

A Comparative Study of Color-Emotion Responses between Individuals with Color Vision Deficiency and Generative AI

Su Jin Park¹, Jeongmin Kho², Jeeun Kim³, Sangwon Lee⁴, Jeehyun Lee⁴, Long Xu^{5*}

¹Interdisciplinary Program of Human Life & Innovation Design, Postdoctoral Researcher, Yonsei University, Seoul, Korea

²Interdisciplinary Program of Human Life & Innovation Design, Research Professor, Yonsei University, Seoul, Korea

³Department of Integrated Design, Visiting Professor, Yonsei University, Seoul, Korea

⁴Department of Integrated Design, Professor, Yonsei University, Seoul, Korea

⁵Department of Product Design, Associate Professor, Qingdao University of Technology, Qingdao, China

Abstract

Background Color emotion research has predominantly focused on normal color vision populations, leaving color vision deficiency (CVD) individuals' color emotion characteristics largely unexplored. Studies on artificial intelligence (AI) capacity to simulate human color emotion responses have similarly been limited to normal color vision groups, leaving open the question of how faithfully AI can reflect CVD individuals' responses.

Methods Using 130 color stimuli from the Nippon Color & Design (NCD) 130 color system and 28 emotional adjectives, color emotion ratings from a CVD group were compared against responses generated by three large language models (LLMs) prompted with CVD type-specific personas. Intraclass correlation, Pearson correlation, one-way ANOVA, and principal component analysis were conducted to examine within-group consistency, pattern similarity, mean differences, and dimensional structure.

Results AI showed high pattern similarity with the CVD group on perceptually grounded dimensions but markedly low similarity on context-dependent dimensions. AI exhibited systematic bias, underrating dynamic emotions while overrating calm and natural ones, and generated excessively uniform responses relative to the CVD group. Although both groups shared the same three-dimensional structure, CVD participants prioritized the arousal dimension while AI prioritized the rusticity-modernity dimension.

Conclusions AI limitations stem not from CVD-specific factors but from a structural inability to capture context-dependent dimensions, a pattern consistent across both normal and CVD populations. The reversal in dimensional priority suggests that CVD's luminance-dependent perceptual characteristics may influence the organization of color emotion space itself. In inclusive design, a collaborative framework assigning AI to physical attribute-based screening and human designers to contextual and emotional judgment warrants consideration.

Keywords Color Emotion, Color Vision Deficiency, Large Language Model, Generative AI, Semantic Differential

Copyright : This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5A2A03012004).

*Corresponding author: Long Xu (xulong@qut.edu.cn)

Citation: Park, S., Kho, J., Kim, J., Lee, S., Lee, J., & Xu, L. (2026). A Comparative Study of Color-Emotion Responses between Individuals with Color Vision Deficiency and Generative AI. *Archives of Design Research*, 39(3), 243-257.

<http://dx.doi.org/10.15187/adr.2026.08.39.3.243>

Received : Jun. 01. 2026 ; **Reviewed** : Jul. 08. 2026 ; **Accepted** : Jul. 21. 2026
pISSN 1226-8046 **eISSN** 2288-2987

1. 연구의 배경 및 목적

색채는 디지털 인터페이스, 제품 디자인, 브랜드 아이덴티티 등 다양한 디자인 맥락에서 감성적 의미를 전달하는 핵심 매체로 기능하며, 색채 자극이 직접적으로 유발하는 감성 반응인 색채 감성에 대한 체계적 연구는 Kobayashi(1981)와 Ou et al.(2004) 등을 통해 오랜 기간 축적되어 왔다. 그러나 이러한 연구들은 공통적으로 일반 색각자(Normal Color Vision; NCV)를 표준 집단으로 가정하고 있으며, 전 세계 약 2억 명이 영향을 받는 색각 이상(Color Vision Deficiency; CVD) 집단의 색채 감성 특성은 상대적으로 간과되어 왔다(Machado et al., 2009). 색각자는 단순한 색채 변별 능력의 저하에 그치지 않고 색채 감성 반응의 구조 자체에도 영향을 미치기 때문에(Sato & Inoue, 2016; Jonauskaitė et al., 2021), 다양한 시각 특성을 가진 사용자를 포괄하는 포용적 디자인을 위해서는 색각자 집단의 고유한 색채 감성 특성에 대한 이해가 필요하다.

이러한 흐름과 함께, 최근 대규모 언어 모델(Large Language Model; LLM)의 발전으로 AI의 색채 감성 재현 가능성에 대한 관심이 높아지고 있다. Zheng & Xu(2025)는 GPT-4o의 색채 감성 반응이 물리적 속성 기반 차원에서는 인간과 유사하지만 문화, 맥락적 차원에서는 괴리가 크다는 것을 확인하였으나, 일반 색각 집단만을 대상으로 하였다는 한계를 지닌다. Hayashi et al.(2025) 역시 AI가 색각 이상에 관한 사실적 지식은 보유하지만 실제 지각적 결과를 재현하지 못함을 보고하였으나, 감성 반응 차원의 비교는 다루어지지 않았다. 결과적으로 AI가 색각자 집단의 색채 감성 반응을 얼마나 충실하게 반영할 수 있는가라는 질문은 여전히 검토되지 않은 연구 공백으로 남아 있다.

이에 본 연구는 생성형 AI가 색각자의 실제 지각을 재현하는지 여부를 검증하기보다는, 색각자 감성 데이터와 비교하여 AI의 언어적 색채 감성 응답이 어느 정도 일치하는지를 분석하는 것을 목적으로 한다. Kho et al.(2026)의 색각자 집단 색채 감성 데이터셋을 기반으로, GPT-4o, Claude 3.5 Sonnet, Gemini 1.5 Pro에 색각자 페르소나를 부여하여 NCD 130 표색계의 130개 색 표에 대한 28개 감성 형용사 5점 척도 평가 데이터를 수집하였다. ICC, Pearson r, ANOVA, PCA의 네 가지 분석을 통해 집단 내 일치도, 패턴 유사성, 평균 차이, 차원 구조를 비교하였다.

본 연구는 기존 AI 색채 감성 연구를 색각자 집단으로 확장하여 AI의 시뮬레이션 가능성을 검토하고, 감성 차원별 가능성과 한계를 구체적으로 규명함으로써 AI 기반 포용적 디자인 도구 개발을 위한 경험적 근거를 제공한다. 나아가 색각자 사용자를 위한 색채 디자인에서 AI와 인간 디자이너 간의 역할 분담 모델을 제안하는 실무적 함의를 함께 제시한다.

2. 이론적 고찰

2.1. 색채 감성의 이론적 기반

색채 감성은 색채 자극이 직접적으로 유발하는 감성 반응으로, 색채의 상징과 문화적 의미를 다루는 색채 의미론과 구별된다(Schloss, 2024). 색채 감성 연구는 Osgood et al.(1957)의 의미미분법을 중심으로 발전해왔으며, 이후 Wright & Rainwater(1962), Kobayashi(1981), Ou et al.(2004)을 거치며 색채 감성의 보편적 구조가 체계화되었다. 이 연구들을 종합하면, 색채 감성은 색의 물리적 속성에 직접 대응하는 지각적 차원(예: warmth, weight, lightness)과 개인의 경험·문화적 맥락에 의해 형성되는 맥락 의존적 차원(예: modernity, elegance, naturalness)으로 구분된다. 전자는 문화권을 초월하여 비교적 일관되게 나타나는 반면, 후자는 집단에 따라 상당한 변동성을 보인다(Jonauskaitė et al., 2020, 2024). 이러한 이분법적 구조는 AI의 색채 감성 처리 능력을 평가하는 이론적 기준으로 기능하며, 구체적으로는 AI가 각 차원에서 인간과 유사한

반응을 보이는지, 아니면 분명한 차이를 보이는지를 분석하는 틀을 제공한다. 그와 동시에 색채 지각 방식이 다른 집단에서도 이러한 구조가 동일하게 성립하는지에 대한 질문을 남긴다.

2. 2. 색약자 집단의 색채 감성 지각

색약자는 색채 변별 능력의 저하에 그치지 않고 색채 감성 반응에도 직접적인 영향을 미친다. Sato & Inoue(2016)는 색약자 집단이 warmth 평가 시 red-green opponent channel 대신 luminance cone-contrast에 주로 의존함을 보고하였으며, 이러한 명도 의존적 지각 전략은 각성(arousal)이나 강도(intensity)와 같이 명도 변화에 민감한 감성 차원을 상대적으로 현저하게 지각하게 만들 가능성이 있다. 또한 색약자 집단은 고채도 적녹 배색에서 지각적 대비의 감소로 인해 일반 색각자보다 정적인 인상을 경험하며(Sakamoto & Ichihara, 2019), 색채와 감성 연합의 패턴과 집단 내 개인 간 변동성이 일반 색각자와는 부분적으로 상이하다(Jonauskaitė et al., 2021). 이처럼 색약자는 단일한 지각 특성을 가진 집단이 아니며, 기존의 색채 감성 체계인 Kobayashi의 Color Image Scale과 Ou et al.의 3차원 분류 체계가 이들의 지각적 다양성을 충분히 반영하기는 어렵다. 때문에, 색약자 집단의 색채 감성을 이해하기 위해서는 그들의 고유한 지각 특성에 기반한 독립적 탐색이 필요하며, 이 과정에서 AI가 색약자 집단의 경험을 얼마나 재현할 수 있는지에 대한 검토가 병행될 필요가 있다.

2. 3. AI의 색채 감성 처리 능력과 한계

AI는 인간과 유사한 감성 범주 구조를 보유하면서도 복잡한 감성적 뉘앙스를 포착하는 것에는 한계가 있으며(Miotto et al., 2022; Elyoseph et al., 2023), 이는 색채 감성 영역에서도 일관되게 확인된다. Zheng & Xu(2025)는 GPT-4o가 warm-cool, heavy-light 등 물리적 속성 기반 차원에서는 인간의 수준을 상회하는 일관성을 보이는 반면, modern-classical, active-passive 등 맥락 의존적 차원에서는 인간과의 괴리가 크게 나타남을 보고하였다. 주성분 분석에서도 AI는 물리적 속성(heavy-light)을 중심으로 색채 감성 공간을 조직한 반면 인간은 문화적이고 맥락적인 차원(active-passive)을 우선적 축으로 처리함을 확인하였다. AI가 집단 내에서 높은 일관성을 보이는 것은, 신뢰성의 지표라기보다는 개인마다 다르게 나타나는 색채 감성의 다양성을 시뮬레이션하지 못하는 구조적인 한계로 해석된다.

그러나 이러한 연구들은 여전히 일반 색각 집단만을 대상으로 한다는 공통적인 한계를 지닌다. 색약자 집단을 대상으로 한 AI 시뮬레이션 연구에서도 유사한 한계가 확인되는데, AI는 색각 이상에 대한 사실적 지식은 보유하지만 실제 지각적 결과를 재현하지 못한다(Hayashi et al., 2025). 또한, 색약자를 위한 디자인은 단순히 색채 조정을 넘어 사용자의 감성적 경험을 반영하는 포용적 접근을 필요로 한다(Tapang & Espineda, 2024). 이에 본 연구는 AI가 색약자 집단의 색채 감성 반응을 얼마나 충실하게 반영할 수 있는지를 검토함으로써 AI 기반 포용적 디자인 도구 개발을 위한 경험적 근거를 제공하고자 한다.

3. 연구 방법

3. 1. 연구 설계

본 연구는 생성형 AI 모델이 색약자의 색채 감성 반응을 얼마나 충실하게 반영할 수 있는지를 검토하기 위해 비교 연구 설계를 채택하였다. Zheng & Xu(2025)는 일반 색각 집단과 AI의 색채 감성을 비교하였으나, 색약자와 같이 시각적 다양성을 가진 집단을 대상으로 한 연구는 아직 이루어지지 않았다. 이에 본 연구는 색약자 집단의 색채 감성 반응 데이터를 세 가지 LLM(ChatGPT, Claude, Gemini)이 동일한 색채 자극에 대해 생성한 응답 데이터와 비교하는 것을 핵심 구조로 삼는다. 수집된 데이터는 집단 내 평가자 간 일치도, 집단 간 패턴 유사성, 집단 간 평균 차이, 색채-감성 차원 구조의 순으로 분석하였다.

3. 2. 인간 데이터 출처

인간 집단의 색채 감성 데이터는 Kho et al.(2026)의 학술발표에서 구축된 색약자 감성평가 데이터셋을 색약자 기준 데이터로 활용하였다. 해당 연구는 NCD 130 표색계를 자극으로 사용하여 색약자 집단을 대상으로 색채 감성 평가를 수집한 것으로, 색약자 집단은 이시하라 검사 및 Farnsworth-Munsell 100색상 검사를 통해 색각 이상이 확인된 20~60세의 성인 27명으로 구성되어 있다. 데이터는 참여자 × 130(색 표) × 28(형용사)의 3차원 평가 행렬 형태로 구성되었으며, 각 셀의 값은 해당 참여자가 특정 색 표에 대해 28개 감성 형용사를 5점 척도로 평가한 값을 나타낸다.

3. 3. 실험 자극

실험 자극으로는 NCD 130 표색계를 사용하였다. 이 색 표 세트는 색상(hue), 명도(lightness), 채도(chroma)의 다양한 조합을 체계적으로 포괄하는 색채심리 연구의 표준 자극으로, 광범위한 색 공간을 대표하는 타당성이 확보되어 있다. AI 실험에서는 직접적인 시각 입력이 불가능하므로, 각 색 표를 NCD 색이름, RGB 값(0~255), CIE L*a*b* 좌표의 세 가지 색 공간 수치 정보로 변환하여 제공하였다. 이와 같은 다중 색 공간 인코딩 방식은 LLM 모델마다 색채 내부 표상의 다양성을 고려한 것으로, 세 모델 모두에게 동일한 조건에서 일관된 경험을 제공하고 충분한 색채의 정보를 제공하기 위해 수행하였다.

3. 4. 측정 도구

감성 어휘 측정 도구는 Kho et al.(2026)의 연구에서 도출된 28개 감성 형용사로 구성되었다. 해당 어휘 목록은 색채 자극에 대한 자유 기술 과정에서 충분한 빈도로 출현한 단어들을 추출하여 확정된 것으로, 세 개의 LLM 모델의 결과를 동일한 기준으로 비교할 수 있는 공통된 측정 틀로 기능한다. 선정된 28개 형용사는 Figure 1의 y축 레이블과 같다.

3. 5. AI 데이터 수집 절차

AI 비교 집단으로는 OpenAI의 ChatGPT(GPT-4o), Google의 Gemini(Gemini 1.5 Pro), Anthropic의 Claude(Claude 3.5 Sonnet)를 선정하였다. 이들 모델은 현재 가장 광범위하게 활용되는 LLM으로서 공개 API를 통한 접근이 가능하며, 다양한 추론 과제에서 높은 성능을 보인 것으로 보고되어 왔다. 다음으로, AI에게 색약자의 지각적 관점을 부여하기 위해, 제1형(protanopia), 제2형(deutanopia), 제3형(tritanopia) 색각 이상 유형별로 각 30개의 페르소나 프로필을 설정하였다. 각 프로필은 해당 색각 이상 유형의 지각적 조건에서 색 표를 경험하는 것처럼 감성을 평가하도록 지시하였으며, AI 모델별로 총 90개의 프로필 응답을 수집하였다. 각 프롬프트는 단일 색 표에 대한 수치 정보(NCD 색채명, RGB, CIE L*a*b* 값)를 제시하고, 28개 감성 형용사 각각에 대해 5점 척도로 평가하도록 구성하였다. 최종적으로 각 AI 모델당 90개 프로필 × 130색 × 28형용사의 327,600개, 총 982,800개의 응답 데이터를 수집하였다.

3. 6. 자료 분석

수집된 데이터는 Python 3.11을 사용하여 네 가지의 통계 방법으로 분석하였다. 1)집단 내 평가자 간 일치도(ICC 3,1), 2)집단 간 패턴 유사성(Pearson r), 3)집단 간 평균 차이(One-way ANOVA, Tukey HSD), 4)색채-감성 차원 구조(PCA)의 순으로 분석하였다. 형용사별 28회의 ANOVA에 따른 다중검정 오류율은 Benjamini-Hochberg FDR 절차($q < .05$)로 통제하였으며, 보정 후에도 28개 형용사 모두 유의하였다. 사후검정에는 Tukey HSD를 사용하여 각 형용사 내 쌍별 비교의 family-wise error를 통제하였다.

먼저, ICC 해석 기준은 Koo & Li(2016)에 따라 0.50 미만은 낮음, 0.50~0.75는 보통, 0.75~0.90은 좋음, 0.90 이상은 매우 좋음으로 분류하였다. 본 연구에서는 평가자가 각 집단 내에서 고정되어 있고, 해당 집단 내 일치도만을 분석 대상으로 삼았으므로, ICC(3,1)모델을 적용하였다. 다음으로, 색채와 감성 간의 구조적 관계를 파악하기 위해 집단별 주성분 분석을 실시하였다. 130개 색채를 분석 대상으로 삼았으며, 28개의 형용사 평가 점수를 변수로 설정하였다. 주성분의 수는 누적 설명분산과 해석 가능성을 고려하여 3개로

선정하였다. Human, GPT, Claude, Gemini의 첫 세 개 주성분은 각각 77.5%, 81.9%, 87.9%, 87.2%의 누적 설명력을 보여 대부분의 분산을 설명하였다. 반면 네 번째 주성분 이후의 추가 설명력은 4.4~8.4% 수준으로 상대적으로 작았으므로, AI 모델과 인간 집단 간 차원 구조를 일관되게 비교하기 위해 3개의 주성분을 사용하였다. 마지막으로 집단 간 비교를 위해 색약자 집단의 loading 값은 내림차순 정렬하였다.

4. 연구 결과

4. 1. 집단 내 일치도 분석

Figure 1은 28개 형용사에 대한 네 집단의 ICC(3,1) 값을 나타낸다. 전반적으로 실제 사람인 색약자 집단이 가장 낮은 ICC 값을 보였으며, GPT, Gemini, Claude 순으로 집단 내 일치도가 높아지는 위계적인 패턴이 확인되었다.

색약자 집단은 대부분의 형용사에서 평가자 간 일치도가 낮음 ($ICC < 0.50$)으로 나타났으며($M = 0.20$), 이는 색채-감성 평가가 개인마다 크게 다르며, 반응이 다양하다는 것을 확인할 수 있었다. 다만 예외적으로 Light ($ICC = 0.51$)와 Heavy ($ICC = 0.50$)만이 보통 수준에 도달하였는데, 이는 명도처럼, 누구나 비교적 뚜렷하게 느끼는 속성일수록 개인별 평가 결과의 일치도가 상대적으로 높게 나타나는 것으로 해석할 수 있다.

반면에, AI 집단 중 Claude는 대부분의 형용사에서 좋음에서 매우 좋음 수준의 ICC를 보였으며($M = 0.70$), 90개의 시뮬레이션 페르소나가 매우 일관된 응답을 생성하였음을 나타낸다. 특히 Soft ($ICC = 0.87$), Light ($ICC = 0.88$), Heavy ($ICC = 0.83$)에서 매우 좋음 수준에 도달하였다. Gemini는 보통에서 좋음 수준의 일치도를 보였으며($M = 0.57$), GPT는 전반적으로 보통 수준에 머물렀다($M = 0.49$). 한편 Emotional, Natural, Modern과 같이 문화적 맥락이나 추상적 의미를 내포하는 형용사에서는 모든 집단에서 상대적으로 낮은 ICC가 관찰되었다. 이는 지각적 근거가 약한 차원에서는 AI 모델 역시 집단 내 일관성이 저하됨을 보여준다. 이러한 결과는 AI 모델, 특히 Claude가 색약자 집단의 평가자 간 합의 수준을 크게 상회하는 응답적인 균일성을 유지하고 있음을 나타낸다. 그러나 AI의 높은 ICC는 응답의 안정성을 반영하는 동시에, 색약자 집단의 색채-감성 지각에 내재된 개인 간 다양성을 충분히 시뮬레이션하지 못할 가능성을 배제할 수 없다.

4. 2. 집단 간 패턴 유사성 및 평균 차이

Figure 2와 3은 색약자 집단과 AI 3종 간의 색채-감성 평가 패턴 유사성(Pearson r) 및 집단 간 평균 차이(One-way ANOVA, Tukey HSD)를 나타낸다.

Pearson r 분석 결과(Figure 2), 세 AI 모델의 색약자 집단과의 패턴 유사성은 전반적으로 보통 수준에 머물렀다(GPT: 28개 형용사 평균 $r = 0.544$, Claude: $r = 0.480$, Gemini: $r = 0.492$). $r = .70$ 이상의 높은 상관을 보인 형용사는 GPT 8개, Claude 7개, Gemini 6개에 그쳤다. 형용사별로는 Light (GPT: $r = 0.91$, Claude: $r = 0.93$, Gemini: $r = 0.89$), Heavy (GPT: $r = 0.89$, Claude: $r = 0.90$, Gemini: $r = 0.89$), Soft (GPT: $r = 0.81$, Claude: $r = 0.83$, Gemini: $r = 0.82$)와 같이 지각적으로 명확한 차원에서 세 모델 모두 높은 상관이 일관되게 관찰되었다. 반면 Classic (GPT: $r = 0.10$, Claude: $r = 0.06$, Gemini: $r = 0.19$), High-tech (GPT: $r = 0.23$, Claude: $r = 0.08$, Gemini: $r = 0.11$), Natural (GPT: $r = 0.33$, Claude: $r = -0.01$, Gemini: $r = 0.20$) 등의 형용사에서는 세 모델 모두 색약자 집단과의 패턴 유사성이 현저히 낮았다. 특히 Claude는 네 개 형용사(Alluring, Natural, Classic, High-tech)에서 $r < .10$ 으로 색약자 집단과의 패턴 유사성이 사실상 없는 수준을 나타냈으며, 이는 세 모델 중 가장 많았다. 형용사별 일원분산분석 결과, 28개 형용사 전체에서 집단 간 유의한 차이가 확인되었다(모든 형용사 $p < .001$).

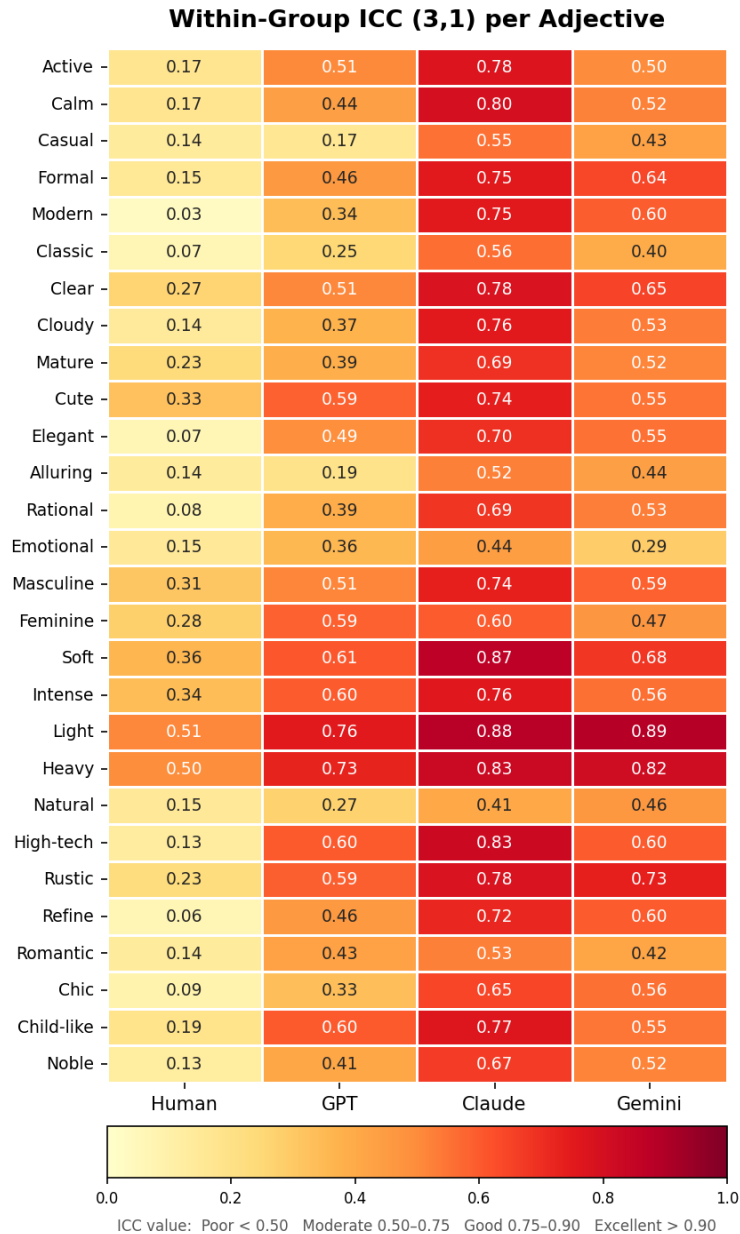


Figure 1 Within-group ICC (3,1) values for 28 color-emotion adjectives across Human and three AI groups (GPT, Claude, Gemini). Higher values indicate greater consistency among raters within each group.

Tukey HSD 사후 검정 결과(Figure 3), Human vs Claude 쌍은 28개 형용사 중 96.4%에서, Human vs Gemini는 92.9%에서, Human vs GPT는 71.4%에서 유의한 평균 차이가 확인되었다($p < .05$). 평균 차이의 방향을 살펴보면, AI 모델들은 High-tech (평균 차이 = 1.16), Intense (0.99), Active (0.65), Emotional (0.54), Alluring (0.51) 등의 형용사에서 색약자 집단에 비해 현저히 낮은 평균값을 보였다. 반면 Soft (-0.70), Natural (-0.66), Calm (-0.63), Cloudy (-0.48), Classic (-0.36) 등에서는 AI가 색약자 집단보다 높은 평균값을 나타냈다. 즉 AI는 강렬하거나 역동적인 감성 차원에서 색약자 집단보다 낮게, 정적이거나 자연적인 차원에서는 높게 평가하는 경향이 있었다. 이는 AI가 색채-감성 평가에서 색약자 집단과 유사한 순위 패턴을 일부 공유하더라도, 절대적 평가 수준에서는 체계적인 편향이 존재하는 것으로 해석할 수 있다.

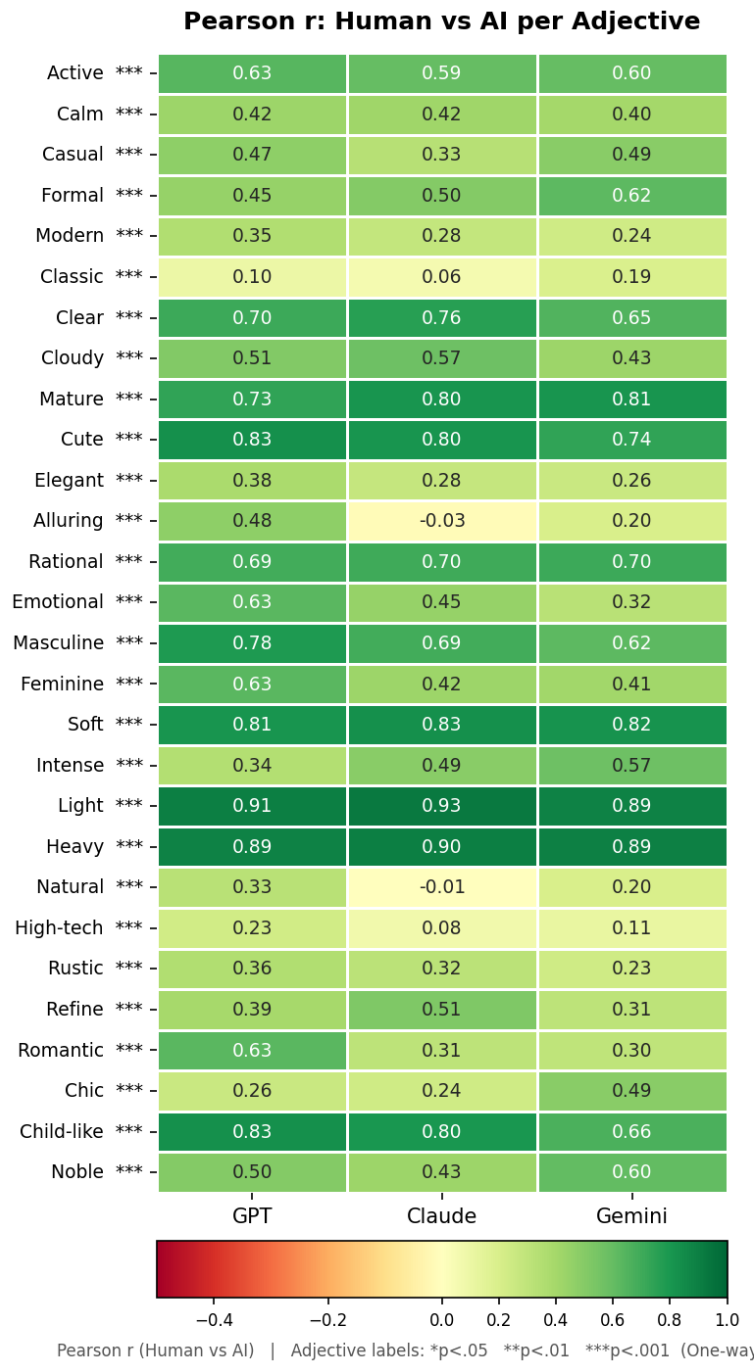


Figure 2 Pearson r between Human and AI group means per adjective across 130 color stimuli (GPT, Claude, Gemini).

Mean Color-Emotion Ratings by Group
error bar = SD | *p<.05 **p<.01 ***p<.001 (Tukey HSD, Human vs AI)

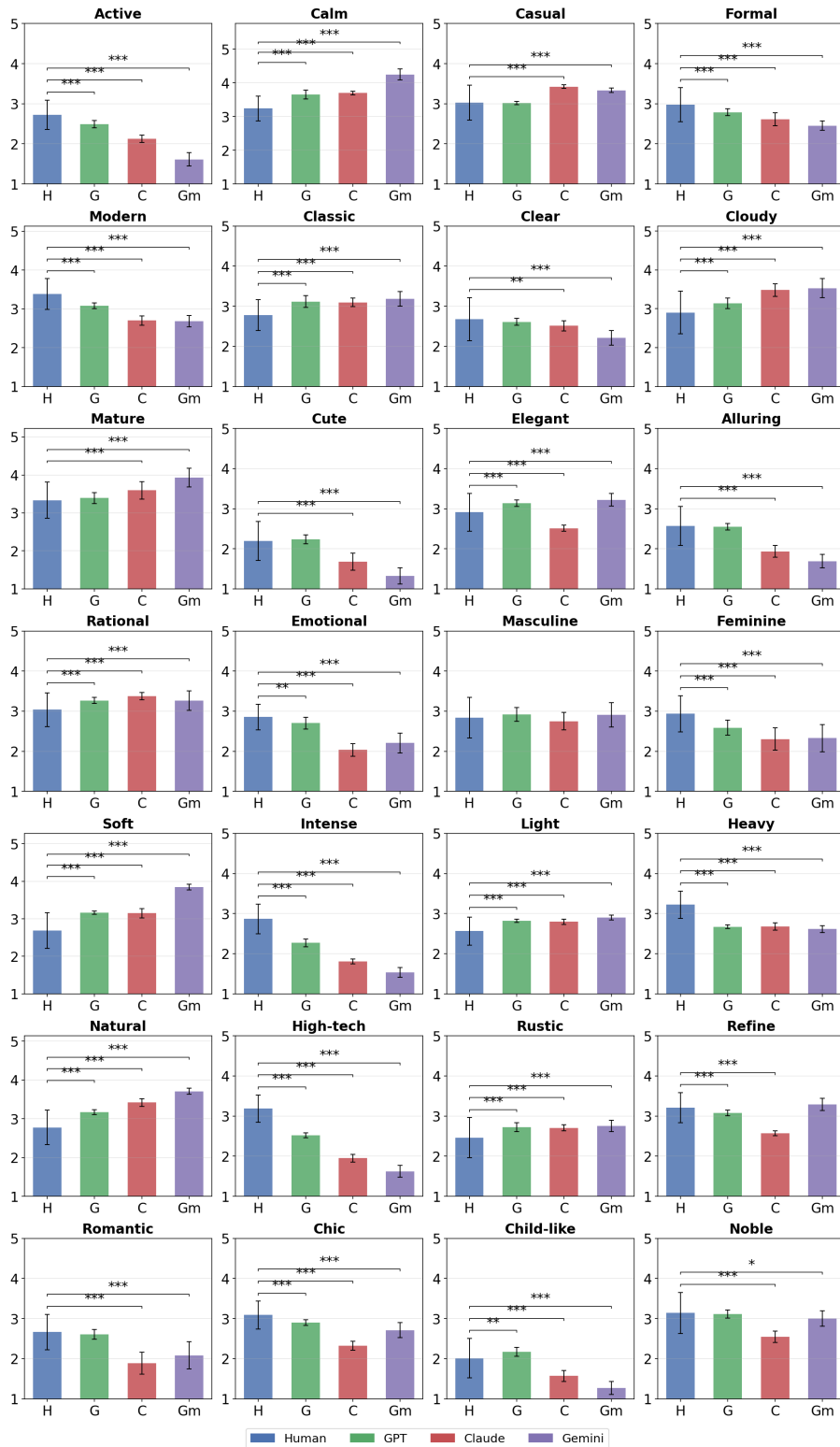


Figure 3 Mean color-emotion ratings per adjective by group (H = Human, G = GPT, C = Claude, Gm = Gemini). Error bars indicate SD. Brackets indicate significant differences between Human and each AI group (Tukey HSD; *p < .05, **p < .01, ***p < .001).

4. 3. 색채-감성 차원 구조

Figure 4-6은 네 집단의 PCA 결과를 나타낸다. 3개 주성분의 누적 설명 비율은 색약자 집단 77.5%, GPT 81.9%, Claude 87.9%, Gemini 87.2%로 나타났다. 색약자 집단의 응답은 복잡하고, 다양한 반면, AI 모델은 색약자 집단보다 더 적은 수의 차원으로 색채-감성 구조를 압축적으로 설명한다. 이는 앞서 ICC 분석에서 확인된 AI의 높은 집단 내 일관성 경향과 동일한 양상을 보인다.

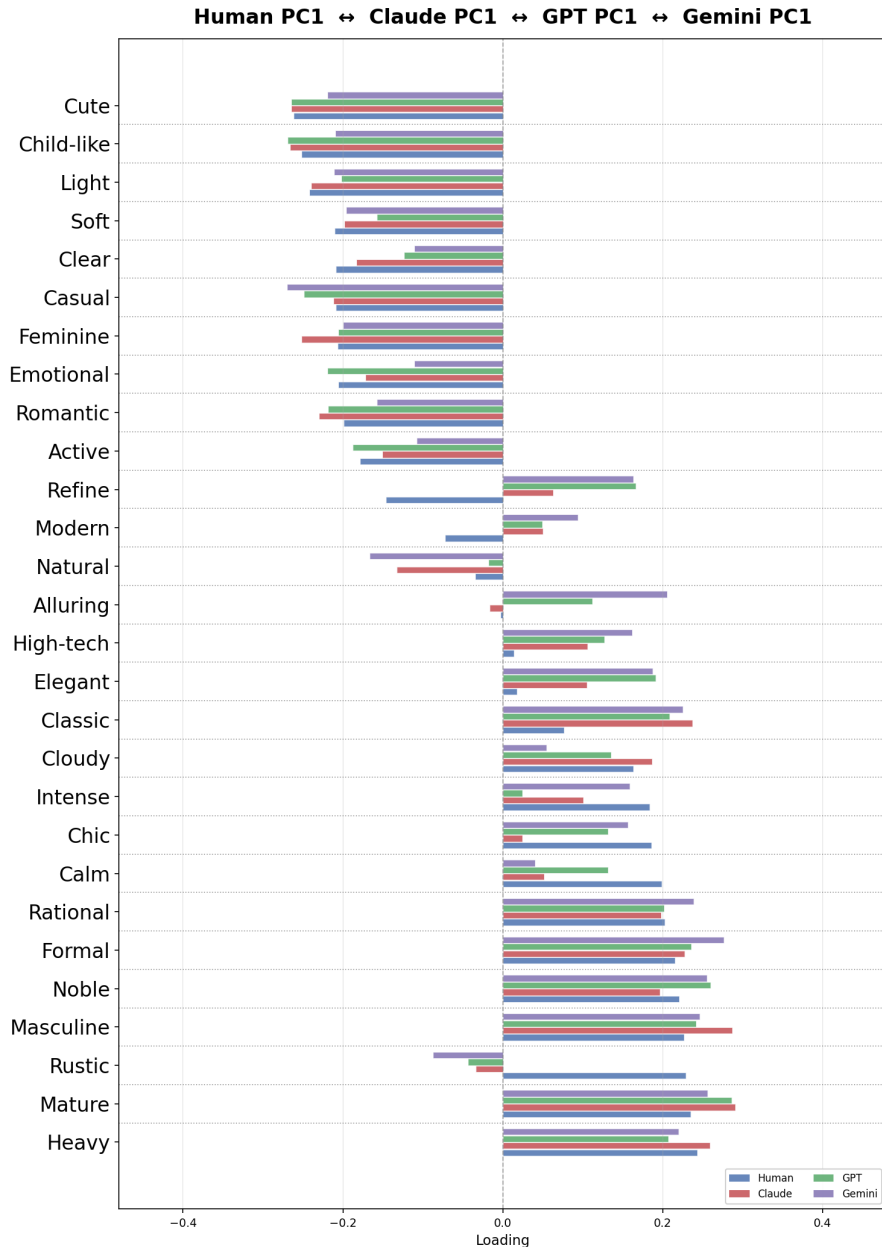


Figure 4 PCA loadings for PC1 (Gravitas vs. Cuteness) across Human and three AI groups (GPT, Claude, Gemini). Adjectives are sorted by Human PC1 loading in descending order.

그러나, 주성분 분석 결과, 색약자 집단과 AI 3종은 동일한 감성 차원을 공유하는 것으로 나타났다. 다만, 색약자 집단에서 2순위였던 차원이 AI에서는 3순위, 3순위였던 차원이 2순위로 나타나는 등, 중요도 순서에는 차이가 있었다.

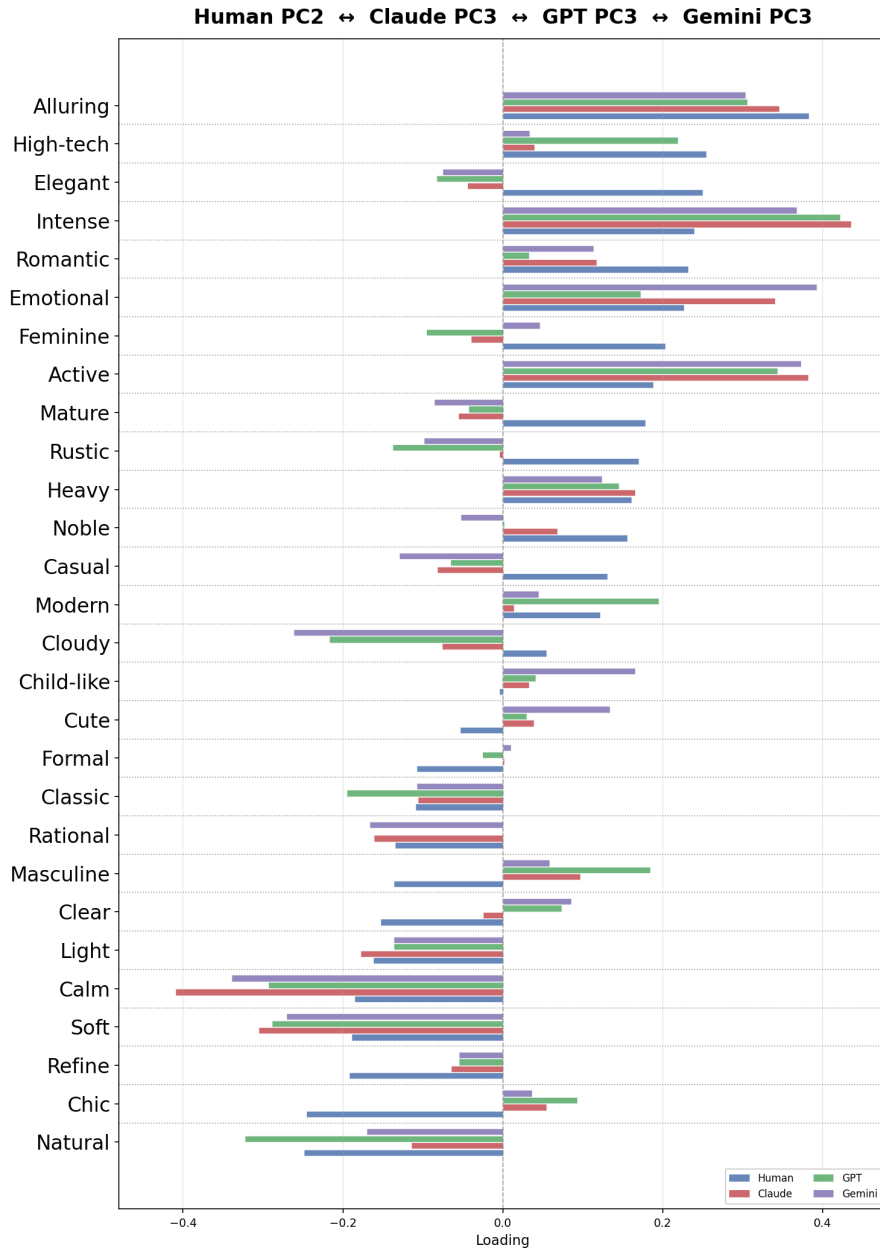


Figure 5 PCA loadings for PC2/PC3 (Arousal vs. Serenity) across Human and three AI groups. This dimension corresponds to Human PC2 and AI PC3.

먼저, 첫 번째 차원(Figure 4)은 Mature, Heavy, Masculine, Formal, Noble 등의 형용사가 한 극을 이루고, Cute, Child-like, Light, Soft, Clear 등이 반대 극을 형성하는 Gravitas vs. Cuteness(중후함 대 귀여움) 차원으로, 네 집단의 loading 패턴이 가장 유사하게 나타나, 일관된 구조를 보임을 확인하였다. 때문에 PC1 차원은 색채의 무계감 지각에 있어, 사람과 AI 간의 일치도가 높으며, 모든 집단에서 공통적으로 나타나 색채-감성 평가의 가장 보편적인 구조임을 확인할 수 있었다.

다음으로, 두 번째 차원(Figure 5)은 색약자 집단 PC2와 AI PC3에 대응되며, Intense, Active, Alluring, Emotional이 양(+)의 부하량을, Calm, Soft, Natural이 음(-)의 부하량을 보여 Arousal vs. Serenity(각성

대 평균) 차원으로 해석된다. 집단 간 구조적 유사성이 높았으나, 색약자 집단에서는 이 차원이 PC2(설명율 19.6%)로 두 번째로 중요한 반면 AI에서는 PC3(14-15%)로 세 번째에 위치하였다. AI 3종(ChatGPT, Claude, Gemini) 간에는 뚜렷한 차이가 나타나지 않았다.

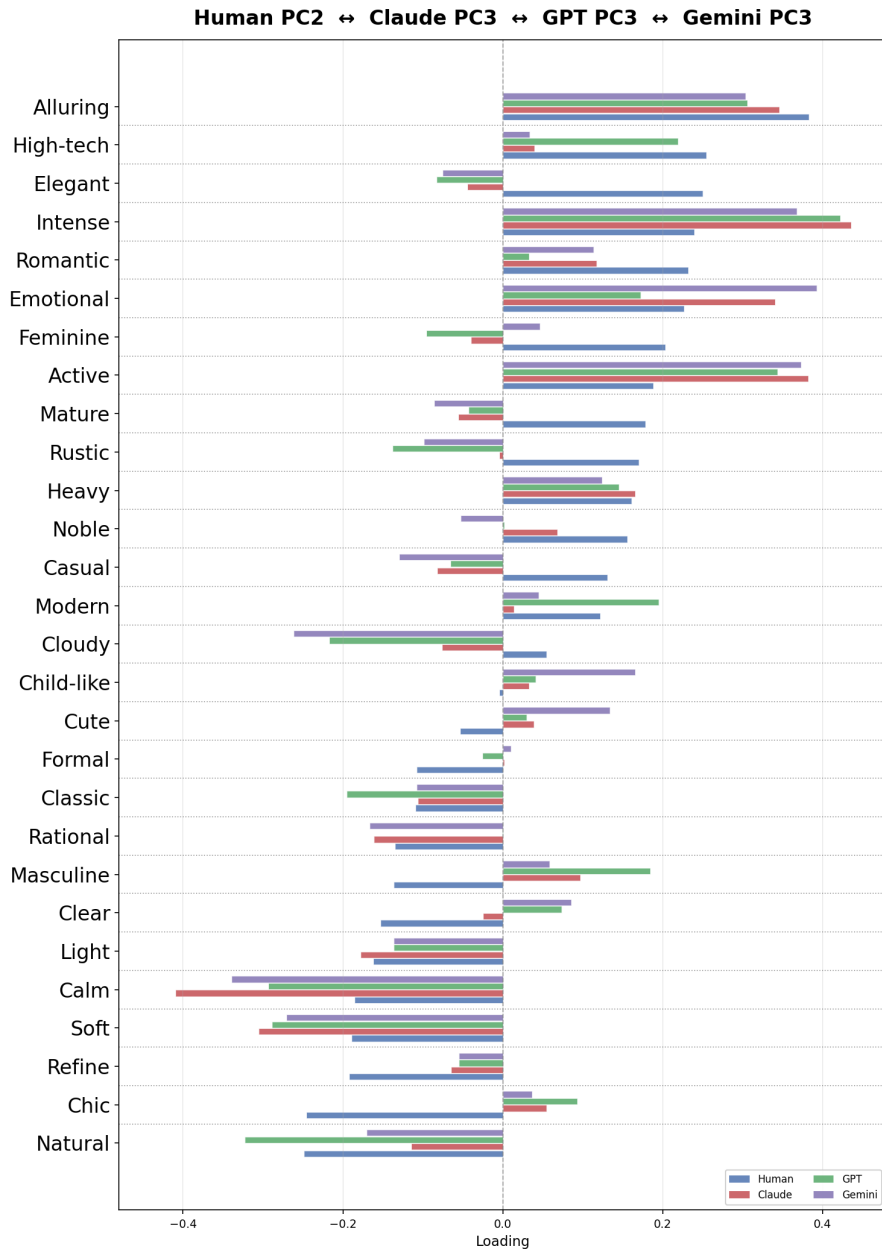


Figure 6. PCA loadings for PC3/PC2 (Rusticity vs. Modernity) across Human and three AI groups. This dimension corresponds to Human PC3 and AI PC2.

세 번째 차원(Figure 6)은 색약자 집단 PC3와 AI PC2에 대응되며, Classic, Natural, Rustic이 양(+), Modern, High-tech, Refine, Elegant가 음(-)의 부하량을 보여 Rusticity vs. Modernity(소박함 대 현대성) 차원으로 해석된다. AI 집단에서도 동일한 구조가 확인되었으며, 이 차원은 AI에서 PC2(29-34%)로 두 번째로 중요하게 나타났다. AI 3종 간 구조는 전반적으로 유사하였다.

이러한 두 번째와 세 번째 주성분의 순서 역전은 색약자 집단과 AI의 색채-감성 공간 조직 방식의 차이를 반영한다. 색약자 집단은 Arousal vs. Serenity 차원을 Rusticity vs. Modernity 차원보다 우선적으로 지각하는 반면, AI는 Rusticity vs. Modernity의 대비를 보다 중요한 구조적 축으로 인식하는 경향이 있었다.

5. 논의

5. 1. 색약자 집단과 AI의 색채 감성 평가 일치도

ICC 분석 결과, 색약자 집단은 대부분의 형용사에서 낮은 수준의 일치도를 나타냈다($M = 0.20$). Jonauskaitė et al.(2021)은 적녹 색각 이상자의 색채-감성 연합이 일반 색각자와 부분적으로 다르며 집단 내 개인 간 변동성도 크다고 보고하였다. 이는 색약자를 단일 집단으로 간주하기 어려우며, 색각 이상의 유형과 정도에 따라 색채 감성 반응이 다양하게 나타나는 것으로 해석할 수 있다. 예외적으로 Light($ICC = 0.51$)와 Heavy($ICC = 0.50$)에서만 보통 수준의 일치도가 관찰되었다. Sato & Inoue(2016)는 적녹 색각 이상자가 warmth 평가 시 luminance cone-contrast에 의존하는 경향이 있음을 보고하였는데, 명도 기반 차원에서 상대적으로 높은 일치도가 나타난 본 연구의 결과가 이와 관련될 가능성이 있다.

이와 대조적으로 AI는 현저히 높은 집단 내 일치도를 보였다(Claude: $M = 0.70$). 색약자 집단($M = 0.20$)과 AI 간의 ICC 차이는 AI가 색약자 집단의 개인 간 다양성을 시뮬레이션하지 못하고 있음을 시사하며, Bojic et al.(2025)의 결과와도 일치한다. AI의 색채 감성이 "평균"에 수렴한다는 점에서, AI 응답의 균일성을 색약자 집단의 대리 지표로 수용하는 것은 적절하지 않다. 브랜드 색채나 UI 배색처럼 일관성이 요구되는 작업에서는 AI가 유용한 참조 기준을 제공할 수 있으나, 색약자 사용자 대상의 감성 평가에서는 당사자 참여 기반의 검증이 반드시 병행되어야 한다.

5. 2. AI의 색약자 색채 감성 시뮬레이션 가능성과 한계

Pearson r 분석 결과, AI는 Light, Heavy, Soft 등 물리적·지각적 속성이 명확한 차원에서 색약자 집단과 높은 패턴 유사성을 보인 반면, Classic, High-tech, Natural, Alluring 등 맥락 의존적 형용사에서는 유사성이 현저히 낮았다. 이러한 비대칭성은 일반 색각 집단을 대상으로 한 Zheng & Xu(2025)에서도 동일하게 보고되었다. 단, 두 연구는 측정 도구가 달라 수치의 직접 비교는 어려우며 패턴 수준의 비교로 제한된다.

두 연구를 종합하면, AI가 색약자 집단의 색채 감성을 특별히 못 따라가는 것이 아니라, 일반 색각·색약자 집단 모두에서 맥락 의존적 차원을 포착하지 못하는 AI의 구조적 한계가 공통적으로 나타나는 것으로 해석된다. ANOVA에서 확인된 체계적 편향, 즉 AI가 강렬·역동적 차원은 낮게, 정적·자연적 차원은 높게 평가하는 경향 역시 이 한계의 일부다. Haase & Pokutta(2024)의 인간-AI 공동창작 모델에서처럼, AI는 지각적 기반의 색채 구조 제안에, 인간 디자이너는 문화적·감성적 맥락 조율에 각각 강점을 발휘하는 협업 체계가 색약자 대상 색채 디자인에서도 유효할 것으로 전망할 수 있다.

5. 3. AI와 색약자 집단의 색채 감성 공간 구조 비교

PCA 결과, 색약자 집단과 AI 모두 Gravitas vs. Cuteness, Arousal vs. Serenity, Rusticity vs. Modernity의 세 차원을 공유하였다. 그러나 두 번째와 세 번째 주성분의 순서가 역전되어, 색약자 집단은 Arousal vs. Serenity를, AI는 Rusticity vs. Modernity를 더 우선적으로 처리하였다. Sato & Inoue(2016)에 따르면 색약자 집단은 색채 감성 평가 시 luminance와 blue-yellow 채널에 의존하는 경향이 있는데, 이러한 지각적 특성이 명도 변화에 민감한 Arousal vs. Serenity 차원을 더 현저하게 지각하게 만들 가능성이 있다. Elyoseph et al.(2023)은 LLM이 인간과 유사한 정서적 범주 구조를 보유하면서도 처리 방식에서 차이를 보인다고 지적하였는데, 본 연구는 이를 색약자의 색채 감성 영역에서 구체적으로 확인하였다. 다만 이러한 순서 역전이 색약자 특유의 지각 특성에서 기인하는지는 일반 색각 집단과의 비교 연구를 통해 검증될 필요가 있다.

디자인 관점에서, AI는 범주적 색채 구조 분류에 강점을 가지며, 색약자 사용자의 각성·정서 반응이 중요한 단계에서는 인간 디자이너의 판단이 이를 보완하는 역할 분담이 적합하다.

5. 4. 색약자를 위한 접근성 디자인에서의 함의

앞선 분석 결과를 종합하면, AI는 색약자 집단의 색채 감성을 부분적으로만 반영할 수 있음이 확인된다. AI는 집단 내 균일성이 지나치게 높아 색약자 집단의 개인 간 다양성을 반영하지 못하며(5.1), 문화적·맥락적 차원에서 색약자 집단과의 패턴 유사성이 낮고(5.2), 색채 감성 공간의 차원 우선순위에서도 차이를 보인다(5.3). Hayashi et al.(2025)은 대규모 시각-언어 모델이 색각 이상에 대한 사실적 지식은 보유하고 있지만, 색각 이상자가 실제 경험하는 지각적 결과를 재현하지 못하고 표준적 색채 지각 방식으로 수렴하는 경향이 있음을 보고하였다. 본 연구의 결과 역시 AI가 색약자 집단의 luminance 의존적 색채 지각 방식이나 감성 공간의 우선순위 차이를 충분히 내재화하고 있지 못함을 시사한다.

포용적 디자인 관점에서, 이러한 결과는 AI를 색약자 대상 디자인 평가 도구로 활용할 때 신중한 접근이 필요함을 시사한다. Tapang & Espineda(2024)가 강조하였듯, 색약자를 위한 디자인은 단순한 색채 조정을 넘어 당사자의 감성적 경험을 반영하는 포용적 접근이 필요하다. AI가 제공하는 색채 감성 평가는 무게감, 명도 대비 등 물리적 속성 기반의 초기 스크리닝 단계에서는 유용하게 활용될 수 있으나, 색약자 집단의 감성적 경험을 반영한 최종 검증은 당사자 참여 기반의 사용자 연구로 보완되어야 한다. 결과적으로, 색약자 집단의 색채 감성 데이터를 AI 학습에 포함시키는 방향으로 모델을 개선한다면, AI 기반 접근성 디자인 도구의 신뢰도를 높이는 데 기여할 수 있을 것이다.

6. 결론

본 연구는 색약자 집단의 색채 감성 반응과 세 가지 생성형 AI 모델의 응답을 비교함으로써, AI가 시각적 다양성을 가진 집단의 색채 감성을 대리할 수 있는지를 검토하였다. ICC, Pearson r, One-way ANOVA, PCA 분석을 통해 색약자 집단과 GPT, Claude, Gemini 세 모델을 비교한 결과, AI는 물리적·지각적 속성이 명확한 차원에서는 색약자 집단과 유사한 패턴을 보여 색채 기획의 초기 단계에서 활용 가능성을 확인하였으나, 맥락 의존적 차원에서의 패턴 유사성 부족, 체계적 편향, 집단 내 과도한 균일성, 색채 감성 공간의 차원 우선순위 차이 등의 한계도 확인되었다. 이는 현재의 AI 모델이 색약자 집단의 색채 감성을 충분히 반영하지 못함을 시사하며, 포용적 디자인 맥락에서 AI를 활용할 때는 반드시 당사자 참여 기반의 검증이 병행되어야 한다.

본 연구는 색약자 집단만을 대상으로 하여 일반 색각 집단과의 직접 비교가 이루어지지 않았으며, 단일 문화권 데이터를 활용하였다는 한계를 가진다. 향후 연구에서는 일반 색각 집단과의 비교를 통해 색약자 특수성을 규명하고, 색약자 데이터를 AI 학습에 통합하는 방향의 모델 개선 연구가 수행될 필요가 있다.

References

1. Bojic, L., Kovacevic, P., & Cabarkapa, M. (2025). Comparing large Language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm. *Scientific Reports*, 15, 11477.
2. Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lvovsky, M. (2023). ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology*, 14, 1199058.
3. Haase, J., & Pokutta, S. (2024). Human-AI co-creativity: Exploring synergies across levels of creative collaboration. *arXiv:2411.12527*. <https://arxiv.org/abs/2411.12527>

4. Hayashi, K., Ozaki, S., Sakai, Y., Kamigaito, H., & Watanabe, T. (2025). Diagnosing vision language models' perception by leveraging human methods for color vision deficiencies. *arXiv:2505.17461*. <https://arxiv.org/abs/2505.17461>
5. Jonauskaitė, D., Abu-Akel, A., Dael, N., et al. (2020). Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. *Psychological Science*, 31(10), 1245–1260.
6. Jonauskaitė, D., Camenzind, L., Parraga, C. A., et al. (2021). Colour-emotion associations in individuals with red-green colour blindness. *PeerJ*, 9, e11180.
7. Jonauskaitė, D., Epicoco, D., Al-Rasheed, A. S., et al. (2024). A comparative analysis of colour-emotion associations in 16–88-year-old adults from 31 countries. *British Journal of Psychology*, 115, 275–305.
8. Kobayashi, S. (1981). The aim and method of the color image scale. *Color Research & Application*, 6(2), 93–107.
9. Kho, J., Kim, J., Park, T., Liang, C., & Lee, J. H. (2026, May 22–23). *Comparative analysis of color image perception between normal color vision and color vision deficiency for fashion color planning* [Poster presentation]. 2026 International Conference on Clothing and Textiles, Seoul, South Korea.
10. Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
11. Machado, G. M., Oliveira, M. M., & Fernandes, L. A. (2009). A physiologically-based model for simulation of color vision deficiency. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1291–1298.
12. Miotto, M., Rossberg, N., & Kleinberg, B. (2022, November). Who is GPT-3? An exploration of personality, values and demographics. In *Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+ CSS)* (pp. 218–227).
13. Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.
14. Ou, L. C., Luo, M. R., Woodcock, A., & Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, 29(3), 232–240.
15. Sakamoto, T., & Ichihara, Y. G. (2019, July). Exploring Color-Universal Design Considering Kansei Differences: Color-Vision Types and Impressions of Color Images. In *International Conference on Human-Computer Interaction* (pp. 62–72). Cham: Springer International Publishing.
16. Sato, K., & Inoue, T. (2016). Perception of color emotions for single colors in red-green defective observers. *PeerJ*, 4, e2751.
17. Schloss, K. B. (2024). Color semantics in human cognition. *Current Directions in Psychological Science*, 33(1), 58–67.
18. Tapang, R. K., & Espineda, M. N. E. (2024). Empowering inclusivity: Understanding and designing for color vision deficiency in consumer behavior and product packaging. *Archives of Design Research*, 37(4), 27–41.
19. Wright, B., & Rainwater, L. (1962). The meanings of color. *Journal of General Psychology*, 67(1), 89–99.
20. Zheng, L., & Xu, L. (2025). Comparing AI and human emotional responses to color: A semantic differential and word-color association approach. *Color Research & Application*, 50, 286–300.

생성형 AI와 색각 이상자의 색채 감성 반응 비교 연구

박수진¹, 고정민², 김지은³, 이상원⁴, 이지현⁴, 허룡^{5*}

¹연세대학교 인간생애와 혁신적디자인 융합전공 박사후연구원, 서울, 대한민국

²연세대학교 인간생애와 혁신적디자인 융합전공 연구교수, 서울, 대한민국

³연세대학교 통합디자인학과 객원교수, 서울, 대한민국

⁴연세대학교 통합디자인학과 교수, 서울, 대한민국

⁵칭다오이공대학교 제품디자인과, 칭다오, 중국

초록

연구배경 색채 감성 연구는 일반 색각자를 표준으로 발전해왔으며, 색각 이상(색약자) 집단의 색채 감성 특성은 상대적으로 간과되어 왔다. AI의 색채 감성 재현 능력에 관한 연구도 일반 색각 집단에 한정되어, AI가 색약자 집단의 색채 감성을 얼마나 충실하게 반영할 수 있는지는 검토되지 않은 상태이다.

연구방법 NCD 130 표색계의 130개 색 표와 28개 감성 형용사를 자극 및 측정 도구로 사용하여, 색약자 집단의 색채 감성 평가 데이터를 LLM 3종의 색약자 유형별 페르소나 응답 데이터와 비교하였다. 급내상관계수, 피어슨 상관분석, 일원분산분석, 주성분 분석을 통해 집단 내 일치도, 패턴 유사성, 평균 차이, 차원 구조를 분석하였다.

연구결과 AI는 지각적 속성이 명확한 차원에서 색약자 집단과 높은 패턴 유사성을 보인 반면, 맥락 의존적 차원에서는 유사성이 현저히 낮았다. AI는 역동적 감성을 낮게, 정적이고 자연적인 감성을 높게 평가하는 체계적 편향을 보였으며, 색약자 집단에 비해 과도하게 균일한 응답을 생성하였다. 동일한 3차원 구조를 공유하였으나, 색약자 집단은 각성 차원을, AI는 소박함-현대성 차원을 우선적으로 처리하였다.

결론 AI의 한계는 색약자 특수성보다 맥락 의존적 차원을 포착하지 못하는 AI의 구조적 속성에서 기인하며, 이는 일반 색각·색약자 집단 모두에서 공통적으로 작동한다. 색약자 집단이 각성 차원을, AI가 소박함-현대성 차원을 우선 처리하는 역전 패턴은 색약자의 명도 의존적 지각 특성이 색채 감성 공간의 조직 방식 자체에 영향을 미칠 수 있음을 시사한다. 포용적 디자인에서 AI는 물리적 속성 기반의 초기 스크리닝에, 인간 디자이너는 맥락적이고 감성적인 판단에 각각 강점을 발휘하는 협업 체계가 설계될 필요가 있다.

주제어 색채 감성, 색각 이상, 대규모 언어 모델, 생성형 AI, 의미미분법

*교신저자 : 허룡 (xulong@qut.edu.cn)