

멀티모달 생성형 AI 기반 버추얼 휴먼 관련 기술 및 사업화

□ 오제욱 / 한국버추얼휴먼산업협회

요약

버추얼 휴먼(가상 인간)은 기술 발전과 함께 인간과 구별하기 어려운 외모, 목소리, 움직임을 구현하며 대중과 상호작용하고 있다. 삼성전자의 네온 프로젝트는 이러한 디지털 인류 구현의 기초를 닦았으며, 이후 신디시아, 헤이젠 등 관련 기술이 더욱 발전했다. 생성 AI와 딥페이크 기술은 다양한 콘텐츠 제작에 활용되지만, 딥페이크 범죄와 퍼블리시티권 침해 우려도 제기된다. 버추얼 아이돌 플레이브는 애니메이션 같은 외모와 진정성 있는 소통으로 성공을 거두며, 기술 이상의 인간적 요소의 중요성을 보여주었다. 향후 기술과 인간다움의 조화를 통해 버추얼 휴먼의 발전이 지속될 것으로 전망된다.

1. 서론

2018년 6월 미국 TIME지는 매해 그래왔듯 인터넷에서 가장 영향력 있는 인물 25명을 선정하여 발표했다[1]. 도널드 트럼프, 팝스타 리한나, 가장 수입이 많은 유튜버로 알려진 <Ryan's world>의 라이언, 대한민국의 자랑스러운 BTS까지, 지면을 장식한 25명의 면면은 보는 누구나 고개를 끄덕일 만했다. 단 한 명, Miquela Sousa를 제외하고 말이다. Miquela Sousa, 일명 릴 미켈라(Lil Miquela)로 알려진 이 인물은 어떤 배경으로 25명에 선정된 것인

지 그 설명을 두 번 세 번 다시 읽게 만들었다. 이는 그녀가 바로 실존하는 인물이 아닌 가상의 인물, 버추얼 휴먼이었기 때문이다.

릴 미켈라의 바이럴 이래로 지난 6년여간 전 세계 여러 나라에서 다양한 디지털 존재들이 등장했다. 어떤 존재는 육안으로 구분할 수 없을 정도로 실존 인물과 닮아있고, 어떤 존재는 사람의 목소리를 완벽하게 구사하고, 또 어떤 존재는 사람들 틈에서 자연스럽게 말을 섞으며 뛰어난 지적 능력과 언어 구사력을 보여주었다. 이들의 공통점은 다양한 수단으로 실재하는 인물들과 매우 인격적인 상호작



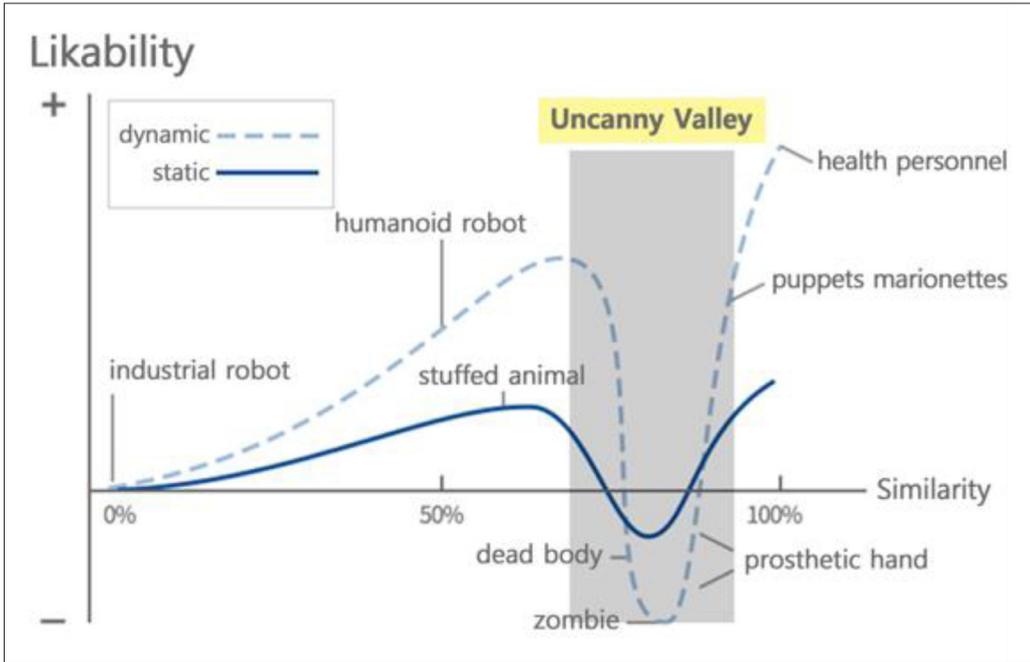
<그림 1> Lil Miquela Courtesy of @lilimiquela (TIME지에 게재된 릴 미켈라의 모습)

용을 주고받는다. 심지어 어떤 존재들은 기존의 인류보다 더 뛰어난 능력을 발휘하거나, 연예인 못지 않은 큰 인기를 누린다. 바야흐로 ‘버추얼 휴먼’이라 불리는 존재들의 시대가 열린 것이다.

II. 본문

2020년 1월 CES에서 삼성전자는 디지털 신인류 프로젝트 네온(Neon)을 선보였다. 미리 공개된 홍보 영상 속

네 명의 인물은 모두 실제 사람의 모습을 하고 있었는데 그중 한 명만이 실제 사람이고 다른 세 명은 가상의 인물, Artificial Human이라는 내용에 전 세계의 눈이 삼성전자의 부스로 쏠렸다. 네티즌들은 영상 속에서 작은 단서라도 찾기 위해 재생 속도를 늦추거나 화면을 확대해 가며 분석에 열을 올렸지만, 아무리 살펴봐도 네 명 모두 완벽하게 실제 사람처럼 보였다. 삼성전자는 실존 인물의 얼굴을 포함한 전신을 스캔한 데이터로 인물의 외관을 재연하고 TTS(Text To Speech) 기술로 음성 데이터를 생성하며, 음성 데이터에 맞게 입술을 비롯한 얼굴 표정과 자연

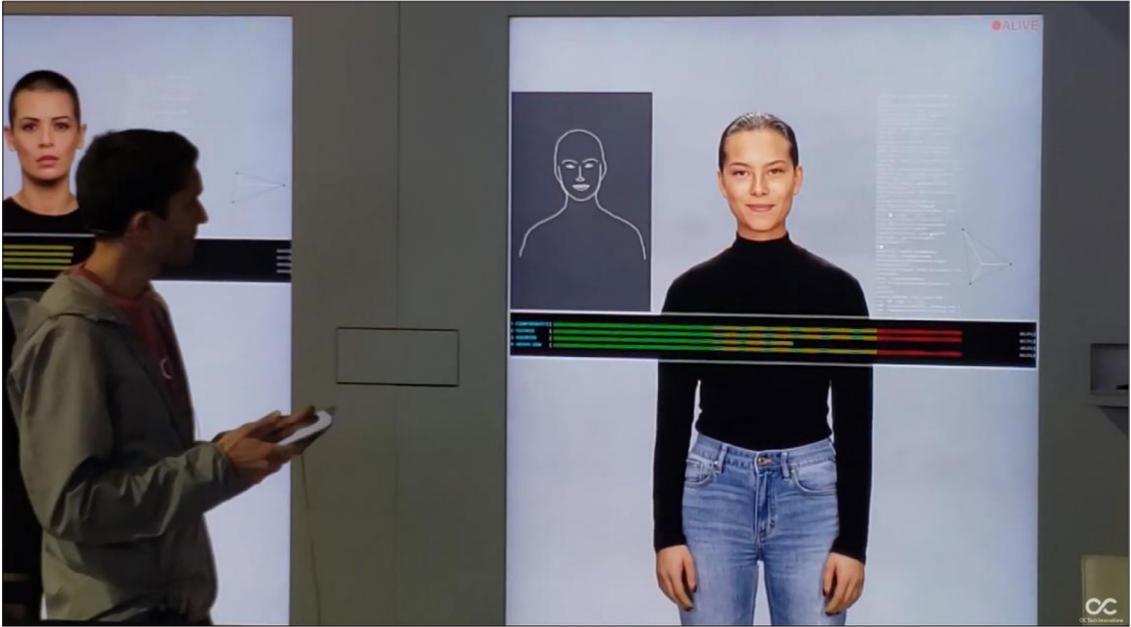


<그림 2> 불쾌한 골짜기 현상(Uncanny Valley) 개념도 @기획재정부

스러운 몸의 움직임에 재연하는 기술로 디지털 신인류, 네온을 탄생시켰다. 이후 삼성전자는 추가로 네온을 활용해 동영상 제작할 수 있는 솔루션 <네온 스튜디오>의 데모 영상을 공개했다. 영상에 따르면 네온으로 동영상을 제작하는 방법은 매우 간단하고 실용적이다. 먼저 희망하는 아바타를 선택하고 미소, 활짝 웃음 등 기본이 될 표정을 선택한다. 이때 볼과 입술, 눈 등 특정 부위의 표정에 대한 커스터마이징 또한 가능하다. TTS 모듈에서 대사를 입력하고 어조와 언어, 발화 속도, 목소리 등 오디오 셋팅을 결정하면 동영상이 생성된다. 아바타의 위치와 크기를 조정하고 배경이나 참고용 이미지 등을 추가하면 동영상 콘텐츠를 완성하게 된다.

인류와 구별이 어려운 디지털 신인류를 구현하려면 먼저 인류가 어떤 존재인지를 규정해야 한다. 인류가 인간을 인간으로 인식하는 데에 필요한 요소는 인간적인 사고, 얼굴의 생김새, 목소리, 몸의 생김새, 표정의 변화, 몸의 움직임 등이 있다. 상대방을 인식할 때 이러한 요소들

이 충족되지 않으면 로봇이나 기계 같다고 느끼거나 심지어 불쾌감을 느끼게 되는데 이러한 현상을 ‘불쾌한 골짜기(Uncanny Valley)’라고 한다. 로봇 공학에서 처음 사용된 이 표현은 인간을 본따 만든 인형, 캐릭터, 로봇 등이 인간과 닮을수록 호감도가 올라가지만, 한 눈에 알아보기 어려울 정도로 인간과 유사해졌을 때에는 완벽한 인간으로 한 눈에 파악되기 전까지는 불쾌한 감각을 불러일으킨다는 이론이다. 네온 프로젝트의 데모 영상을 두고 ‘외계인을 잡아다 가둬놓고 개발시킨 기술’이라는 이야기가 돌았던 이유는 당시의 딥러닝 기술뿐만 아니라 하드웨어 기술 또한 비추얼 휴먼의 불쾌한 골짜기 현상을 완벽하게 극복하기에는 부족함이 많았기 때문이다. 실제로 2020년 1월 CES 현장에서 시연된 네온의 데모 버전은 사고와 발화, 표정과 동작 등 네온을 구성하는 각각의 모달리티(Modality)를 구동하는 모듈들이 하드웨어 자원을 경쟁적으로 점유하면서 버퍼가 일어나고 오디오나 얼굴, 몸의 그래픽 등에서 오류가 발생하는 것이 목격됐다[2]. 이 때문



<그림 3> 시연 중 오류가 발생한 네온 데모 버전 @Samsung StarLabs, YouTube Ch. OC Tech Innovations

에 완벽했던 데모 영상이 오히려 과장된 것이 아니었냐는 의혹이 일어나기도 했었다.

삼성전자의 네온 프로젝트는 아쉽게도 중단되어 네온 스튜디오가 실제로 사용되지는 못했지만 인류의 얼굴과 몸의 생김새, 사고와 발화, 표정과 몸의 움직임까지 육안으로 구별이 어려운 수준까지 재연해내는 데에 성공했다는 점에서 버추얼 휴먼 산업에 있어 중요한 기초를 놓았다고 평가할 수 있다. 네온 스튜디오의 서비스 모델은 이후 신디시아(Synthia)[3]와 헤이젠(HeyGen)[4] 등 디지털 더블 아바타를 개발하는 기업들에 의해 계승되고 더욱 발전되고 있다. 특히 미국 기업인 헤이젠은 선두 기업인 신디시아를 따라잡기 위해 아바타의 립싱크 정확도를 높이고 표정과 몸의 자연스러운 움직임을 재연하는 데에 집중하고 있다. 국내에서도 딥브레인, 이스트소프트, 에이아이파크 등 다수의 기업들이 기술 개발에 박차를 가하고 있다. 국내에서는 네온과 같은 디지털 더블 아바타 기술이 방송사의 버추얼 휴먼 뉴스 앵커를 공급하는 용도로 주로 사용되고 있는데 해외 솔루션은 쉽고 편리한 UI를 제공해

방송사뿐만 아니라 유튜버들과 자체 콘텐츠를 제작하는 여러 기업들에서도 두루 사용되고 있다.

이러한 휴먼 디지털 더블 기술의 상용화에는 목소리 딥페이크 기술의 발전이 큰 기여를 했다. 목소리 딥페이크 기술은 기술적으로는 보통 TTS(Text To Speech) 기술과 보이스 클로닝(Voice Cloning) 기술, 두 가지로 분류된다. TTS는 문장을 입력하면 오디오로 데이터를 변환하여 출력해 주는 기술로 디지털 더블 아바타로 다국어 발화 동영상을 제작하는 데에 사용된다. 보이스 클로닝은 오디오를 특정 인물의 목소리로 변환시켜 주는 기술로 타인의 목소리를 모사하는 성우의 역할을 대신할 수 있는 기술이다. 최근 보이스 클로닝 기술이 누군가의 목소리뿐만 아니라 어기와 어조까지 유사하게 흉내낼 수 있도록 무섭게 발전하면서 딥페이크 범죄에 대한 우려가 커지고 있다. 실제로 2023년 4월 미국에서 딸을 납치했다며 금전을 요구하는 협박 전화를 받은 한 여성의 사연이 국내에도 보도된 바 있다. 결국 협박범들의 납치가 사실이 아니고 전화 통화 속 들려온 울부짖는 딸의 목소리 또한 보이스 클로

닝 기술로 제작된 딥페이크 오디오였음이 확인되었는데, 이때 협박범들이 딸의 목소리를 복제하기 위해 필요로 했던 음성 데이터가 SNS에 올라와 있던 고작 3초에 불과한 분량이었다는 사실이 알려져 SNS에 음성과 얼굴 등을 올리는 것이 과연 안전한지 전 세계 네티즌들에게 경각심을 불러일으켰던 사건이 되었다[5]. 대표적인 목소리 딥페이크 솔루션 서비스로는 일레븐랩스(Eleven Labs)[6]가 널리 알려져 있다.

국내에서는 목소리 딥페이크 기술이 가장 널리 사용된 사례로 AI 커버를 들 수 있다. 음원 속 목소리를 보이스 클로닝 기술로 다른 가수의 목소리로 교체하거나 일반인이 자신의 목소리로 교체하는 방식이다. 많은 사랑을 받은 비비의 <밤양갱>은 AI 커버 또한 다수 제작되었는데 박명수, 악동뮤지션의 이수현, 양희은, 성시경 등 여러 다른 가수들의 목소리로 제작된 음원을 유튜브에서 쉽게 접할 수 있다. 상당수의 네티즌들은 본인이 좋아하는 가수의 목소리로 다양한 곡을 소비할 수 있다는 점에서 이러한 콘텐츠를 거부감 없이 소비하고 또 직접 창작하기도 하고 있지만 정작 목소리를 도용당한 가수들과 원곡자들은 불편한 감정을 표현하기 시작했다. 박명수는 라디오 방송 중

이 사안에 관해 당혹스러운 심경을 숨기지 않았고 고 김광석 등 고인의 목소리까지 AI 커버에 무분별하게 활용되는 점에 사자에 대한 예의와 퍼블리시티권 침해 이슈 또한 불거지고 있다[7].

인류의 외관을 구현하는 것은 대표적인 이미지와 영상 생성 AI 서비스인 스테이블 디퓨전과 미드저니 등을 통해 고도화되고 있다. 특히 미드저니는 2024년 5월 그간 불완전했던 인물의 외모 일관성 유지 기능을 크게 향상시킨 업데이트를 발표했다. 프롬프트를 동일하게 입력해도 인물의 외모가 달라지는 경우가 많아 버추얼 휴먼의 외모 일관성을 담보할 수 없었던 이전 버전과 달리 이번 업데이트 이후의 미드저니는 누구나 가상의 인물을 생성해 내고 그 인물의 서사를 영구적으로, 또 무제한적으로 써 내려갈 수 있게 되었다[8]. PikaLabs[9], Runway[10] 등 동영상 전문 생성 AI 서비스들 또한 프롬프트로 동영상을 생성하는 것은 물론, 특정 이미지나 동영상으로부터 인물의 외관 또는 질감(texture), 스타일 등을 레퍼런스로 추출하여 동영상을 재구성하는 솔루션 등을 경쟁적으로 업데이트하며 빠르게 발전하고 있다.

필자의 디오비스튜디오[11]는 프롬프트를 기반으로 모



<그림 4> 미드저니로 생성한 가상의 인물들 created by Learning Experience Design Consultant Christy Tucker @Christy Tucker



<그림 5> TVING, JTBC 방송 <얼라이브(Alive)> 방송 화면 캡처 @TVING, JTBC

든 이미지를 생성해 내는 방식과 달리 실존 인물을 촬영한 후 얼굴 부분만 AI로 교체하는 딥페이크 기술을 고도화해 왔다. 2020년 10월 디오비스튜디오의 첫 버추얼 휴먼 루이(Rui)는 숙련된 아이돌 지망생을 쉐도우 액터(Shadow Actor)로 고용하여 SNS에서 활동하는 음악 인플루언서로

활동했다[12]. 감쪽같이 구현된 고품질 가상 얼굴과 그녀의 노래 실력, 통통 튀는 매력에 네티즌들은 열광했고 한국관광공사와 문화체육관광부까지 그녀를 명예 홍보대사로 위촉하면서 매력적이면서 사고 칠 염려도 없는 버추얼 휴먼을 십분 활용했다[13]. 루이를 제작할 때 적용한 가상



<그림 6> 2023년 디오비스튜디오가 재현한 찰스 다윈과 마릴린 먼로 @dob Studio

얼굴 생성 기술은 데이터셋의 전처리 방식과 얼굴 생김새의 유사도를 평가, 각 부분을 개별적으로 변경할 수 있는 기능 등을 추가하며 발전을 거듭하여 이제는 레퍼런스 얼굴을 빠르게 복원하고 동영상에 적용할 수 있는 가상 얼굴로 생성해낼 수 있게 되었다. 이러한 딥페이크 기술을 기반으로 2022년 1월 공개된 <얼라이브(Alive)>에서 선보인 고 임윤택의 모습은 유가족과 팬들은 물론 전 세계 네티즌들에게 큰 감동을 선사했다[14].

특정 얼굴을 초고화질로 재구현할 수 있는 기술은 최근 작고하신 고인을 복원하여 유가족을 위한 좀 더 생생한 추모와 기념의 기회를 제공할 뿐만 아니라 수십, 수백 년 전 추억의 셀럽들, 역사 속 위인의 복원 또한 가능하게 했다. 오스카상을 수상한 국민 여배우 윤여정님의 50년 전 모습을 완벽하게 복원해 찬사를 받은 KB라이프생명의 광고로 디오비스튜디오는 대한민국 광고대상 금상을 수상하기도 했다[15]. 또한 찰스 다윈과 마릴린 먼로를 복원하여 제작한 TV 다큐멘터리는 역사 속 셀럽과 위인이 현재와 미래에 어느 때든 생전의 완벽한 모습으로 우리 곁에 돌아올

수 있다는 것을 증명한 사례가 되었다.

앞서 내온을 필두로 버추얼 휴먼이 구현하고자 했던, 인류가 인간을 인간으로 인식하는 데에 필요한 요소를 다시 한 번 짚어보자. 인간적인 사고, 얼굴의 생김새, 목소리, 몸의 생김새, 표정의 변화, 몸의 움직임 등 SNS를 비롯한 디지털 시공간에서 접할 수 있는 버추얼 휴먼의 이러한 구성 요소들은 생성 AI 기술과 응용 솔루션들의 빠른 발전으로 인해 높은 완성도로 구현되고 있다. 하지만 정작 대중들에게 보다 장기적이고 보다 뜨거운 관심과 찬사를 받는 버추얼 휴먼은 이러한 요소들에 있어서 기존의 가설을 뒤집어 엎어버렸고, 그간 인류를 최대한 완벽하게 모사하려고 애썼던 버추얼 휴먼 콘텐츠 창작자들에게 적지 않은 충격과 영감을 선사해 주고 있다. 애니메이션 캐릭터의 모습으로 쟁쟁한 KPOP 아이돌들과 어깨를 나란히 하며 예술적, 상업적인 성공과 함께 차세대 한류의 주역으로 불리우는 버추얼 아이돌 <플레이브>가 그 주인공이다.

플레이브(PLAVE)는 2023년 3월에 데뷔한 5인조 남성 버추얼 아이돌 그룹으로, 리더 예준을 비롯해 노아, 밤비,



<그림 7> 인기 버추얼 아이돌 그룹 플레이브(Plave) @VLAST

은호, 하민으로 구성되어 있다. 이들은 버추얼 아이돌 그룹 최초로 음악방송 프로그램에서 1위를 차지하였고, 첫 단독 콘서트 티켓을 예매 시작 10분 만에 전석 매진시키는 등 선풍적인 인기를 끌고 있다[16]. 여의도 더 현대에서 진행된 팝업스토어에서는 어느 브랜드의 평균 매출을 두 세 배 이상 상회하는 높은 매출을 올린 것으로 알려졌고, 유명 코스메틱 브랜드의 모델로 발탁되기도 했다. 또, 음악 스트리밍 서비스 멜론에서는 ‘멜론의 전당’ 빌리언스 클럽에 역대 최단 기록으로 이름을 올렸다. ‘멜론의 전당’은 앨범 발매 후 24시간 이내에 1백만 회 이상의 스트리밍을 기록한 경우 등재되는 ‘밀리언스 클럽’과, 데뷔 후 누적 10억 회 이상의 스트리밍을 기록한 경우인 ‘빌리언스 클럽’으로 나누어지는데 빌리언스 클럽에 등재된 아티스트는 2024년 7월 기준 총 109개 팀밖에 되지 않는다. 멜론에서 스트리밍을 서비스하는 아티스트 3백만 팀 중에서 플레이브가 109팀 중의 하나로, 그것도 역대 최단 기간인 데뷔 후 494일 만에 등재된 것이다[17].

플레이브의 성공은 분명 육안으로 구별이 어려울 정도로 인류와 닮은 얼굴과 몸 덕분은 아님이 틀림없다. 누가 봐도 사람이라기보다는 애니메이션 캐릭터일 뿐인 그들은 심지어 최첨단의 AI 기술로 구현된 것도 아니다. (물론 플레이브를 제작한 블래스트(VLAST)의 3D 모델링, 모션 캡처, 리얼타임 렌더링을 비롯한 그래픽 기술은 뛰어나고 독창적이다. 다만, 전 세계가 열광하는 생성 AI나 외계인을 잡아다 놓고 고문하며 만들어낸 기술이라고 평가받는 기술은 아니라는 점을 부각하고자 하는 이야기이다.) 평론가들과 대중은 플레이브의 성공 비결을 기술이 아닌 멤버들의 음악적인 재능과 노력, 팬들을 진심으로 대하는 그들의 인간적 됴됨이에서 찾는다. ‘버추얼 휴먼’으로 리스크는 줄이고, 생산성은 높이고, 높은 부가가치의 뉴노멀 메타버스 콘텐츠 산업을 기대하는 모든 자본가, 기업가들

과 창작자들에게 플레이브는 앞으로 버추얼 휴먼을 제작할 때에, 기술뿐만 아니라 기술 너머의 핵심이 되는 무언가를 반드시 우선적으로 고려해야 한다는 점을 명확히 보여주고 있다.

III. 결론

인류를 닮은 인공의 존재를 만들고자 하는 노력은 딥러닝 이전 로봇 공학에서도, 또 더 이전 문학과 설화, 상상력의 영역에서부터 지속되어 온, 인류에게는 익숙하고 자연스러운 시도이다. 앞으로도 우리는 버추얼 휴먼을 만들기 위해 다양한 기술적, 예술적 시도를 지속할 것이고, 챗GPT와 미드저니, 일레븐랩스와 같은 생성 AI 기술의 도움을 받아 아날로그 인류의 수를 넘어서는 셀 수 없이 많은 디지털 페르소나(Persona)들이 등장하고, 활동하고, 소멸될 것이다. 플레이브에 열광하는 젊은 세대는 그들처럼 일부 요소만 AI와 디지털로 변환한 트랜스휴먼(Transhuman)들 뿐만 아니라 오감으로 분간이 어려운 요소들을 구비한 그럴싸한 AI 디지털 로봇들에게도 관심을 기울이고, 그들과 인격적인 상호작용을 주고받으며 관계를 형성하고 서사를 만들어 갈 것이다. 그렇기에 우리는 기술뿐만 아니라 기술 너머의 ‘인간됨’의 핵심을 함께 고려하며 ‘버추얼 휴먼’에 관한 전방위적인 담론을 공론화하고 세대와 국경을 넘어 우리의 새 이웃이 된 디지털 신인류를 어떻게 대할 것인지, 어떻게 어울려 살아갈지를 결정해야 한다. 그들이 바로 우리 자녀 세대의 친구가 되고, 스승이 되고, 심지어 연인이 될 수도 있기 때문이다. 필자가 종사하는 콘텐츠 산업을 넘어 교육계와 종교계, 정치를 비롯한 사회 전반에서 버추얼 휴먼을 중심으로 한 활발한 논의가 시작되기를 기대한다.

참 고 문 헌

- [1] The 25 Most Influential People on the Internet (2018. 6. 28), <https://time.com/5324130/most-influential-internet/>
- [2] NEON, AI based human interaction (2020. 1. 12), <https://youtu.be/zAv5UgxOREk>
- [3] Synthesia, <https://www.synthesia.io/>
- [4] HeyGen, <https://www.heygen.com/>
- [5] “엄마, 살려 줘!” 내 딸 목소린데... 알고 보니 AI 보이스피싱 (2023. 4. 16), <https://www.mk.co.kr/news/society/10713585>
- [6] ElevenLabs, <https://elevenlabs.io/>
- [7] “연예인들 어떻게 하라고” 소름 돋는 ‘이것’...연예계 발칵 뒤집더니 (2024. 7. 4), <https://www.seoul.co.kr/news/newsView.php?id=20240704500085>
- [8] Create Consistent Character Images in Midjourney, <https://christytuckerlearning.com/create-consistent-character-images-in-midjourney/>
- [9] PikaLabs, <https://pika.art/>
- [10] Runway, <https://runwayml.com/>
- [11] 디오비스튜디오 홈페이지, <https://www.dob.world/kr>
- [12] 유튜브 <루이커버리(RuiCovery)> 채널, <https://www.youtube.com/@RuiCovery>
- [13] “가상 인간이 문체부 누리소통대시를?” ‘버추얼 휴먼’ 루이가 전하는 ‘문화강국’ 대한민국..., <https://blog.naver.com/mcstkorea/222542698931>
- [14]故유재하·임윤택 소환한 ‘얼라이브’, 뉴미디어 콘텐츠상 대상 수상, <https://www.dailian.co.kr/news/view/1149965/?sc=Naver>
- [15] AI 기술로 재탄생한 윤여정의 20대 모습, https://youtu.be/r9E0bfQq4gg?si=SPVZF_ZN-NKZ2uJa
- [16] 나의 ‘BUT, YOUR 아티스트’ ... 버추얼아이돌그룹 PLAVE 인기 이유는?, <https://www.d-art.co.kr/news/articleView.html?idxno=625>
- [17] 플레이브, 멜론 ‘빌리언스 클럽’ 최단 기록 입성, https://sports.khan.co.kr/article/202407180924003/?utm_source=urlCopy&utm_medium=social&utm_campaign=sharing

저 자 소개



오 제 목

- 2003년 : 연세대학교 중어중문학, 국어국문학 전공 학사
- 2020년 : 가톨릭대학교 글로벌한류비즈니스학과 석사
- 2020년 ~ 2024년 12월 : 디오비스튜디오 대표이사
- 2024년 7월 ~ 현재 : 한국버추얼휴먼산업협회 공동협회장
- 2024년 12월 ~ 현재 : 디오비스튜디오 고문
- 주관심분야 : 버추얼 휴먼, SNS, 문화 콘텐츠, 미디어 리터러시