식을 통해 상황을 이해하는 멀티-모달 기술이 활용될 수 있다. 예를 들어, Microsoft의 HoloLens는 사용자의 제스처나 음성 명령에 대한 데이터를 넘어, 시선, 생체 신호 등의 데이터를 실시간으로 분석하고 물리적 환경의 3D 스캔 데이터를 결합하여 다양한 AR 객체와의 상호작용을 지원하도록 발전할 수 있다. AR 기반 메타버스에 대한 기술이 물리적 공간과 가상 객체 간의 일관된 상호작용을 할 수 있도록 기술 수준이 발전하게 되면 교육, 제조, 의료 분야에서 생산성과 효율성을 높이는 데 기여할 수 있다.

공간 컴퓨팅은 물리적 공간과 디지털 공간의 상호작용을 가능하게 하는 기술로, 멀티-모달 인식 기술은 이러한 상호작용의 핵심 요소이다. 공간 컴퓨팅 기술은 주변 환경을 3D로 모델링하고, 이를 디지털 데이터와 연계하여 사용자가 공간 내에서 자연스럽게 행동하는 환경 구축에 활용될 수 있다. 멀티-모달 인식 기술은 이러한 환경에서 사

용자 상태와 의도를 정확히 파악하여, 보다 직관적이고 개인화된 사용자 경험을 제공하는 데 필수적이다. 이를 통해 메타버스는 단순한 가상 환경을 넘어, 현실과 가상이 융합된 새로운 차원의 상호작용 플랫폼으로 발전이 가능하다.

Ⅲ. 생성형 AI 기술을 활용한 메타버스 콘텐츠 개발 동향

1. 생성형 AI에 의한 메타버스 서비스 개발 환경의 변화

생성형 AI는 메타버스 환경에서 콘텐츠 제작과 사용자 경험 향상에 있어 다음과 같이 세 가지 주요한 기여를 한다. 첫째로, 사실적 그래픽과 3D 모델 생성 비용의 감소이다. 생성형 AI는 방대한 데이터셋을 학습하여 현실 세계의 물체와 환경을 고해상도로 재현하는 데 활용할 수 있다. 이는 몰입형 공간 구성에 도움이 된다. 기존에는 3D 모델을 제작하는 데 전문가의 수작업이 필수적이었으나, 생성형 AI는 자동화된 모델링 과정을 통해 이 과정을 간소화할 수 있다. 예를 들어, NVIDIA의 Omniverse와 같은 플랫폼은 AI를 활용하여 건물, 자연경관, 도시 풍경 등을 사실적으로 생성하며, 이는 설계와 시뮬레이션 작업의 효율성을 크게 향상시킨다[24][25]. 이러한 기술은 건축, 제조, 게임 및 영화 산업에서 활용되어, 수개월이 걸리던 작업을 며칠 만에 완성할 수 있도록 지원한다. 또한, 비용 절감 효과와 함께 사용자 맞춤형 모델 제작이 가능해져 산업 전반의 창의성과 생산성을 동시에 높이는 데 기여한다. 그래픽 작업 소요 비용이 큰 메타버스 콘텐츠 제작 과정에 있어서의 비용 절감은 더 나은 메타버스 콘텐츠 제작에 자본이 투입될 수 있도록 이끌어 사용자 경험 향상에 도움이 된다.

둘째로, 사용자 상황과 행동에 기반한 동적 환경 조성이다. 사용자 위치에 따라 변화하는 VR 공간과 사용자 인터랙션에 따라 변화하는 VR 및 AR 객체들을 활용할 수 있게 되면 서비스 범위를 확대하고 콘텐츠를 다양하게 구성할 수 있게 된다. 예를 들어, The Matrix라는 시뮬레이터는 1인칭 및 3인칭 관점에서 실시간 반응형 제어를 제공하여 720p급의 고화질 비디오 스트림을 생성한다[26]. 또한, Unity와 Unreal Engine은 AI 기반의 물리 시뮬레이션 및 동적 상호작용 기능을 제공하여 사용자가 특정 물체를 이동시키거나 경로를 따라가는 콘텐츠를 실시간으로 제작할 수 있도록 돕는다. 이는 몰입형 게임과 시뮬레이션 환경에서 사용자 경험을 한층 강화하는 데 기여하며, 군사 훈련, 의료 시뮬레이션, 직업 교육 등 다양한 분야에 적용되고 있다. 특히, 환경 변화와 상호작용이 중요한 생물학적 시뮬레이션이나 가상 실험실에서도 이러한 기술이 활용되며, 현실감을 더욱 높여주는 요소로 자리잡게 될 것이라고 본다.

셋째는, 대화형 인터페이스를 활용한 개인화된 경험 제공이다. AI 기반 대화형 인터페이스는 메타버스 내에서 사용자의 행동 데이터를 분석하여 개인화된 경험을 제공한다[27][28]. 예를 들어, Meta의 Horizon Worlds와 같은 플랫폼은 사용자의 과거 활동 기록과 선호도를 기반으로 맞춤형 콘텐츠를 추천하고, 사용자가 선호하는 경로를 설계한다[29]. 이는 게임 내 맞춤형 캐릭터 생성, 개인화된 학습 콘텐츠 제공, 사용자의 심리적 상태에 맞는 환경 제안 등으로 구체화될 수 있다. 공간 컴퓨팅 기술의 발전은 현실 세계와 디지털 공간 간의 연결성을 강화하며, 현실 데이터를 디지털화하여 메타버스에서 새로운 유형의 콘텐츠 개발을 촉진한다. 이러한 기술은 가상 쇼핑, 원격 협업, 디지털 트윈 구현 등에서 실질적인 가치를 제공하며, 사용자 중심의 디지털 경험 설계 가능성을 높인다.

2. 생성형 AI에 의한 메타버스 콘텐츠 개발 범위의 확장

대화형 인터페이스를 통한 AI 에이전트 활용이 가능해졌다. 대화형 에이전트는 메타버스 환경에서 사용자와의 상호작용을 풍부하게 만들며 다양한 분야에서 활용되고

있다. CodeBaby는 교육 및 의료 분야에서 실시간으로 대화하며 사용자와 인터랙션 가능한 캐릭터를 생성하는 서비스를 제공한다. 대화 처리 및 이해 기술의 발전은 캐릭터의 반응을 보다 자연스럽게 만들어 사용자의 몰입감을 극대화시켜 준다. 또한, 인간 행동 심리학의 적용으로 메타버스 내 캐릭터가 현실 세계와 유사한 사회적 경험을 제공함으로써 사용자에게 정서적 연결감을 줄 수 있다. 예를 들어, 의료 상담에서 환자가 자신의 상태를 설명할 때 대화형 에이전트가 실시간 피드백을 제공하여 심리적 안정을 도울 수 있을 것이다. 반대로, 메타버스에서의 사회·문화·경제 경험 중에 현실 세계에서 발생하는 질투 등의 감정이 다른 방식으로 나타날 수 있기도 하다[30].

멀티-모달 입력 데이터에 대한 처리 및 분석 기술의 발전은 메타버스에서 사용자와 환경을 보다 깊이 이해할 수 있는 기반을 마련했다. AI는 얼굴 인식과 음성 분석을 통해 사용자의 감정을 파악하고, 이에 맞춰서 적합한 반응을 하도록 할 수 있다. 사용자가 스트레스를 받는 상황에서는 AI가 차분한 배경음악을 재생하거나, 시각적으로 안정감을 주는 장면을 생성할 수 있다. 또한, 메타버스 내 사용자의 위치와 행동을 분석하여 상황에 적합한 맞춤형 서비스를 제공한다. 학습 모드에서는 학습 콘텐츠를, 게임 모드에서는 도전적이고 재미있는 환경을 제공하는 방식이 될 수 있다. 이러한 기술은 사용자가 메타버스 공간에서 느끼는 몰입감을 극대화하고, 개인화된 경험을 제공한다.

생성형 AI는 사용자가 원하는 테마나 스토리에 맞춘 콘텐츠를 실시간으로 생성하여 메타버스 내에서 개인화된 경험을 제공하는 데 활용될 수 있다. 메타버스 내에서의 생성형 AI의 활용으로 인해 사용자는 과거에 상상했던 교육 분야 콘텐츠로 가상 공간에서 역사적 사건을 체험하거나 과학 실험을 시뮬레이션하는 콘텐츠를 더 몰입감 높은 방식으로 경험을 하거나, 특정 스토리라인에 따라 자신만의 환경을 커스터마이징할 수 있게 되었다. 엔터테인먼트 분야에서는 사용자가 자신의 아바타를 통해 독창적인 스토리를 만들고 캐릭터를 조작하며, 다양한 감정을 표현할 수 있는 인터랙티브 경험을 제공할 수 있게 되었다. 이러한 맞춤형 콘텐츠는 메타버스의 상업적 가능성을 확대하며, 사용자 참여도를 높이는 데 기여할 수 있다.

Ⅳ. 생성형 AI 기술 활용에 따른 윤리적 이슈

생성형 AI는 메타버스의 혁신적인 발전을 가능하게 하는 핵심 기술로, 현실과 가상의 경계를 허물며 몰입감 있는 디지털 경험을 창출한다. 하지만 이러한 기술의 확산은 심각한 윤리적 문제를 수반한다. 생성형 AI가 학습하는 데이터는 종종 편향된 정보를 포함하며, 이로 인해 메타버스 내에서 인종 차별, 성 차별 등 부당한 결과물이 생성될 가능성이 있다[31]. 더 나아가, AI가 생성한 콘텐츠는 딥페이크를 포함하여 가짜 뉴스나 사기 사건에 악용될 수 있어 신뢰 기반의 디지털 생태계를 위협한다. 특히, 메타버스 환경에서는 사용자들의 개인정보가 다량으로 수집되고 처리되는데, 이러한 데이터가 부주의하게 활용될 경우 개인의 프라이버시와 안전이 심각하게 침해될 위험이 있다.

생성형 AI의 윤리적 문제를 해결하기 위해 주요 국가들은 다양한 정책과 규제를 수립하고 있다. 미국은 '국가 AI 이니셔티브법'과 같은 법안을 통해 AI 기술의 윤리적 활용과 국가 안보를 강화하고 있으며, 영국은 '국가 AI 전략'을 통해 AI 생태계의 지속가능성과 거버넌스를 강화하고 있다. 유럽연합은 AI를 위험 수준에 따라 분류하는 규제를 통해 인권과 안전을 보호하려는 노력을 기울이고 있으며, 캐나다와 싱가포르는 데이터 윤리와 투명성 확보를 위한 구체적인 가이드라인을 제시하고 있다. 이들 국가들은 AI 기술의 발전을 촉진하면서도 기술이 가져올 부작용을 최소화하기 위해 윤리적 원칙과 투명한 거버넌스를 기반으로 하는 정책을 도입하고 있다.

메타버스에서 생성형 AI를 윤리적으로 활용하려면 투명성과 책임성을 강화하는 기술적, 정책적 노력이 병행되어야 한다. 첫째, AI 모델 개발 과정에서 알고리즘의 편향성을 줄이고 데이터의 공정성을 확보해야 한다. 둘째, 사

용자 데이터를 보호하기 위한 강력한 프라이버시 보호 장치와 규제를 마련해야 한다. 셋째, 메타버스에서 발생할 수 있는 윤리적 문제를 사전에 식별하고, 이를 해결하기 위한 다국적 협력 체계를 구축할 필요가 있다. 메타버스가 신뢰할 수 있는 디지털 환경으로 자리 잡을 수 있도록 기술 혁신과 윤리적 균형을 동시에 이루는 지속가능한 발전 방안을 강구해야 한다.

V. 결 론

생성형 AI는 메타버스 발전에 있어 핵심적인 기술로, 현실과 가상의 경계를 허물며 몰입감 있는 디지털 경험을 창출한다. 이 기술은 텍스트, 이미지, 3D 모델, 대화형 에이전트 등의 생성 능력을 통해 메타버스 콘텐츠 개발의 효율성을 극대화하고, AI 기반 자동화는 콘텐츠 제작 비용과 시간을 줄이는 동시에 사용자 맞춤형 서비스와 몰입형 환경을 구현하는 데 기여하고 있다. 또한, 고해상도의 사실적 그래픽 생성, 동적 환경 조성, 개인화된 경험 제공 등에서 중요한 역할을 수행하며, 가상 공간의 상호작용성과 활용성을 높이고 있다. 그러나 이러한 기술적 진보는 윤리적 과제도 수반하며 메타버스의 지속가능한 발전을 위해 해결해야 할 핵심 이슈로 남아있다. 생성형 AI가 메타버스 생태계의 지속가능한 성장을 견인하는 기술적 동력으로 자리잡기를 기대한다.

참 고 문 헌

[1] 방준성, “메타버스 서비스 확대를 위한 인공지능 기술의 활용,” 한국통신학회지(정보와통신) 39(2) (2022), pp. 64-73.
[2] 방준성, 박판근, “디지털 윤리와 UX를 반영한 메타버스 R&D 추진전략,” 방송공학회논문지 27(5) (2022), pp. 703-717. doi: https://doi.org/10.5909/JBE.2022.27.5.703 (accessed Dec. 27, 2024)
[3] 한상열, 이승환, 전이술, 양희태, “메타버스 성장통 극복 방안 연구,” 소프트웨어정책연구소, 연구보고서 RE-173 (2023).
[4] M. M. Soliman, E. Ahmed, A. Darwish, and A. E. Hassanien, “Artificial intelligence powered Metaverse: analysis, challenges and future perspectives,” Artificial Intelligence Review 57(26) (2024). doi: https://doi.org/10.1007/s10462-023-10641-x (accessed Dec. 27, 2024)
[5] S. S. Sengar, A. B. Hasan, S. Kumar, and F. Carroll, “Generative artificial intelligence: a systematic review and applications,” Multimedia Tools and Applications 29(5) (2024), pp. 535-545. doi: https://doi.org/10.1007/s11042-024-20016-1 (accessed Dec. 27, 2024)
[6] 방준성, 조병철, “생성형 인공지능에 의한 콘텐츠 저작과 유통 환경의 변화에 따른 정책 제언,” 방송공학회논문지 28(4) (2023), pp. 400-409. doi: https://doi.org/10.5909/JBE.2023.28.4.400 (accessed Dec. 27, 2024)
[7] 이준우, “메타버스 기술 및 표준 동향과 R&D 추진방향,” TTA저널 200 (2022), pp. 57-66.
[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” Communications of the ACM 63(11) (2020), pp. 139-144. doi: https://doi.org/10.1145/3422622 (accessed Dec. 27, 2024)
[9] D. P. Kingma and Max Welling, “An Introduction to Variational Autoencoders,” Foundations and Trends in Machine Learning 12(4) (2019), pp. 307-392. doi: https://doi.org/10.48550/arXiv.1906.02691 (accessed Dec. 27, 2024)
[10] C. Luo, “Understanding Diffusion Models: A Unified Perspective,” arXiv preprint arXiv:2208.11970 (2022). doi: https://doi.org/10.48550/arXiv.2208.11970 (accessed Dec. 27, 2024)

참 고 문 헌

[11] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion Models: A Comprehensive Survey of Methods and Applications," arXiv:2209.00796v14 (2022).
doi: https://doi.org/10.48550/arXiv.2209.00796 (accessed Dec. 27, 2024)

[12] A. Nichol, H. Jun, P. Dhariwal, P. Mishkin, and M. Chen, "Point-E: A System for Generating 3D Point Clouds from Complex Prompts," arXiv:2212.08751v1 (2022).
doi: https://doi.org/10.48550/arXiv.2212.08751 (accessed Dec. 27, 2024)

[13] Adobe Substance 3D, https://www.adobe.com/products/substance3d.html (accessed Dec. 27, 2024).

[14] J. Gao, T. Shen, Z. Wang, W. Chen, K. Y., Daiqing Li, O. Litany, Z. Gojcic, and S. Fidler, "GET3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images," NeurIPS Conference (2022).
doi: https://openreview.net/forum?id=GAUwreODU5L (accessed Dec. 27, 2024)

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems (2017), pp. 6000 - 6010.
doi: https://doi.org/10.48550/arXiv.1706.03762 (accessed Dec. 27, 2024)

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 1 (2019), pp. 4171-4186.
doi: https://doi.org/10.18653/v1/N19-1423 (accessed Dec. 27, 2024)

[17] S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, and J. Gao, "Large Language Models: A Survey," arXiv:2402.06196v2 (2024).
doi: https://doi.org/10.48550/arXiv.2402.06196 (accessed Dec. 27, 2024)

[18] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, W. Ye, Y. Zhang, Y. Chang, P. S. Yu, Q. Yang, and X. Xie, "A Survey on Evaluation of Large Language Models," ACM Transactions on Intelligent Systems and Technology 15(3) (2024), pp. 1-45.
doi: https://doi.org/10.1145/3641289 (accessed Dec. 27, 2024)

[19] NVIDIA's Avatar Cloud Engine (ACE), https://developer.nvidia.com/blog/tag/omniverse-ace/ (accessed Dec. 27, 2024).

[20] Epic Games's MetaHuman Creator, https://dev.epicgames.com/documentation/en-us/metahuman/metahuman-creator-overview (accessed Dec. 27, 2024).

[21] S. Rao, W. Xu, M. Xu, J. Leandro, K. Lobb, G. DesGarennes, C. Brockett, and B. Dolan, "Collaborative Quest Completion with LLM-driven Non-Player Characters in Minecraft," arXiv:2407.03460v1 (2024).
doi: https://arxiv.org/html/2407.03460v1 (accessed Dec. 27, 2024)

[22] Soul Machines, https://www.soulmachines.com/ (accessed Dec. 27, 2024).

[23] H. Chen, X. Wang, Y. Zhou, B. Huang, Y. Zhang, W. Feng, H. Chen, Z. Zhang, S. Tang, and W. Zhu, "Multi-Modal Generative AI: Multi-modal LLM, Diffusion and Beyond," arXiv:2409.14993v1 (2024).
doi: https://doi.org/10.48550/arXiv.2409.14993 (accessed Dec. 27, 2024)

[24] NVIDIA Omniverse for Developers, https://developer.nvidia.com/omniverse (accessed Dec. 27, 2024).

[25] H. A. Alhaija, J. Lucas, A. Zook, M. Babcock, D. Tyner, R. Rao, and M. Shugrina, "Interactive AI Material Generation and Editing in NVIDIA Omniverse," *ACM SIGGRAPH'23* 4 (2023), pp. 1-2.
doi: https://doi.org/10.1145/3588430.3597248 (accessed Dec. 27, 2024)

[26] R. Feng, H. Zhang, Z. Yang, J. Xiao, Z. Shu, Z. Liu, A. Zheng, Y. Huang, Y. Liu, and H. Zhang, "The Matrix: Infinite-Horizon World Generation with Real-Time Moving Control," arXiv:2412.03568v1 (2024).
doi: https://doi.org/10.48550/arXiv.2412.03568 (accessed Dec. 27, 2024)

참 고 문 헌

[27] 방준성, 이병탁, 박판근, "대규모 언어 모델을 사용하는 인공지능 기반 대화형 챗봇의 편향성 평가 프레임워크 개발 방법," 방송공학회논문지 28(5) (2023), pp. 545-563.
doi: https://doi.org/10.5909/JBE.2023.28.5.545 (accessed Dec. 27, 2024)
[28] 방준성, 이상민, "인공지능 기반 학습자 맞춤형 교육을 위한 형평성과 편향성 연구," 멀티미디어언어교육 27(4) (2024), pp. 70-86.
doi: https://doi.org/10.15702/mall.2024.27.4.70 (accessed Dec. 27, 2024)
[29] Meta Horizon, https://horizon.meta.com/ (accessed Dec. 27, 2024).
[30] 방준성, 이상민, "소셜 메타버스에서의 디지털 질투의 영향과 측정," 방송공학회논문지 29(5) (2024), pp. 535-545.
doi: https://doi.org/10.5909/JBE.2024.29.5.535 (accessed Dec. 27, 2024)
[31] 방준성, 김보라, 조순정, "인공지능 윤리에 관한 문화다양성 적용의 필요성과 가이드라인 개발 전략," 문화예술융합연구 5(2) (2024), pp. 97-115.
doi: https://doi.org/10.47415/IRAC.5.2.7 (accessed Dec. 27, 2024)

저 자 소 개

방준성

- 2013년 : 광주과학기술원(GIST) 정보통신공학과 공학박사
- 2013년 ~ 현재 : 한국전자통신연구원(ETRI) 디지털융합연구소 책임연구원
- 2016년 ~ 현재 : 과학기술연합대학원대학교(UST) 인공지능학과 교수
- 2022년 ~ 현재 : 한양대학교 과학기술윤리법정책센터 기술전문위원
- 2023년 ~ 현재 : 주식회사 와이매틱스 대표이사/CEO
- ORCID : https://orcid.org/0000-0003-1446-7755
- 주관심분야 : Contextual Computing, AI Safety, Conversational Bot, XR, Computer Vision, Metaverse

이상민

- 2008년 ~ 현재 : 경희대학교 메타버스학과 교수
  한국연구재단 심의위원
- 2023년 ~ 현재 : 교육부 데이터 기반 실증연구 위원
- 2024년 ~ 현재 : Language Learning & Technology 에디터
- ORCID : https://orcid.org/0000-0002-7686-3537
- 주관심분야 : XR, Metaverse, AIEd, Computational Creativity, Machine Translation