논문 2025-2-11 http://dx.doi.org/10.29056/jsav.2025.06.11

생성형 인공지능이 생성한 얼굴 표정의 활용성 분석

지은미*, 심윤식*†

Analysis of the usability of facial expressions generated by generative artificial intelligence

Eun-Mi Ji*, Yoonsik Shim*†

요 약

본 논문에서는 현재 개발된 LLM 모델에 기반한 생성형 인공지능이 아바타 형태로 사람과 대화할 때, 문맥에 맞게 아바타가 적합한 표정을 갖는 아바타를 생성할 수 있는지 분석하고자 한다. 또한 생성된 아바타의 표정이 다양한 성별, 표정, 나이, ID 별로 기존 표정 인식 데이터와 비교 분석하여 적합하게 생성되었는지 활용성을 평가하고자 한다. 마지막으로 이러한 비교 평가는 기존에 존재하는 다양한 인공지능 간의 결과물의 차이와 특징을 비교하고 분석함으로써, 현재 개발된 인공지능의 표정 생성 능력에 대한 활용 가능성을 제시하고자 한다.

Abstract

This paper aims to analyze whether generative AI based on currently developed large language models (LLMs) can generate avatars that display appropriate facial expressions in context when interacting with humans in avatar form. Furthermore, it evaluates the usability of the generated facial expressions by comparing them with existing facial expression recognition datasets across various categories such as gender, expression type, age, and identity. Lastly, this comparative evaluation seeks to identify and analyze the differences and characteristics among the results of various existing AI models, thereby assessing the potential applicability of current AI systems in generating realistic and contextually appropriate facial expressions.

한글키워드 : 생성형 인공지능, 아바타, 표정 생성, 표정 인식, 딥러닝

keywords : Generative-AI, avatar, expression generation, expression recognition, deep learning

1. 서 론

최근 ChatGPT[1]와 같은 LLM 모델에 기반한 생성형 인공지능 성능이 급속히 발전함에 따라 인간과 인공지능 간의 일상적인 대화가 가능하게 되었다. 이에 따라 인공지능이 가상의 아바타를 생성하여, 자연스럽게 인간과 대화하는 GUI가 머지않아 현실화할 것으로 기대된다. 인간과 인간 사이의 대화(communication)란 생각, 정보, 감정 등을 다른 사람과 주고받는 과정을 의미한다.

* 배재대학교 게임멀티미디어공학과
† 교신저자: 심윤식(email: ysshim@pcu.ac.kr)

접수일자: 2025.06.03. 심사완료: 2025.06.13.
게재확정: 2025.06.20.

이를 위해 단순히 주고받는 문맥뿐만 아니라 제스처나 표정 등이 실제 의미 전달에 중요한 역할을 담당한다. 마찬가지로 인간과 인공지능이 대화할 때, 자연스러운 대화를 위해선 인공지능이 생성한 아바타가 의미와 맞는 적합한 표정을 표현할 수 있어야 한다.

그러므로 본 연구에서는 최신의 생성형 AI 도구들을 이용해 다양한 표정을 생성하고, 생성된 표정 이미지의 퀄리티 차이를 성별, 표정, 나이, identity 별로 분석하여 생성된 표정이 실제 대화 아바타로 활용 시 얼마나 유용한지를 분석하고자 한다.

이를 위해 표정 생성을 위해 사용된 대표적인 4종류의 생성형 AI 도구는(ChatGPT[1], Co-Pilot[2], Gemini[3]. Pincel[4])이고, 일반적으로 표정 인식할 때 사용되는 무표정 및 대표적인 6가지 표정[5](기쁨, 슬픔, 화남, 놀람, 공포, 혐오)을 생성하고, 각 생성형 AI 도구 간의 생성된 표정 이미지들의 특성을 파악하고 장단점을 비교할 것이다.

일반적으로 생성형 이미지 모델에서 생성한 결과물을 평가할 때는 기존 방식과는 다른 기준을 적용한다. 가장 많이 사용하는 기준은 시각적 품질(visual quality)이나 사실성(realism), 프롬프트 일치율(prompt alignment) 등의 기준이 있으며. 최근에는 공평성(fairness)이나 공격성(toxicity) 같은 AI 안전성(safety) 요소를 평가 내용에 포함하기도 한다[6]. 이러한 생성형 모델의 성능을 평가하는 방법은 크게 정성적 방법과 정량적 방법으로 나눌 수 있다. 정성적 방법은 사람이 자신의 주관적 기준으로 결과물을 검토해 성능을 결정하는 방법이다. 이 방법은 사람이 직접 결과물을 평가하기 때문에 정량적 평가가 반영하지 못하는 세부적인 품질 차이까지 반영할 수 있다는 장점이 있으나, 각 사람의 기준에 따라 결과가 다르게 나올 수 있어 일관성을 확보하기 어렵다는 단점이 있다. 또한 무엇보다 여러 사람이 직접 확인해야 한다는 점에서 시간과 비용이 많이 소모된다. 이에 반해 정량적 방법은 일관적인 평가 기준을 만들어 이에 따라 모델의 성능을 객관적인 지표와 수치로 평가하는 방법이다. 이 방법은 숫자로 표현할 수 없는 부분은 평가 시 고려하기 힘들다는 단점이 있지만, 결과가 일관적이어서 모델 간 비교가 용이하고, 효율적으로 모델의 성능을 평가할 수 있다는 장점이 있다. 본 연구에서는 정량적 평가 방법으로 현재 뛰어난 성능을 보이는 기 개발된 인공지능 기술을 이용해서 생성형 모델의 성능을 객관적인 지표와 수치로 평가하는 방법을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 최신의 표정 생성 및 표정 인식에 관한 연구 내용을 기술한다. 3장에서는 생성형 AI 도구 기반 표정 생성 결과를 기술하며, 4장에서는 생성된 표정의 정량적 분석 방법 및 결과를 제시한 후, 마지막 5장에서 결론을 기술한다.

2. 표정 생성 및 표정 인식

2.1 표정 생성

표정 생성(Facial Expression Generation)은 주어진 얼굴 이미지에 특정 감정을 반영하여 새로운 표정을 합성하는 기술로, 가상 캐릭터 애니메이션, 감정 인공지능, 게임, 증강 현실, 딥 페이크 탐지 연구 등 다양한 응용 분야에서 중요성이 점점 커지고 있다.

딥러닝이 등장하기 전에는 표정 생성을 위해 주로 3D 얼굴 모델링 또는 모핑(morphing) 기술을 활용해 표정을 생성하였다. 대표적인 방법으로는 3D Morphable Model (3DMM)[7], Blendshape 모델링[8], Facial Action Coding System (FACS) 기반 애니메이션[9]이 있다.

이러한 방법들은 정밀한 랜드마크 위치와 근육 기반 제어를 통해 다양한 표정을 생성할 수 있으며, 영화 및 게임 산업에서 여전히 널리 사용된다. 그러나 이들 방식은 고해상도 모델링과 전문적인 조정이 필요하며, 학습 기반 자동화에는 한계가 있었다.

한편, Autoencoder 및 Variational Autoencoder(VAE)[10]와 같은 생성형 모델 기반의 접근도 활발히 연구되었다. 특히 VAE 기반 모델은 분포 학습 측면에서 이점을 가지지만, 생성 이미지의 선명도나 품질 면에서는 GAN 기반 방법에 비해 다소 떨어질 수 있다. 또한, 얼굴 랜드마크나 3D 정점(vertex) 정보를 조건으로 사용하는 geometry-guided generation 방식[11]도 활용되고 있다. 이러한 방법은 표정 생성 과정의 해석 가능성을 높이고, 특정 영역(예: 눈, 입)의 제어를 가능하게 한다는 점에서 유용하다.

딥러닝의 발전과 함께 GAN(Generative Adversarial Network)[12]의 등장은 고해상도 이미지 생성을 가능하게 하며 표정 생성 분야에서도 중요한 전환점을 제공하였다. GAN 기반 모델은 일반적으로 생성기(generator)와 판별기(discriminator)로 구성되어, 생성기가 입력된 얼굴에 새로운 감정을 입힌 이미지를 만들어내고, 판별기는 그것이 실제 같은지 판별하면서 양자 간의 경쟁을 통해 성능을 향상시킨다. StarGAN[13]은 다중 감정 레이블을 조건으로 사용하여 다양한 표정을 하나의 모델에서 생성할 수 있도록 하였고, 이 외에도 감정의 강도까지 조절 가능한 ExprGAN[14], 조건부 GAN(cGAN)[15] 기반의 다양한 변형 모델들이 제안되면서 표정 생성의 정밀성과 다양성이 크게 확장되었다.

최근에는 다양한 LLM(Large Language Model)들과 GAN 기반 모델들의 결합으로 상당히 정밀한 표정을 생성하고, 이를 출력으로 제시한다. 본 논문에서는 최신의 대표적인 생성형 AI 모델을 이용하여 무표정과 6가지 대표 표정을 생성하고 이를 분석하고자 한다.

2.2 표정 인식

표정은 인간의 감정을 전달하는 가장 직관적이고 보편적인 비언어적 신호 중 하나로, 사람 간의 의사소통에서 핵심적인 역할을 한다. 이러한 인간의 표정을 자동으로 인식하는 기술인 표정 인식(Facial Expression Recognition, FER)은 인간-컴퓨터 상호작용(HCI), 감정 기반 사용자 인터페이스, 의료 및 보안 분야에서 중요한 기술로 간주된다. FER의 궁극적인 목표는 정적인 이미지 혹은 동영상 내의 얼굴로부터 인간의 감정을 자동으로 인식하는 것이다.

딥러닝의 부상과 함께 FER 분야 역시 보다 강력하고 정교한 데이터 기반 학습 모델인 CNN(Convolutional Neural Network)을 중심으로 급속한 진화를 이루었다. CNN은 얼굴 이미지로부터 의미 있는 특징을 자동으로 추출하고 이를 바탕으로 감정 클래스를 분류한다. VGGNet[16], ResNet[17], Inception[18] 등의 구조가 FER에 적용되어 우수한 성능을 보였으며, FER2013[19], RAF-DB[20] AffectNet[21], CK+[22]와 같은 대규모 공공 데이터셋을 통해 효과적으로 학습되었다. 이러한 모델들은 과거의 수작업 특징 기반 방식보다 뛰어난 인식 능력을 가지지만, 여전히 데이터 편향과 일반화 문제를 안고 있다. 최근에는 Transformer 기반 모델과 Attention 메커니즘[23]이 도입되어 FER의 성능을 더욱 향상시키고 있다. Self-Attention 구조는 이미지의 전역적인 문맥을 효과적으로 포착할 수 있으며, 특정 표정 특징에 집중함으로써 미세한 감정 차이를 구분하는 데 도움을 준다. 특히 Spatio-temporal Attention[24]은 동영상 기반 FER에서 시간 및 공간의 중요한 정보를 동시에 고려할 수 있도록 한다.

FER 연구에 사용되는 대표적인 데이터셋은 표 1과 같다. FER2013은 가장 널리 사용되는 정적 이미지 기반 데이터셋으로, 48x48 해상도의 흑백 얼굴 이미지와 7개의 기본 감정 라벨을 제공한다. RAF-DB는 실세계 이미지와 더욱 정교한 주석 정보를 포함하며, AffectNet은 100만 개 이상의 표정 이미지와 함께 arousal-valence 차원의 연속 감정 라벨을 포함한다. CK+는 실험실 환경에서 촬영된 고품질 동영상 시퀀스를 제공하며, 표정의 시작부터 정점까지의 변화 과정을 포함하고 있다. 마지막으로 국내에는 AI 허브[25] 싸이트에 한국인 감정인식을 위한 복합 영상 표정 인식 데이터 Set이 존재한다.

표 1. 표정인식 데이터 셋
Table 1. results of evaluation

Dataset	Type	Size	Emotion
FER2013	Static	35,887	7
RAF-DB	Static, real-world	~30,000	7
AffectNet	in-the-wild	1M+	8
CK+	Video clips	593	7
AI-Hub	Static	500,000	7

앞서 살펴본 바와 같이, 표정 인식을 위해 다양한 데이터셋과 기술적 접근 방법이 제안되어 왔으며, 본 논문에서는 이러한 표정 인식 기술을 활용하여 생성형 AI가 생성한 얼굴 이미지에 표현된 감정을 정량적으로 분석하고자 한다. 이를 위해 얼굴 검출(face detection), 랜드마크 추출(face landmark estimation), 표정 인식(facial expression recognition), 연령 및 성별 추정(age and gender estimation), 개인 식별(identity recognition) 등 다수의 얼굴 분석 기능에서 실용적인 정확도를 보이면서도 웹 환경에서의 적용에 최적화된 일체형 라이브러리인 face-api.js[26]를 활용하였다. 생성형 AI가 생성한 특정 표정의 얼굴 이미지에 대해 해당 라이브러리를 통해 표정 인식 결과를 도출하고, 이를 비교·측정함으로써 이미지 생성 모델이 표정 및 감정 인식 측면에서 어느 정도의 정량적 정확도를 보이는지 확인하고자 한다.

3. 생성형 AI기반 표정생성

본 연구에서는 LLM을 통한 표정 생성이 가능한 대표적인 글로벌 IT 기업 서비스들인 ChatGPT(OpenAI), Gemini(Google), Co-pilot (Microfost)와 추가로 표정 생성에 특화된 서비스를 제공하는 Pincel을 이용하여, 표정 인식에 사용되는 대표적인 7가지의 기본 감정 표현[5](무표정, 노멀, 기쁨, 슬픔, 놀람, 공포, 화남, 혐오)을 생성하여 아래 그림 1과 그림 2와 같이 제시하였다. 이때 생성형 AI에 입력된 프롬프트는 그림 1은 “20대 남성의 무표정, 기쁨, 슬픔, 화남, 놀람, 공포, 혐오 표정을 생성해 줘“ 이고 그림 2는 “20대 여성의 무표정, 기쁨, 슬픔, 화남, 놀람, 공포, 혐오 표정을 생성해 줘“ 이었다. 다만, 이 4개의 생성형 AI 도구 모두 7개의 표정을 한번의 프롬프트로는 생성하지 못하며, 한번에 하나의 표정을 생성하도록 요청했을 때 원하는 표정이 생성되었다. 이때 Gemini로 생성된 표정 얼굴은 동일인이 아닌 다른 사람의 얼굴로도 표정이 생성되었다. Pincel의 경우는 사용자로부터 입력받은 참조(reference) 표정을 기반으로 표정을 생성하는 기능을 활용하였는데, 입력 표정으로 AI가 생성한 이미지를 사용하였으며, 혐오 표정 생성은 지원하지 않아 결과를 표시하지 못했다. 또한, 그림 1, 2에서 생성한 표정을 좀더 다양한 인종이나 나이에서 생성하도록 프롬프트를 제시하였으나, 대부분의 모델들이 인종이나 나이 변화에 적합한 표정을 생성하지는 못했다.

	무표정	기쁨	슬픔	화남	놀람	공포	혐오
ChatGPT							
Co-pilot							
Gemini							
Pincel							None

그림 1. 대표적인 4종류의 생성형 AI로 생성된 남성 표정 얼굴
Fig. 1. Male Facial Expressions Generated by Four Representative Generative AIs

	무표정	기쁨	슬픔	화남	놀람	공포	혐오
ChatGPT							
Co-pilot							
Gemini							
Pincel							None

그림 2. 대표적인 4종류의 생성형 AI로 생성된 여성 표정 얼굴
Fig. 2. Female Facial Expressions Generated by Four Representative Generative AIs

4. 생성된 표정의 분석

본 장에서는 앞 절에서 생성된 얼굴 표정이 얼마나 정확히 생성되었는지에 대한 분석을 수행하고자 한다. 앞서 설명한 바와 같이 생성된 표정을 평가하기 위한 방법으로 정성적 평가(사람

이 수기로 분석)과 정량적 평가(기계적으로 계산된 수치 평가)로 나눌 수 있다.

본 연구에서는 정량적 평가 방법을 제안하였으며 이를 위해 사람이 미리 평가한 대용량 학습 데이터에 기반해 얼굴의 모든 정보를 정밀하게 인식하는 인공지능에 기반한 face-api.js 라이브러리 싸이트[26]에 접속하여 생성형 모델이 생성한 얼굴 표정을 입력으로 제시한 후 출력으로 제시된 수치를 비교 평가하는 방법을 사용한다.

표정 인식을 위한 face-api.js 라이브러리는 얼굴 전체 이미지 입력 → 얼굴 검출 → 얼굴 crop → 표정 CNN → 감정 분류 확률 결과 도출의 과정을 거치며, 각 단계마다 모델 경량화와 성능의 균형을 고려한 효율적인 서브모듈을 적용하고 있다.

일반적으로 7가지 기본 감정(무표정, 기쁨, 슬픔, 놀람, 공포, 화남, 혐오) 분류를 목표로 하는 face-api.js의 표정인식 시스템은 FER2013[19] 및 AffectNet[21] 등의 공개 데이터셋으로 학습된 CNN 기반의 다중 클래스 분류 모델을 적용한다. 이 과정에서 표정 이미지를 경량화된 합성곱 신경망에 입력하여 특징 벡터를 추출한 후, fully-connected layer 및 softmax 함수를 통해 각 감정 클래스에 대한 확률 출력을 생성한다[27]. 나이 추정 역시 회귀 기반 CNN 모델을 사용하며, 입력된 얼굴 이미지로부터 나이와 성별의 연속값 및 확률을 추정한다. 본 모델은 IMDB-WIKI[28] 등 대규모 얼굴 나이 데이터셋을 활용하여 사전 학습된 것으로, soft-label regression 및 ordinal regression 기반의 나이 추정 모델링 기법이 적용된다[29]. 개인 식별 및 유사도 계산은 얼굴 특징 벡터(face descriptor)를 생성하는 임베딩 모델을 통해 이루어지는데, dlib 및 FaceNet[30] 구조에서 영향을 받은 128차원 얼굴 임베딩을 생성하여 두 얼굴 간의 유사도를 L2 distance 기반으로 계산한다. 이러한 딥러닝 기반 임베딩 방식은 기존의 고전적인 PCA, LBP 기반 방법보다 높은 식별 성능과 변형 불변성을 제공한다.

face-api.js 라이브러리 싸이트에 접속하면 그림 3와 같은 GUI를 제공하며, 입력으로 얼굴 이미지를 파일로 로딩하면, 출력으로 표정 인식 정확도 스코어, 나이 인식 정확도 스코어 및 identity 인식 유사도 스코어를 측정할 수 있다.

먼저 표정 인식의 정확도 스코어를 측정하기 위해, 그림 1과 2에서 4개의 생성형 AI가 생성한 6가지 표정 얼굴에 대해 정확도를 측정한 결과를 표 2에서 볼 수 있다. 표 2에서 4종류의 AI 생성 도구 중 Pincel은 혐오 표정을 생성하지 못해 None으로 표시되어 있어, 정확도 값을 산정하지 못했다. 표 2의 정확도 값은 0~1 사이의 값을 가지며 1에 근사할수록 표정이 잘 생성된 것을 의미한다. 표 3은 표 2와 같은 방법으로 기 생성된 표정에 대해 나이를 측정한 정확도 결과를 보여주며, 표 4는 기 생성된 표정 얼굴들이 identity가 유지되는지를 0~1 사이의 유사도 측정치로 표시한다.

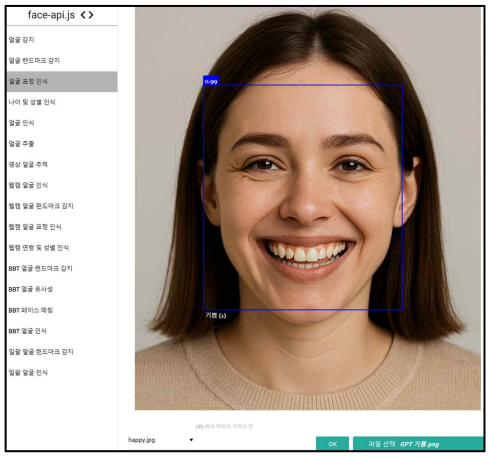

그림 3. 생성형 AI의 결과물을 평가하기 위한 AI 인식 결과(face-api.js 싸이트[26])
Fig. 3. AI recognition results for evaluating the outcomes of Generative AI

표 2의 표정 인식 유사도 결과를 분석해 보면, 4종류의 생성형 AI 모델중 비교적 표정을 잘 생성한 모델은 평균값이 높게 측정된 ChatGPT, Co-pilot 모델임을 알 수 있다. Gemini는 여자의 경우는 표정을 잘 생성하나, 남자의 표정은 공포나 혐오 표정은 잘 생성하지 못했음을 알 수 있다. 마찬가지로 혐오 표정은 Co-pilot이 남녀 구분 없이 잘못 생성하고 있음을 알 수 있어, 아직은 생성형 AI 모델이 다른 감정에 비해 공포나 혐오 표정과 같은 부정적인 표정 생성에 제한을 둔 것을 알 수 있다.

표 2. 표정 인식 유사도
Table 2. results of facial expression

AI	성별	기쁨	슬픔	화남	놀람	공포	혐오	평균
ChatGPT	남	1.00	0.99	0.99	0.99	0.98	0.99	0.99
	여	0.99	0.99	0.99	0.91	0.99	0.99	0,98
Co-pilot	남	1.00	0.99	0.97	0.99	0.99	0.00	0.82
	여	1.00	1.00	0.99	0.97	0.99	0.00	0.83
Gemini	남	1.00	0.93	0.99	0.99	0.06	0.00	0.66
	여	1.00	0.98	0.99	1.00	0.99	0.97	0.99
Pincel	남	1.00	0.00	0.00	0.99	0.00	-	0.33
	여	1.00	0.00	0.00	0.99	0.00	-	0.4

표 3. 나이 인식 유사도
Table 3. results of age estimation

AI	성별	기쁨	슬픔	화남	놀람	공포	혐오	평균/ 표준편차
ChatGPT	남	23	36	35	29	39	54	36.0/9.59
	여	22	24	28	25	25	31	25.8/2.91
Co-pilot	남	36	44	34	30	33	35	35.3/4.31
	여	24	22	33	26	35	28	28.0/4.65
Gemini	남	26	21	25	24	29	32	26.2/1.85
	여	25	22	25	25	24	31	25.3/2.75
Pincel	남	28	20	21	31	27	-	28.4/1.85
	여	21	20	21	23	21	-	21.2/0.98

표 4. identity 인식 유사도
Table 4. results of identity estimation

AI	성별	기쁨	슬픔	화남	놀람	공포	혐오	평균/ 표준편차
ChatGPT	남	0.51	0.45	0.45	0.51	0.36	0.5	0.46/0.03
	여	0.52	0.40	0.40	0.50	0.30	0.55	0.50/0.09
Co-pilot	남	0.34	0.39	0.35	0.41	0.38	0.40	0.38/0.03
	여	0.48	0.48	0.00	0.43	0.00	0.41	0.30/0.21
Gemini	남	0.55	0.49	0.00	0.46	0.59	0.00	0.35/0.25
	여	0.51	0.57	0.55	0.55	0.59	0.00	0.46/0.21
Pincel	남	0.28	0.26	0.56	0.44	0.33	-	0.37/0.11
	여	0.00	0.49	0.00	0.56	0.49	-	0.30/0.25

표 3의 나이 인식 유사도 결과를 분석해 보면, 대부분의 AI 모델이 20~30대 연령을 잘 유지하고 있음을 알 수 있다. 다만, ChatGPT와 Co-pilot, Gemini의 경우 생성된 혐오 표정 이미지에서는 나이 추정 값이 급격히 상승하는 경향을 보였으며, Co-pilot의 공포 표정 이미지에서도 전반적으로 나이가 높게 예측되는 경향이 나타났다. 이를 분석해 보면 생성형 AI 모델이 공포나 혐오 표정의 생성 시 표정을 과장되게 생성하는 과정에서 주름 등에 의해 나이가 많게 생성하고 있음을 알 수 있다. 그러나, 이러한 경향은 생성형 AI가 아닌 일반적인 사람의 표정에서도 공포나 혐오 표정을 지을 때도 얼굴을 찌푸리면서 주름이 많이 나타나므로 유사하게 예측된 나이가 증가할 것으로 추정할 수 있다.

마지막으로, 표 4에 제시된 생성된 각 표정과 무표정 얼굴 간의 identity 유사도 결과를 분석해 보면(동일한 얼굴일 경우 유사도는 1에 수렴하고, 다른 얼굴일 경우 유사도는 0에 근접함), ChatGPT가 생성한 얼굴의 identity 유사도가 가장 높은 것으로 나타났다. 즉, ChatGPT가 생성한 얼굴은 모두 같은 사람의 얼굴로 인식되었으나, Co-pilot의 경우는 여성의 경우 화남이나 공포 얼굴의 경우는 다른 사람의 얼굴로 인식되었으며, Gemini는 혐오 표정의 경우 모두 다른 사람의 얼굴을 생성하였으며, Pincel은 여성인 경우 기쁨, 화남 얼굴은 다른 사람의 얼굴로 판단하였다. 다만, ChatGPT의 경우에도 모든 표정이 동일 identity를 가진 얼굴로 판정되었으나, 유사도 평균값이 0.5~0.55 정도이므로 완벽한 동일 얼굴 생성으로 판별하기에는 아쉬운 결과로 보인다.

5. 결 론

본 논문에서는 최근 급속히 발전하고 있는 대

규모 언어 모델(LLM) 기반 생성형 AI가 생성한 얼굴 표정에 대해 감정전달 측면에서의 유용성을 얼굴 표정 인식에 특화된 AI를 활용한 정량적 평가 도구를 통해 분석하였다. 생성된 얼굴 이미지를 표정 인식, 나이 추정, 그리고 identity 유지성 측면에서 다각도로 측정함으로써 생성형 AI의 얼굴 표정 생성 기술 수준을 종합적으로 평가하였다.

표정 인식 정확도 측면에서는 전체적으로 매우 높은 일치도를 기록하였으며, 특히 ChatGPT 기반 생성 결과는 약 98~99%에 이르는 정합도를 보여 현재 생성형 AI가 정서적 표정 생성에 있어 상당히 높은 수준의 전달력을 가지는 것으로 나타났는데, 이는 향후 아바타 기반 인간-인공지능 대화에서 감정 표현의 자연스러움을 크게 향상시킬 가능성을 시사한다.

반면, 나이 추정 결과에서는 공포·혐오와 같은 부정적 감정 생성 시 표정의 과장된 표현으로 인해 나이 예측값이 상승하는 경향이 나타났다. 이는 주름, 찡그림 등 표정 근육의 물리적 특징이 나이 판단에 영향을 미치는 복합적인 현상을 반영한 것으로 볼 수 있다.

Identity 유지성 평가에서는 일부 생성형 AI 모델이 특정 표정에서 동일인의 얼굴을 유지하지 못하고 특징 붕괴 현상이 발생하였다. 특히 공포·화남·혐오와 같이 얼굴 형태 변화가 큰 표정에서 이러한 경향이 두드러졌다. ChatGPT 모델은 상대적으로 가장 안정적인 identity 유지 성능을 보였으나, 유사도 스코어가 0.5~0.55 수준에 머물러 아직 완벽한 동일인 재현 성능에는 미달하고 있음을 알 수 있었다.

이러한 결과는 생성형 AI의 얼굴 표정 생성 기술이 감정 표현 정확도 측면에서는 상당히 실용적 수준에 도달했으나, identity 유지 및 감정 강도에 따른 연령 왜곡 문제 등에서는 여전히 개선의 여지가 있음을 시사한다. 향후 생성형 AI의 표정 생성을 위해 보다 정교한 표정 근육 제어, 3D 기반 표현 보정, multi-task 기반 joint learning 기법 등이 적용될 경우 이러한 한계를 보완하고 아바타 기반의 감정표현형 인공지능 시스템의 신뢰성과 자연스러움을 더욱 향상시킬 수 있을 것으로 기대된다.

나아가 본 연구는 향후 생성형 인공지능 기술의 고도화에 따라 표정 생성 능력에 대한 정량적 분석을 지속적으로 확장·심화함으로써 기술 발전 수준을 체계적으로 평가하고, 이를 토대로 인간과 AI 간 정서적 상호작용의 구현 가능성에 대한 보다 실질적이고 구체적인 통찰을 제공할 수 있을 것으로 기대된다.

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2020-NR053396).

참 고 문 헌

[1] OpenAI. (2023). ChatGPT response to a query about AI ethics [Large language model]. https://chat.openai.com/chat (Accessed May 27, 2025).

[2] Copilot AI 모델, Microsoft, 2025, https://copilot.microsoft.com..

[3] Google. (2024). Gemini response to a question about machine learning applications [Large language model]. https://gemini.google.com (Accessed May 27, 2025).

[4] Pincel. (2025). AI Image Editor [AI photo editing tool]. https://pincel.app/tools/ai-editor

[5] Ekman, Paul, and Wallace V. Friesen. “Constants across cultures in the face and emotion“, Journal of Personality and Social Psychology, vol. 17, no. 2, 1971, pp. 124 -

129. https://doi.org/10.1037/h0030377
[6] Cui, S., Yang, R., Zhou, Y., Xu, J., Wang, Y., Huang, Y., Guo, W., Tan, W., & Liu, X. (2023). FFT: Towards harmlessness evaluation and analysis for LLMs with factuality, fairness, and toxicity. arXiv. https://arxiv.org/abs/2311.18580
[7] Otberdout, N., Ferrari, C., Daoudi, M., Berretti, S., & Del Bimbo, A. (2021). Sparse to dense dynamic 3D facial expression generation. arXiv. https://arxiv.org/abs/2105.07463
[8] Lee, S., Kim, J., & Kim, Y. (2024). SAiD: Speech-driven blendshape facial animation with diffusion. arXiv. https://arxiv.org/abs/2401.08655
[9] Cuculo, V., & D'Amelio, A. (2019). OpenFACS: An open source FACS-based 3D face animation system. In International Conference on Intelligent Human Systems Integration (pp. 173 - 179). Springer
[10] Yeh, R., Liu, Z., Goldman, D. B., & Agarwala, A. (2016). Semantic facial expression editing using autoencoded flow. arXiv preprint arXiv:1611.09961
[11] Song, L., Lu, Z., He, R., Sun, Z., & Tan, T. (2017). Geometry guided adversarial facial expression synthesis. arXiv preprint arXiv:1712.03474.
[12] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Advances in Neural Information Processing Systems (Vol. 27).
[13] Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., & Choo, J. (2018). StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8789 - 8797). https://doi.org/10.1109/CVPR.2018.00916
[14] Ding, H., Sricharan, K., & Chellappa, R. (2018). ExprGAN: Facial Expression Editing with Controllable Expression Intensity. In Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).
[15] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. https://arxiv.org/abs/1411.1784
[16] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1409.1556
[17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp.770 - 778). https://doi.org/10.1109/CVPR.2016.90
[18] Zhou, H., Huang, S., & Xu, Y. (2023). IncepTR: Micro-expression recognition integrating inception-CBAM and vision transformer. Multimedia Systems, 29, 3863 - 3876. https://doi.org/10.1007/s00530-023-01164-0
[19] Goodfellow, I., Erhan, D., Luc Carrier, P., Courville, A., Mirza, M., Hamner, B., Chetlur, S., & Bengio, Y. (2013). Challenges in representation learning: Facial expression recognition challenge. Kaggle. https://www.kaggle.com/competitions/challenges-in-representation-learning-facial-expression-recognition-challenge/data
[20] Li, S., Deng, W., Du, J., Wang, L., & Lu, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2852 - 2861). https://doi.org/10.1109/CVPR.2017.304
[21] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A database for

facial expression, valence, and arousal computing in the wild. In IEEE Transactions on Affective Computing, 10(1), 18 - 31. https://doi.org/10.1109/TAFFC.2017.2740923
[22] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops (CVPRW) (pp. 94 - 101). IEEE. https://doi.org/10.1109/CVPRW.2010.5543262
[23] Xue, F., Wang, Q., & Guo, G. (2021). TransFER: Learning relation-aware facial expression representations with transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 3602 - 3611).
[24] J. Son, J. Park and K. Kim, "CSTA: CNN-based Spatiotemporal Attention for Video Summarization", 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2024, pp. 18847-18856, doi: 10.1109/CVPR52733.2024.01783.
[25] 시Scenario-based facial expression 3D data, AI Hub https://www.aihub.or.kr/
[26] Malte, H. (n.d.). face-api.js: JavaScript API for face detection and recognition in the browser. https://ailearn.space/modules/faceApi/faceExpressionRecognition.html
[27] Li, S., & Deng, W. (2020). Deep Facial Expression Recognition: A Survey. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3834 - 3855.
[28] Rothe, R., Timofte, R., & Van Gool, L. (2016). Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision, 126(2 - 4), 144 - 157.
[29] Chen, B.C., Chen, C.S., & Hsu, W.H. (2013). Ordinal regression with multiple output CNN for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013) (pp. 4920 - 4928).
[30] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015) (pp. 815 - 823).

저 자 소 개

지은미(Eun-Mi Ji)

1988.2. 숭실대학교 전자계산학과
1990.2. 숭실대학교 전자계산학 공학석사
2002.2. 충북대학교 이학박사
2025.3. ~ 현재 배재대학교 게임멀티미디어 공학과 석사과정
1991.4 ~ 1995.2 한국과학기술연구원 정보전자연구부 연구원
<주관심분야> 영상처리, 암호화/보안, 사용자 인증

심윤식(Yoonsik Shim)

2002.2 고려대학교 기계공학과
2004.2 고려대학교 컴퓨터학 석사
2013.6 Univ. of Sussex, Informatics 박사
2013.8-2016.8 Univ. of Sussex, Research Fellow
2018.3-2020.2 : 고려대학교 컴퓨터정보통신대학원 연구교수
<주관심분야> 인공지능, 컴퓨터그래픽스, 인공생명, 로보틱스, 신경망, 시뮬레이션