

# Chat GPT vs Human: Chat GPT 의 창작 주체 가능성

Chat GPT vs Human: Can Chat GPT also become a creative agent?

최이정

Ihjeong Choi

가톨릭대학교 심리학과

Department of Psychology, the  
Catholic University of Korea  
2jcchii@gmail.com

윤지호

Jiho Yun

가톨릭대학교 심리학과

Department of Psychology, the  
Catholic University of Korea  
jhk010319@naver.com

정윤경

Yoonkyung Jeong

가톨릭대학교 심리학과

Department of Psychology, the  
Catholic University of Korea  
benijeong@catholic.ac.kr

## 요약문

인공지능 기술의 발전은 예술 분야에서도 혁신적인 가능성과 도전을 제시하고 있다. 본 연구에서는 Chat GPT 가 예술 창작의 주체로 여겨질 수 있는지를 탐색하기 위해 연구를 설계하였다. 이를 위하여 인간과 Chat GPT 가 창작한 시를 각각 제시하면서 창작 주체를 진실 또는 거짓으로 제안하여 창의성, 즐거움, 추천가능성의 정도를 비교하였다. 연구 결과, 실제 창작 주체가 누구인지는 창의성과 즐거움에는 영향을 미치지 않았지만, 추천 가능성에서는 유의미한 차이가 나타났다. 인공지능이 쓴 시에 대해서는 창작자가 인간이라고 했을 때 창의성에 대해서는 조금 더 높게 지각되었다. 이러한 결과는 인공지능과 인간 창작물이 공존 가능성을 시사하고, 창작 주체 정보가 예술 작품 평가에 미치는 영향을 다루는 데 기여한다. 더불어 Chat GPT 의 다양한 예술 분야에서의 창작 가능성에 대한 미래적 탐구가 필요함을 제언한다.

## 주제어

Chat GPT, AI (Artificial Intelligence), artwork, Art, HCI (Human-Computer Interaction), HAI (Human-AI Interaction)

## 1. 서론

최근 몇 년 동안 인공지능 기술은 상당한 발전을 이루었고, 이에 따라 예술 분야에서도 새로운 가능성과 도전이 제기되었다. 과거 사람들은 예술과 관련된 독창성, 창의성, 감수성은 인간만이 가진 것이라 여겼기에[1][2] 예술 작품 창작을 인간 고유의 영역으로 생각했다. 그러나 ‘컴퓨터가 소설을 쓰는 날’이라는 인공지능이 쓴 문학작품이 일본의 신문사가 주관하는 문학상인 ‘호시 신이치 상’ 1 차 심사예 통과하는 등[3] 최근의 인공지능은 인간 고유의 영역으로 취급되던 예술의 영역에도 진출하기 시작했다.

인공지능이 만든 예술 작품에 대한 연구는 지속적으로 다양한 측면에서 이루어져 왔다. 다양한 연구 중 한 가지를 예로 들면, 인공지능이 만든 예술 작품과 인간이 만든 예술 작품을 사람들이 잘 구별해내지 못한다는 것을 발견한 연구들이 있다[4, 5, 6]. 특히 [6] 연구의 경우, 인간 창작물과 인공지능 창작물, 그리고 인공지능 창작하고 인간이 수정한 창작물을 비교했다. 실험 결과 사람들이 각 작품의 창작자를 식별하지는 못했으나, 인공지능의 창작 본을 인간이 수정한 창작물을 가장 좋게 평가했다는 결과를 보고하며 인간과 인공지능의 예술 분야에서 협업 가능성을 제시했다. 하지만 [4]의 경우, 인공지능이 만든 제품들이 창의성, 복잡성, 의도성 등과 같은 측면에서 인간이 만든 제품보다 더 높게 평가받았다고 밝혔다.

예술 작품에 대한 사람들의 평가와 인식은 매우 주관적인 부분이기 때문에[7, 8], 인간 창작 작품과 인공지능 창작 작품의 평가를 비교하는 연구에서 다양한 연구 결과를 보여왔다. 또한 창작자 정보가 인간인지 인공지능인지에 따라 사람들의 평가가 달라진다는 연구들이 존재한다. 해당 부분을 연구한 논문들에서는 같은 작품이어도 창작자가 인간이라고 한 경우 더 창의적이라고 평가한 연구가 있는 반면[9], 창작자가 인간이라고 했을 때와 인공지능이라고 했을 때 유의미한 평가 차이를 보이지 않았다는 결과를 보인 연구도 존재한다[10]. 한편, [11]에서는 AI 가 작곡한 음악에 대해 부정적인 인식이 실제 AI 작곡 음악 청취 후의 평가에 큰 영향을 미치지 않았다는 것을 밝혔다.

인공지능 기술의 발전은 예술 분야에서 혁신적인 가능성과 도전을 제시하고 있고, 다양한 예술 스타일을 이해하고 모방하여 창작할 수 있는 능력을 갖춘 Chat GPT 가 2022 년 등장하였다. 본 연구에서는 Chat GPT 가 만든 예술 작품의 경우 사람들이 어떻게 평가하는지 조사하고자 한다. 또한 창작자의 정보가 ChatGPT 일 때와 인간일 때에 따라 평가가 달라지는지 알아보고자 다음과 같은 연구 문제를 설정하였다.

연구질문 1: 실제 창작 주체(인간 vs Chat GPT)가 작품의 창의성, 즐거움, 추천 가능성 평가에 영향을 미치는가?

연구질문 2: 제시된 창작 주체 정보가 창의성, 즐거움, 추천 가능성 평가에 영향을 미치는가?

## 2. 연구 방법

### 2.1 연구 설계 및 참여자 모집 방법

연구 문제를 검증하기 위해 2(창작 주체: 인간 vs Chat GPT) X 2(창작 주체 정보공개: 진실 vs 거짓) 설계 실험을 실시하였다. 인간 창작시이고 창작 주체를 진실 그대로 받은 집단(이하 A1), 인간 창작시이지만 창작 주체를 Chat GPT 이라고 거짓으로 받은 집단(이하 A2), Chat GPT 창작시이고 창작 주체를 진실 그대로 받은 집단(이하 B1), Chat GPT 창작시이지만 창작 주체를 인간이라고 거짓으로 받은 집단(이하 B2)로 집단을 나누었다.

실험 참여 신청 폼 링크를 포함한 공고문을 대한민국 내의 다양한 대학 커뮤니티에 작성해 모집했다. 신청자들을 각각의 조건을 가진 4 개의 집단에 무작위로 배정하여 각 참가자의 집단에 맞는 실험 참여 링크를 이메일과 문자 메시지로 전송하였고, 신청자 88 명 중 81 명이 실험을 완료하였다. 시와 설문지가 한국어로 작성되었음을 고려하여 대한민국 국적을 소유하고 있고 한국어를 모국어로 사용하는 사람들만을 대상으로 모집하였다. 참가자들의 연령은 20 세부터 58 세까지로 다양하게 이루어져 있었고(Mean = 25.33, SD = 8.90), 성별의 비중은 여성(N=62; 76.5%)이 남성(N=19; 23.4%)보다 많았다.

### 2.2 실험물 및 실험 절차

창작 주체가 인간인 시를 본 집단, 즉 A1 과 A2 집단의 참가자들에게는 경산 류시호 시인의 '추억 속의 봄길'이라는 시를 제공하였다. 창작 주체가 Chat GPT 인 시를 본 집단, 즉 B1 과 B2 집단의 참가자들에게는 시인이 직접 선정한 작품의 키워드인 '아름다운 추억, 풋풋한 그리움, 강변 독길, 어느 봄날, 지방에서 고등학교 다니며 자전거 타고 통학, 여자친구와 자전거 데이트, 세월이 지나'만을 주고 Chat GPT 에게 시 창작을 요청했다. Chat GPT 가 창작한 시 3 편과 류시호 시인의 시를 제시한 뒤, 자신이 느끼기에 우수하다고 생각되는 순서대로 시의 순위를 정하도록 하는 사전 실험을 온라인 인터뷰를 통해 10 명에게 수행하였다. 사전 실험 결과 가장 높은

순위를 차지한 Chat GPT 창작 시를 Chat GPT 창작시로 B1 과 B2 집단 참가자들에게 제시하였다.

본 실험 또한 온라인을 통해 진행하였다. 참가자 모집 공고 글에 명시된 구글 폼 링크를 통해 자발적으로 실험에 신청한 참가자들은 각각의 집단에 맞는 실험 참여 구글 폼 링크를 문자 메시지 및 이메일로 전송받았다. 실험 참여 링크의 첫 페이지에 정보동의서를 첨부하여 본 실험 시작 전 참가자들의 동의를 얻고 실험을 시작하였다. 실험 참여에 동의한 참가자들은 제시된 창작자 정보와 시를 감상한 후, 시에 대한 느낌을 평가하는 설문지를 작성하였다.

### 2.3 측정 항목

#### 2.3.1 창의성(Creativity)

참가자들이 느끼는 시의 창의성을 측정하기 위해 SP Besemer & K O'Quin [12]이 최초로 개발한 CPSS 를 사용하였다. 설문지는 영어를 한국어로 번역하였을 때 의미가 완전히 일치하는 문항들은 제외하고(e.g., "Priceless - worthless" & "Valuable - worthless" => "가치있는 - 가치없는") 총 47 문항으로 구성되었고, 7 가지 의미 차별 척도를 사용하였다. (e.g., "따분한 - 흥미로운", "예상 가능한 - 참신한") 높은 점수일수록 창의성이 높다는 것을 의미한다. SP Besemer & K O'Quin (1986)에서는 신뢰도 계수(Cronbach's  $\alpha$ )는 .83 이었으며 본 연구에서는 신뢰도 계수(Cronbach's  $\alpha$ )는 .94 로 나타났다.

#### 2.3.2 즐거움(Enjoyment)

참가자들이 시를 보고 느낀 즐거움을 측정하기 위해 Justin B. Moore [13]이 최초로 개발한 PACES 를 사용하였다. 설문지를 한국어로 번역해서 사용했고, 신체적 힘들음을 측정하는 문항을 제외하고 기분을 물어보는 문항들만(e.g., "나는 이 시가 즐겁다", "나는 이 시가 지루하게 느껴진다") 사용하여 총 15 문항으로 구성되었고, 7 점 리커트 척도(7 = 매우 그렇다, 1 = 전혀 그렇지 않다)를 사용했다. 총점이 높을수록 즐거움이 높다는 것을 의미한다. Justin B. Moore [13]에서는 신뢰도 계수 (Cronbach's  $\alpha$ )는 .93 이었으며 본 연구에서는 신뢰도 계수(Cronbach's  $\alpha$ )는 .92 로 나타났다.

표 1. ANOVA 분석

창작주제	인간 창작시				Chat GPT 창작시					
	A1		A2		B1		B2		F	P
	인간	SD	Chat GPT	SD	Chat GPT	SD	인간	SD		
제시된 정보	M	SD	M	SD	M	SD	M	SD		
창의성	4.17	1.58	4.15	1.57	3.96	1.47	4.34	1.50	1.49	0.22
즐거움	3.90	1.80	3.51	1.84	3.70	1.67	3.89	1.71	0.58	0.63
추천 가능성	3.83	2.34	4.07	2.62	4.14	1.97	4.68	2.38	4.47	0.04

### 2.3.3 추천 가능성(Likelihood to Recommend)

참가자들이 해당 시를 다른 사람들에게 얼마나 추천하고 싶은지를 측정하기 위해 FF Reichheld[14]이 최초로 개발한 NPS 수치를 시의 추천 가능성을 평가하기 위한 질문으로 약간의 수정을 하여 사용하였다. "이 시를 다른 사람들에게 얼마나 추천하고 싶으신가요?", "이 시를 친구나 가족과 공유하고 싶은 정도는 어떻게 되시나요?", "이 시를 다른 사람에게 추천하시겠습니까?"와 같은 질문을 사용하였고, 해당 논문에서 제시된 대로 10 단계 consistent scale(0 = 0%, 10 = 100%)를 사용했다. 숫자가 높을수록 추천 가능성이 높다는 것을 의미한다. 본 연구에서는 신뢰도 계수(Cronbach's  $\alpha$ )는 .96 으로 나타났다.

## 3. 연구 결과

결과를 분석하기 위해 본 연구에서는 Python 의 NumPy 와 SciPy 라이브러리를 통해 자료의 통계처리를 실시하였다. 각 측정 항목 별로 ANOVA 분석 및 독립표본 T 검정을 통해 통계분석을 실시하였다. 각각의 대한 결과는 상단의 표 1 과 같다.

먼저, '창의성'에 대한 결과를 살펴보면, ANOVA 분석 결과 '창의성'에 대해서는  $F(3,77) = 1.49$ ,  $p = 0.22$ 로 네 집단 간 유의미한 차이가 나타나지 않았고, '즐거움'에 대한 분석 결과  $F(3,77) = 0.58$ ,  $p = 0.63$ 로 즐거움에서 또한 네 집단 간 유의미한 차이가 나타나지 않았다. '추천 가능성'에 대한 결과를 살펴보면,  $F(3,77) = 4.47$ ,  $p = 0.04$ 로 추천 가능성에서는 네 집단 간 유의미한 차이가 나타났다. 종합하면, 분석결과 창작 주제와 창작 주제 정보 공개 간의 차이가 창의성과 즐거움에는 유의미한 영향을 미치지 않았으나, 추천 가능성에는 유의미한 영향을 미치는 것으로 나타났다.

추천 가능성에 대한 사후 분석 결과, A1 과 B2 간의 평균 차이가 유의미하게 발생한 것을 확인했다.

조금 더 심도 깊은 분석을 위해 T-test 를 시행하였고, T-test 결과를 표로 표기하면 아래와 같다.

표 2. 인간 창작시 집단 T-test

참여자에게 주어진 창작자 정보	A1 인간	A2 ChatGPT	T	P	df
창의성	4.17 (1.58)	4.14 (1.57)	0.12	0.90	96
즐거움	3.89 (1.79)	3.51 (1.84)	1.18	0.24	28
추천 가능성	3.83 (2.33)	4.06 (2.61)	-1.2	0.29	4

표 3. 인공지능 창작시 집단 T-test

참여자에게 주어진 창작자 정보	B1 ChatGPT	B2 인간	T	P	df
창의성	3.96 (1.46)	4.33 (1.50)	-1.98	0.04	96
즐거움	3.70 (1.66)	3.89 (1.84)	-0.53	0.59	28
추천 가능성	4.14 (1.97)	4.68 (2.37)	-1.92	0.12	4

표 4. 정보공개 진실집단 T-test

참여자에게 주어진 창작자 정보	A1 인간	B1 ChatGPT	T	P	df
창의성	4.17 (1.58)	3.96 (1.46)	1.57	0.11	96
즐거움	3.89 (1.79)	3.70 (1.66)	0.58	0.56	28
추천 가능성	3.83 (2.33)	4.14 (1.97)	1.57	0.11	4

표 5. 정보공개 거짓집단 T-test

참여자에게 주어진 창작자 정보	A2 ChatGPT	B2 인간	T	P	df
창의성	4.14 (1.57)	4.33 (1.50)	-0.62	0.53	96
즐거움	3.51 (1.84)	3.89 (1.84)	-1.11	0.27	28
추천 가능성	4.06 (2.61)	4.68 (2.37)	-2.19	0.09	4

동일한 인간 창작시를 감상한 A1 과 A2 집단 T-test 결과 창작 주체의 정보에 따라 창의성 측면에서 유의미한 차이가 나타나지 않았고( $T = 0.12$ ,  $p = 0.90$ ), 즐거움 측면에서도 유의미한 차이가 나타나지 않았으며( $T = 1.18$ ,  $p = 0.24$ ), 추천 가능성 측면에서 또한 유의미한 차이가 나타나지 않았다( $T = -1.2$ ,  $p = 0.29$ ). 즉, 동일한 인간 창작시를 감상한 집단 간에는 창작 주체 정보에 따라 시에 대한 유의미한 평가 차이가 발생하지 않았기 때문에, A1 과 A2 집단에서는 연구문제 2 에 대한 명확한 답변을 도출하기는 어렵다.

동일한 GPT 창작시를 감상한 B1 과 B2 집단 T-test 결과 창작 주체의 정보에 따라 창의성 측면에서 유의미한 차이가 나타났다 ( $T = -1.98$ ,  $p = 0.04$ ). 즐거움 측면에서는 유의미한 차이가 나타나지 않았고( $T = -0.53$ ,  $p = 0.59$ ), 추천 가능성 측면에서도 유의미한 차이가 나타나지 않았다( $T = -1.92$ ,  $p = 0.12$ ). 즉, 동일한 GPT 창작시를 감상한 집단에서는 창작 주체가 인간이라고 정보를 받았을 때 GPT 라고

정보를 받았을 때보다 창의성 측면에서는 더 좋게 평가받았으나, 즐거움과 추천 가능성 측면에서는 창작 주체 정보에 따라 유의미한 평가 차이가 발생하지 않았다. 따라서 B1 과 B2 집단에서는 창의성 부분은 B2 집단에서 더 좋게 평가받았으나, 즐거움과 추천 가능성에서는 연구문제 2 에 대한 명확한 답변을 도출하기는 어렵다.

정보 공개를 진실로 받고 시를 감상한 A1 과 B1 집단 T-test 결과 창작 주체의 정보에 따라 창의성 측면에서 유의미한 차이가 나타나지 않았고( $T = 1.57$ ,  $p = 0.11$ ), 즐거움 측면에서도 유의미한 차이가 나타나지 않았으며( $T = 0.58$ ,  $p = 0.56$ ), 추천 가능성 측면에서 또한 유의미한 차이가 나타나지 않았다( $T = 1.57$ ,  $p = 0.11$ ). 즉, 창의성과 즐거움 그리고 추천 가능성 모두 두 집단 간에 유의미한 평가 차이가 나타나지 않았기에 연구문제 1 에 대해 명확한 답변을 도출하기는 어렵다.

정보 공개를 거짓으로 받고 시를 감상한 A2 와 B2 집단 T-test 결과 창작 주체의 정보에 따라 창의성 측면에서 유의미한 차이가 나타나지 않았고( $T = -0.62$ ,  $p = 0.53$ ), 즐거움 측면에서도 유의미한 차이가 나타나지 않았으며( $T = -1.11$ ,  $p = 0.27$ ), 추천 가능성 측면에서 또한 유의미한 차이가 나타나지 않았다( $T = -2.19$ ,  $p = 0.09$ ). 따라서 창의성과 즐거움 그리고 추천 가능성 모두 두 집단 간에 유의미한 평가 차이가 나타나지 않았기에 연구문제 1 에 대해 명확한 답변을 도출하기는 어렵다.

#### 4. 결론

본 연구는 인간이 Chat GPT 가 생성한 창작물을 어떻게 인식하고 평가하는지에 대한 통찰과 인간과 인공지능의 상호작용에 대한 이해를 높일 수 있는 기회를 제공한다. 아울러 다음과 같은 연구 결론을 도출할 수 있었다.

본 연구에서는 Chat GPT 의 창작물, 특히 문학 작품 창작물에 대한 사람들의 인식을 조사했다. 조사 결과 인간 창작시 집단에서는 창작자의 정보가 영향을 미치지 않았으나, Chat GPT 창작시 집단에서는 창작자의 정보가 창작물의 창의성에 대한 지각에서는 영향을 미친 것을 확인했다. 이를 바탕으로 같은 창작물이어도 창작자가 인간이라고 했을 때 창의성을 더 좋게 평가하며, 창의성을 인간 고유의 특징이라고 무의식적으로 인식하고 있음을 암시한다. 하지만 우리의 연구의 표본이 집단별로 충분하지 않았기에, 더 정확히 조사하기 위한 추후 연구가 필요할 것이다.

또한 통계 분석 결과 각 집단별로 시에 대한 유의미한 평가 차이가 없던 것으로 보아 Chat GPT 가

인간스러운 글(Human-like Text)을 구사해 글을 쓸 수 있다는 것을 확인할 수 있다. 더불어 본 연구에서는 창작 주체에 따른 실험 참여자들의 평가를 거시적으로만 조사했기에, 연령대나 성별과 같은 특수성까지 미시적으로 조사하여 질적 연구의 차원을 추후 연구에서 반영한다면 인간-컴퓨터 상호작용 분야에서 더욱 유의미한 기여가 가능할 것이다.

한편, Coeckelbergh[2]는 인공지능이 주체적으로 창작물을 만들 수 있는가에 대한 질문은 과학적인 질문이 아니라 철학적인 질문이라고 말했다. 또한 대화형 인공지능이라는 Chat GPT 의 특성과 [2]의 연구를 종합적으로 고려했을 때, Chat GPT 가 독립적인 창작 주체가 될 수 있다고 우리 연구만으로 확정적인 결론을 내리는 것은 다소 어렵다. 하지만 Chat GPT 가 인간스러운 글(Human-like Text)을 구사한다는 것은 분명하기 때문에 인간이 Chat GPT 가 창작한 문학작품을 보완한다면, 즉 인간과 Chat GPT 가 협업한다면 더 우수한 작품이 탄생할 수 있을 것이다.

#### 참고 문헌

1. Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of consciousness Studies*, 6(6-7), 15-51
2. Coeckelbergh, M. (2017). Can machines create art?. *Philosophy & Technology*, 30(3), 285-303.
3. Matayama, Ryohei, Satoshi Satoshi, and Takuya Matsuzaki. (2017). "Automatic Generation of Novels from Werewolf Game Logs." *Proceedings of the 23rd Annual Meeting of the Association for Natural Language Processing*, 32-35.
4. Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative Adversarial Networks generating "art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*.
5. Samo, A., & Highhouse, S. (2023). Artificial intelligence and art: Identifying the aesthetic judgment factors that distinguish human-and machine-generated artwork. *Psychology of Aesthetics, Creativity, and the Arts*.
6. Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2023). Does human-AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior*, 139, 107502.
7. Styve ´ n, M. (2007). Exploring the online music market: Consumer characteristics and value perceptions. Doctoral dissertation, Lulea ´ Tekniska Universitet.
8. Chen, Y. (2009). Possession and access: Consumer desires and value perceptions regarding contemporary art collection and exhibit visits. *Journal of Consumer Research*, 35, 925-940.
9. 정태현, & 신형덕. (2021). 창작자 정보가 미술품의 창의성 평가에 미치는 영향: 인공지능과 인간 창작자의 비교. *한국 HCI 학회 논문지*, 16(4), 25-31.
10. 박주연, 박우승, & 사영준. (2023). 인공지능 창작 작품의 예술성 평가: 인공지능 마음인식의 조절된 매개효과를 중심으로. *정보사회와 미디어*, 24(1), 85-112.
11. Tigre Moura, F., & Maw, C. (2021). Artificial intelligence became Beethoven: how do listeners and music professionals perceive artificially composed music?. *Journal of Consumer Marketing*, 38(2), 137-146.
12. Besemer, S. P., & O'Quin, K. (1986). Analyzing creative products: Refinement and test of a judging instrument. *The Journal of Creative Behavior*.
13. Moore, J. B., Yin, Z., Hanes, J., Duda, J., Gutin, B., & Barbeau, P. (2009). Measuring enjoyment of physical activity in children: validation of the physical activity enjoyment scale. *Journal of applied sport psychology*, 21(S1), S116-S129.
14. Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review*, 81(12), 46-55.