

# ChatGPT 사용자 경험 측정모형의 구성타당도와 이해타당도 평가

## An Assessment of Construct Validity and Nomological Validity of the Measurement Model for ChatGPT UX Evaluation

주 저 자 : 정다운 (Jeong, Da Un)

홍익대학교 국제디자인전문대학원 디자인학 박사과정

교 신 저 자 : 나 건 (Nah, Ken)

홍익대학교 국제디자인전문대학원 디자인경영학과 교수  
knahidas@gmail.com

<https://doi.org/10.46248/kidrs.2024.4.108>

접수일 2024. 11. 25. / 심사완료일 2024. 12. 01. / 게재확정일 2024. 12. 09. / 게재일 2024. 12. 30.

## Abstract

The present study aims to evaluate the construct validity and nomological validity of the measurement model for ChatGPT UX evaluation. A careful literature review led to the identification of a total of thirty-six UX evaluation metrics.  $N=16$  respondents participated in Delphi and nine ChatGPT UX evaluation metrics were selected, including performance expectancy, effort expectancy, usability, capability, communication, co-creation, effectiveness, innovativeness, and interaction quality. A survey study was conducted and  $N=212$  respondents rated their ChatGPT UX via fifty-six measurement items. An exploratory factor analysis was performed on the collected survey data and the result indicates that two UX metrics of usability and effectiveness were excluded. A confirmatory factor analysis was conducted to assess the construct validity of the measurement model, and the results reveal that the validity level of the measurement model proposed in the present work is fairly satisfactory. Lastly, the results from the test of nomological validity of the measurement model indicate that, the seven UX evaluation metrics collectively constitute a single second-order construct representing ChatGPT UX.

## Keyword

챗지피티(ChatGPT), 인공지능 사용자 경험(Artificial intelligence UX), 이해타당도(Nomological validity)

## 요약

본 연구의 목적은 ChatGPT UX 평가지표로 이루어진 측정모형의 타당도를 평가함에 있다. 이를 위하여 본 연구는 문헌고찰을 수행하여 총 36개의 UX 평가지표를 도출하였다. 그리고  $N=16$ 명이 참여한 델파이를 통해 얻어진 내용 타당도 지수를 기반으로 성과기대와 노력기대, 유용성, 능력, 커뮤니케이션, 공동창작, 효과성, 혁신성, 상호작용 품질을 포함하는 9개의 ChatGPT UX 평가지표를 잠재적으로 선정하였다. 그 후  $N=212$ 명의 조사 대상자가 설문조사에 참여하여 총 56개의 설문문항을 평가하였다. 수집된 자료에 대해 탐색적 요인분석을 수행한 결과, 유용성과 효과성이 제외되어 UX 평가지표의 수는 총 7개로 축소되었다. ChatGPT UX의 측정모형에 관한 구성타당도를 평가하기 위하여 확인적 요인분석을 수행한 결과, 본 연구가 제시한 측정모형의 타당도 수준은 양호한 것으로 나타났다. 마지막으로, 1차 구성개념과 2차 구성개념 간의 이해타당도를 통계적 유의수준에서 검증한 결과, 7개의 UX 평가지표는 ChatGPT UX를 단일개념으로 구성하는 것으로 나타났다.

## 목차

### 1. 서론

- 1-1. 연구의 배경 및 필요성
- 1-2. 연구의 목적 및 방법

### 2. 문헌고찰

- 2-1. ChatGPT의 이해
- 2-2. AI 서비스의 사용자 경험

### 3. 연구 방법

- 3-1. ChatGPT 사용자 경험 평가지표 도출
- 3-2. 델파이 연구
- 3-3. 설문도구 설계
- 3-4. 조사 대상자 및 자료의 수집
- 3-5. 자료의 분석 방법

### 4. 연구 결과

- 4-1. 기술통계량

- 4-2. 내적 일관성 평가 결과
- 4-3. 탐색적 요인분석 수행 결과
- 4-4. 확인적 요인분석 수행 결과

## 5. 결론

### 참고문헌

#### 부록

## 1. 서론

### 1-1. 연구의 배경 및 필요성

대화형 인공지능(artificial intelligence; AI) 모델 중의 하나인 ChatGPT는 2022년도에 출시한 이후 잠재적인 게임 체인저로 평가받으며 인간의 삶에 깊이 개입하고 있다. ChatGPT는 기존의 대화형 사용자 인터페이스와는 다르게, 최신의 인공지능 학습 언어 모델을 기반으로 광범위한 주제를 가지는 사용자의 프롬프트로부터 구체적인 응답을 생성할 수 있다는 특징을 가진다.<sup>1)</sup> ChatGPT가 제공하는 폭넓은 기능과 정교한 자연어 처리 능력을 통해 사용자가 기술과 상호작용하는 방식을 근본적으로 변화시키고 있다.<sup>2)</sup>

ChatGPT를 포함하는 대화형 AI 서비스가 인간의 삶에 미치는 잠재적인 영향력을 고려한다면, 사용자가 이러한 서비스를 어떻게 경험하는지에 대한 초기 정보를 수집하는 것은 매우 중요하다. 이는 서비스의 성공 여부가 AI 기술의 우수성 뿐만 아니라 사용자 경험(user experience; UX)의 질적 수준에 의해 결정되기 때문이다.

ChatGPT는 전통적인 소프트웨어와 달리 고도로 비정형적이고 동적이며, 사용자의 질문 유형과 맥락, 그리고 사용자의 의도에 따라 결과물을 생성하는 특징을 가진다. 이와 더불어, ChatGPT는 단순히 정답을 제공하는 도구가 아니라, 사용자의 학습과 창작, 문제 해결을 지원하는 지능적인 협업자로 기능한다. 이러한 맥락에서 사용자가 서비스를 얼마나 유용하게 느끼는지, 그

리고 주어진 맥락에서 생성된 응답이 적합하고 신뢰할 수 있는지 등을 평가하는 것이 핵심이다.<sup>3)</sup> 이와 함께, ChatGPT의 UX는 인간-인공지능 간의 커뮤니케이션, 그리고 사용자의 기대와 요구를 얼마나 잘 충족시키는지에 따라 궁극적으로 결정된다. 이에 따라 ChatGPT로부터 사용자가 체험하는 UX의 복합적인 특성을 다루기 위해서는 ChatGPT의 고유한 특징을 반영하는 정교한 UX 평가지표의 설계가 필수적이다.

그럼에도 불구하고 ChatGPT UX의 평가지표 설계에 관한 연구의 사례는 소수의 관련 문헌에서만 관찰되고 있다.<sup>4)</sup> 예를 들어서, ChatGPT의 초기 사용자를 대상으로 인터뷰 기반의 질적 연구를 수행한 Skjuve et al.(2023)<sup>5)</sup>은 주제분석(thematic analysis)의 수행 결과를 통해 실용적-쾌락적 경험이 어떻게 형성되는지에 대한 통찰력을 제공한다는 점에서 의의를 가진다고 평가할 수 있다. 그러나 ChatGPT 서비스는 실용 및 쾌락을 넘어 사용자가 기대하는 심층적인 이해와 창의적인 문제 해결 능력, 그리고 공감적인 상호작용 등의 요인을 포함할 정도로 서비스 경험이 광범위하다.<sup>6)</sup> 따

- 1) Niu, Ben, and Gustave Florentin Nkoulou Mvondo. 'I Am ChatGPT, the ultimate AI Chatbot! Investigating the determinants of users' loyalty and ethical usage concerns of ChatGPT.' *Journal of Retailing and Consumer Services*, 2024, vol.76, p.02
- 2) HO, Brittany, et al. 'ChatReview: A ChatGPT-enabled natural language processing framework to study domain-specific user reviews.' *Machine Learning with Applications*, 2024, 03, Vol.15, p.03

- 3) Kim, Jeong Soo, Minseong Kim, and Tae Hyun Baek. "Enhancing User Experience With a Generative AI Chatbot." *International Journal of Human-Computer Interaction*, 2024, 02, Vol40, No.22 pp.1-13
- 4) 안무정, 강태임. '디지털 트랜스포메이션 경영을 위한 챗GPT 사용자 경험(UX) 디자인 평가: 오픈AI 챗GPT와 마이크로소프트 Bing 챗GPT 교차활용을 중심으로'. *한국디자인문화학회지*, 2023, 06, Vol.29, No.2, pp237-247
- 5) Skjuve, Marita, Asbjørn Følstad, and Petter Bae Brandtzaeg. 'The user experience of ChatGPT: findings from a questionnaire study of early users.' *Proceedings of the 5th international conference on conversational user interfaces*. 2023. pp.1-10.
- 6) Ashfaq, Muhammad, et al. 'I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service

라서 Skjuve et al.(2023)이 다루었던 평가 지표만으로는 ChatGPT가 제공하는 UX의 본질을 평가하기 어렵다는 한계를 가진다. ChatGPT의 UX를 평가하기 위해서는 정량적인 데이터를 기반으로 하는 평가지표의 탐구가 필수적이다. 이는 단순히 사용성(usability)의 수준을 측정하는 것을 넘어, 인간-AI 상호작용의 본질을 탐색하고, 이를 통해 사용자 중심의 혁신을 이끌어 내기 위한 필연적인 연구 과제이기도 하다.

## 1-2. 연구의 목적 및 방법

본 연구의 목적은 ChatGPT UX 평가지표를 탐색함에 있다. 그리고 이들 평가지표로 이루어진 ChatGPT UX 측정모형의 구성타당도(construct validity)를 확인함에 있다. 이와 더불어, 다수의 UX 평가지표가 ChatGPT UX를 단일개념으로 구성할 수 있는지를 확인하기 위하여 1차 구성개념(first-order construct)과 2차 구성개념(second-order construct) 간의 이해타당도(nomological validity)를 통계적 유의수준에서 검증함에 있다.

이에 관한 구체적인 연구 방법은 다음과 같다. 첫째, 문헌고찰과 델파이(Delphi)의 수행을 통해 ChatGPT UX를 평가할 수 있는 잠재적인 평가지표를 도출하였다. 둘째, 디자인 실무 경력을 가지는 N=212명의 조사대상자로부터 설문자료를 수집하였다. 셋째, 수집된 자료에 대하여 탐색적 요인분석(exploratory factor analysis)을 수행하여 평가지표의 잠재요인 구조를 확인하였다. 넷째, 확인적 요인분석(confirmatory factor analysis)을 수행하여 ChatGPT UX 측정모형의 구성타당도와 이해타당도를 검증하였다.

## 2. 문헌고찰

### 2-1. ChatGPT의 이해

현대 기술은 인간의 사고를 모방하고 다양한 방식으로 인간을 돕는 AI에 크게 의존하고 있다. AI는 오랜 역사를 가지고 있지만, 최근 몇 년 동안 획기적인 발전이 이루어져 왔다. 이러한 발전은 AI 기술을 기반으로 하는 챗봇의 개발과 출시로 구체화되었으며, 특히 최근에는 ChatGPT의 사례를 통해 AI의 발전 수준을 가늠할 수 있다.

ChatGPT는 OpenAI가 개발한 언어 모델인 GPT-3의 개선된 형태를 기반으로 개발되었다(Sohail, 2023). GPT-3에 사용된 딥러닝 신경망(Deep Learning Neural Network, DLNN)은 약 1,750억 개의 머신러닝(ML) 매개변수를 포함하고 있다. GPT-3가 설계되기 이전에 마이크로소프트의 Turing-Natural Language Generation(T-NLG) 프레임워크가 100억 개의 매개변수를 포함하고 있다는 점을 고려한다면, GPT-3는 현재까지 구축된 가장 큰 신경망으로 평가되고 있다.<sup>7)</sup>

OpenAI의 ChatGPT는 채팅 환경에서 입력되는 사용자의 데이터를 바탕으로 인간이 답변하는 방식과 유사한 텍스트 응답을 생성하도록 설계되었다. 그리고 방대한 양의 데이터를 학습 데이터로 활용하여 다양한 주제와 질문에 응답할 수 있도록 개발되었다. 그러나 생성형 AI 언어 모델은 학습 데이터를 기반으로 학습한 패턴과 구조에 기반하기 때문에 인간이 수행하는 맥락적 인식 및 이해는 부족하다는 점이 지적되고 있다.<sup>8)</sup> 그럼에도 불구하고 ChatGPT는 현재 약 100개국의 언어를 지원하며 전 세계 사용자에게 언어 번역과 데이터 분석, 디자인 생성, 콘텐츠 제작, 텍스트 생성 등의 분야에서 활용되고 있다.

ChatGPT는 비즈니스 세계와 학문적 영역 모두에서 큰 관심을 불러 일으키고 있다(Farhat et al., 2023). ChatGPT가 도입된 이후 이를 주제로 삼은 연구가 폭발적으로 증가하고 있으며, ChatGPT가 공동 저자로 참여한 사례도 관찰되고 있다(King & ChatGPT, 2023). ChatGPT의 주요 역할은 인간 사용자가 제공하는 명령 또는 입력을 기반으로 인간 대화를 모방하는 것이다. 이는 Google의 Alexa 또는 Apple의 Siri와 같은 AI 비서 및 시스템 앱과 유사하다. 또한 ChatGPT는 새로운 데이터를 학습함으로써 지속적으로 개선될 수 있는 머신러닝 모델을 기반으로 설계되었다. 그리고 사용자는 특정 응답에 대한 피드백을 제공함으로써 ChatGPT의 개선에 기여할 수 있다.<sup>9)</sup>

agents. Telematics and Informatics, 2020, 11, Vol.54, p.14

7) Gupta, B., Mufti, T., Sohail, S. S., & Madsen, D. Ø. 'ChatGPT: A brief narrative review. Cogent Business & Management' 2023, 12, Vol.10, No.3, pp.1-17

8) Bozkurt, Aras. 'Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift.' Asian Journal of Distance Education. 2023. Vol.18, No.1, p.198

9) Hosseini, M., & Horbach, S. P. 'Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other

ChatGPT는 출시 이후 대중들로부터 큰 인기를 얻고 있으나, 일부 비판자들은 ChatGPT와 같은 AI 생성형 모델이 인간 문명에 위협이 될 수 있다고 경고하고 있다.<sup>10)</sup> 또한 ChatGPT가 가지는 다수의 한계가 문제점이 지적되고 있는데, 예를 들어서 ChatGPT는 완벽하지 않으며, 학습 알고리즘에 의존하기 때문에 결과물의 정확도가 낮아질 수 있다는 점이 제기되고 있다. 또한, 편향되거나 허위 정보를 제공할 가능성이 있으며, ChatGPT의 사용이 여러 가지 형태의 윤리적 논란을 불러 일으킬 가능성이 또한 논의되고 있다.<sup>11)</sup>

## 2-2. AI 서비스의 사용자 경험

AI 서비스의 UX는 오늘날 디지털 환경에서 중요한 연구 주제로 부각되고 있다.<sup>12)</sup> 이는 인간의 일상 활동에서 AI가 다양한 형태로 사용자와의 상호작용이 보편화된 관행으로 자리를 잡고 있기 때문이다. AI 서비스에서의 UX는 단순히 AI 기술의 기능적 효용성을 넘어서, 심리적, 사회적, 그리고 정서적 체험을 종합적으로 고려해야 하는 다차원적이고 복잡한 개념이다. 따라서 AI 서비스의 UX 연구에 관해서는 인간과 기술의 상호작용과 함께 이 과정에서 체험하는 UX의 특징을 이해할 수 있는 다학제적 접근이 요구된다.

예를 들어서, AI 서비스에 관한 UX 평가에서 가장 먼저 고려되어야 할 점은 사용자와 AI 기술 간의 소통 방식이다. 자연어 처리 기술을 활용한 인터페이스는 사용자가 서비스를 직관적으로 이해하고 사용할 수 있도록 설계된다. 그러나 이러한 기술이 고도화될수록 사용자는 기술의 의도를 오해하거나 예상치 못한 결과를 경험할 가능성이 높아진다.<sup>13)</sup> 이로 인하여 '신뢰'라는

large language models in scholarly peer review.' Research Integrity and Peer Review, 2023, 05, Vol.8. No.1, p.4

10) Yuval Noah Harari argues that AI has hacked the operating system of human civilisation.(2024, 11.10) [www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation](http://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation)

11) Rahimi, F., & Abadi, A. T. B. 'ChatGPT and publication ethics'. Archives of Medical Research, 2023. 04, Vol.54, No.3, pp.272-274

12) 안무정, 강태임. '디지털 트랜스포메이션 경영을 위한 챗GPT 사용자 경험(UX) 디자인 평가: 오픈AI 챗GPT와 마이크로소프트 Bing 챗GPT 교차활용을 중심으로'. 한국디자인문화학회지, 2023. Vol.29, No.2, pp.237-247

요소가 부각될 수 있으며, 이는 AI의 행동이 투명하고 예측 가능한지를 나타내는 UX 평가지표로서의 역할을 담당할 것으로 예상된다.

또한, AI 서비스의 UX는 감정적 차원의 경험을 포함한다. 사용자는 AI에 대한 기술적 효율성뿐만 아니라, 서비스와의 상호작용에서 정서적 공감이나 만족감을 체험하기를 원한다.<sup>14)</sup> 예를 들어서, AI가 사용자 의도를 정확히 파악하거나, 인간다운 대화의 톤을 유지할 수 있을 때, 사용자는 높은 수준의 만족도를 체험할 가능성이 커지게 된다. 주목해야 할 점은, AI와의 정서적 상호작용이 AI가 제공하는 문제 해결 능력보다 더 강력한 사용자 충성도를 만들어낼 수 있다는 점이다.<sup>15)</sup> 예컨대, 챗봇 서비스가 단순한 답변을 제공하는 것을 넘어 사용자가 정서적 불안이나 스트레스를 느끼는 상황에서 동료처럼 반응하는 능력을 가진다면 강력한 사용자 경험을 창출할 것으로 예상된다.<sup>16)</sup>

한편, 기술적인 유용성과 효과성은 AI 서비스의 UX 평가에서 간과할 수 없는 요소이다. 모든 사용자, 특히 디지털 격차에 놓인 계층도 서비스에 쉽게 접근하여 주어진 업무를 성공적인 수행하도록 AI 서비스가 충분한 역량을 가지는지의 여부를 판단할 수 있는 UX 평가 지표를 마련하는 것이 중요하다. 마지막으로, 지속 가능성 측면에서의 UX도 주목할 만하다.<sup>17)</sup> 지속 가능한 UX는 사용자의 피로도를 줄이고 장기적인 기술 적응을 도모하는 것을 목표로 한다. 이는 AI 서비스가 처음에는 흥미롭고 매력적으로 느껴지더라도, 장기적으로

13) Abbas, A., Imran, K., & Ting, C-Y. 'User experience design using machine learning: A systematic review.' IEEE Access, 2022, Vol.10, pp.51501-51514

14) Srinivasan, R. & gonzalez, B. S. M. 'The role of empathy for artificial intelligence accountability.' Journal of Responsible Technology. 2022. 04, Vol.9, p.100021

15) Seyitoğlu, F. & Ivanov, S. 'Robots and emotional intelligence: A thematic analysis.' Technology in Society, 2024. 06, Vol.77, p.102512

16) Andrade-Arenas, L. & Yactayo-Arias, C. 'Chatbot with ChatGPT technology for mental wellbeing and emotional management.' IAES International Journal of Artificial Intelligence, 2024. 09, Vol.13, No.3, p.2635

17) Park, A. & Lee, S. B. 'Examining AI and systemic factors for improved chatbot sustainability.' Journal of Computer Information Systems, 2024. 11, Vol.64, No.6, pp.728-742

사용자의 일상에 조화를 이루지 못하면 이탈로 이어질 수 있음을 시사한다. 지속 가능한 UX를 위해서는 사용자의 UX 수집과 반영이 필수적이며, 이는 서비스 설계와 개선의 순환 과정을 통해 달성된다.

### 3. 연구 방법

#### 3-1. ChatGPT 사용자 경험 평가지표 도출

선행연구로서 문헌을 조사한 결과 본 연구 범위의 ChatGPT UX평가지표를 수집하기에 연구자료의 수가 매우 한정적이라는 제약이 있었다. 따라서 ChatGPT를 포함한 AI 서비스에 관한 '사용 의도'의 선행 요인(antecedent)을 탐색하였다. 본 연구 목적인 ChatGPT 사용 의도의 선행 요인은 사용자가 AI 서비스에 접근하고 반복적으로 사용할 의지를 형성하게 하는 데 결정적인 역할을 하며, 이는 곧 ChatGPT UX의 수준을 평가하는 중요한 단서로 작용할 것으로 판단하였기 때문이다. 예를 들어서, 기술 수용 모델(Technology Acceptance Model)<sup>18)</sup>을 포함하는 사용자 수용 이론에서 사용 의도는 유용성과 사용 용이성에 의해 크게 영향을 받으며, 이는 UX 디자인에서 사용성을 평가하는 주요 기준으로 이미 널리 채택되고 있다. 유용성과 사용 용이성은 UX의 핵심적인 측면으로, 기술이 사용자의 문제를 얼마나 효율적으로 해결하는지와 얼마나 직관적으로 설계되었는지를 평가할 수 있는 중요한 지표이다. 따라서 사용 의도를 형성하는 선행 요인들은 UX를 정량적으로 평가할 수 있는 타당한 지표로 식별될 수 있으며, 평가지표 설계의 토대를 형성한다. [표 1]은 선행연구로부터 도출한 36개의 AI 서비스 사용의도의 선행요인을 나타내고 있다.

[표 1] ChatGPT 사용자 경험 평가지표 도출 결과

선행연구	사용자 경험 평가지표
Andrews et al.(2021), Cabrera-Sanchez et al. (2021),외 17건 Cao et al.(2021), Chi et al.(2020), Damerji & Salimi(2021),Fan et al.(2020), Gansser & Reich(2021), Gursoy et al.(2019), Lin et al.(2020), Ma & Huo(2023), Manzoor et al.(2024),	사회적 영향

18) Davis, F. D. 'Perceived usefulness, perceived ease of use, and user acceptance of information technology.' MIS Quarterly, 1989. 09, Vol.13, No.3, pp.319-340

Meyer-Waarden & Cloarec(2022), Prakash & Das(2021), Ribeiro et al.(2022), Roy et al.(2024), Sohn & Kwon(2020), Song & Kim(2020), Upadhyay et al.(2022)	
Cabrera-Sanchez et al. (2021) Chi et al.(2020), Choi(2021), Gansser & Reich(2021), Gursoy et al.(2019), Lin et al.(2020), Ma & Huo(2023), Manzoor et al.(2024), Meyer-Waarden & Cloarec(2022), Pillai et al.(2020), Ribeiro et al.(2022), Roy et al.(2024), Sohn & Kwon(2020), Upadhyay et al.(2022)	쾌락적 동기
Chi et al.(2020), Gursoy et al.(2019), Lin et al.(2020), Liu & Tao (2022), Ma & Huo(2023), Manzoor et al.(2024), Pelau et al.(2021), Roy et al.(2024), Song & Kim(2020)	의인화
Andrews et al.(2021), Cabrera-Sanchez et al. (2021), Cao et al.(2021), Chatterjee & Bhattacharjee(2020), Chi et al.(2020), Damerji & Salimi(2021), Fan et al.(2020), Gansser & Reich(2021), Gursoy et al.(2019), Lin et al.(2020), Ma & Huo(2023), Manzoor et al.(2024), Meidute-Kavaliauskiene et al.(2021), Meyer-Waarden & Cloarec(2022), Prakash & Das(2021), Ribeiro et al.(2022), Roy et al.(2024), Sohn & Kwon(2020), Upadhyay et al.(2022)	성과 기대
Andrews et al.(2021), Cabrera-Sanchez et al. (2021), Cao et al.(2021), Manzoor et al.(2024), Chatterjee & Bhattacharjee(2020), Chi et al.(2020), Damerji & Salimi(2021), Fan et al.(2020), Gursoy et al.(2019), Lin et al.(2020), Ma & Huo(2023), Meidute-Kavaliauskiene et al.(2021), Meyer-Waarden & Cloarec(2022), Prakash & Das(2021), Ribeiro et al.(2022), Roy et al.(2024), Sohn & Kwon(2020), Upadhyay et al.(2022)	노력 기대
Chi et al.(2020), Gursoy et al.(2019), Lin et al.(2020), Manzoor et al.(2024), Ribeiro et al.(2022), Roy et al.(2024)	감정
Ma & Huo(2023)	참신함
Andrews et al.(2021), Kashive et al.(2020),Gao & Huang(2019), Chatterjee & Bhattacharjee(2020), Chatterjee et al.(2021), Chiu et al.(2021), Gado et al.(2021), Kim et al.(2020), Liang et al.(2020), Ma & Huo(2023), Mohr & Kühl(2021), Sohn & Kwon(2020), Upadhyay et al.(2022)	태도
Chatterjee & Bhattacharjee(2020), Liang et al.(2020), Prakash & Das(2021),	지각된 위험

Ribeiro et al.(2022), Song & Kim(2020)	
Chatterjee et al.(2021), Choi(2021), Fan et al.(2020), Liu & Tao (2022), Meyer-Waarden & Cloarec(2022), Prakash & Das(2021),Ribeiro et al.(2022)	신뢰
Atwal et al.(2021)	지각된 이익
Atwal et al.(2021)	외부압력
Cabrera-Sanchez et al. (2021), Gansser & Reich(2021), Meidute-Kavaliauskiene et al.(2021), Sohn & Kwon(2020)	가치
Cao et al.(2021), Meyer-Waarden & Cloarec(2022)	행복
Cao et al.(2021)	자기발달
Chatterjee et al.(2021), Vu & Lim(2021) Damerji & Salimi(2021), Gado et al.(2021), Gao & Huang(2019), Kashive et al.(2020), Kim et al.(2020), Liang et al.(2020), Liu & Tao (2022), Mohr & Kühl(2021), Pillai et al.(2020), Sohn & Kwon(2020), Song & Kim(2020),	유용성
Chatterjee et al.(2021), Damerji & Salimi(2021), Gado et al.(2021), Gao & Huang(2019), Kashive et al.(2020), Kim et al.(2020), Liang et al.(2020), Liu & Tao (2022), Mohr & Kühl(2021), Pillai et al.(2020), Sohn & Kwon(2020), Song & Kim(2020),	용이성
Chiu et al.(2021), Choi(2021)	능력
Choi(2021)	명확성
Choi(2021), Fan et al.(2020)	복잡성
Gansser & Reich(2021), Upadhyay et al.(2022)	편리함
Gansser & Reich(2021)	지속성
Gao & Huang(2019), Liu & Tao (2022)	개인화
Gao & Huang(2019), Kim et al.(2020)	커뮤니케이션
Gao & Huang(2019)	공동창작
Kashive et al.(2020)	효과성
Kashive et al.(2020), Kim & Kim(2020)	만족감
Kim & Kim(2020)	품질
Kim & Kim(2020), Upadhyay et al.(2022)	불확실성 회피
Kim & Kim(2020), Meyer-Waarden & Cloarec(2022), Mohr & Kühl(2021), Song & Kim(2020)	혁신성
Kim & Kim(2020), Mohr & Kühl(2021), Sohn & Kwon(2020)	지각된행동통제
Kim & Kim(2020), Meyer-Waarden & Cloarec(2022), Mohr & Kühl(2021)	사회적규범
Kim & Kim(2020), Meyer-Waarden & Cloarec(2022)	공감
Kim & Kim(2020), Meyer-Waarden & Cloarec(2022)	상호작용품질
Kim & Kim(2020), Sohn & Kwon(2020)	기술성
Kim & Kim(2020), Song & Kim(2020)	의존성

### 3-2. 델파이 연구

본 연구는 문헌 연구를 기반으로 도출한 UX 평가지표의 타당성을 평가하고, 이들 지표가 ChatGPT의 고유한 특징과 사용자 맥락을 반영할 수 있는지의 여부를 조사하기 위하여 델파이를 수행하였다. 델파이 기법은 주제에 대한 전문성과 경험을 갖춘 전문가 집단의 의견 수렴을 통해 합의를 도출하는 방식으로, 객관적이고 체계적인 검증 과정이라는 점에서 널리 활용되고 있다<sup>19)</sup>

특히, UX 평가지표와 같이 주관적 판단이 개입될 수 있는 항목에 대해, 다양한 전문가들의 관점과 의견을 수집함으로써 지표의 실효성을 강화할 수 있다. 이 과정은 연구자의 개인적인 해석에 따른 편향을 최소화하고, 전문가 집단의 합의에 기반하여 최적화된 결과물을 도출한다는 점에서 높은 수준의 방법론적인 가치를 가진다.<sup>20)</sup>

본 연구에서는 5년 이상의 디자인 관련 경력을 가지고 있는 16명의 실무자 집단이 델파이에 참여하였다. 델파이에 참여한 전문가는 자기평가 기반의 5점 리커트(Likert) 척도(예., 1=전혀 타당하지 않다, 5=매우 타당하다)를 통해 36개의 UX 평가지표에 관한 타당도를 평가하였다. 수집한 자료에 대하여 내용 타당도 지수(content validity index; CVI)를 계산하여 .8 이상의 값을 가지는 지표를 최종 UX 평가지표로 선정하였다. 각각의 평가지표에 대한 CVI 값을 계산한 결과, 성과기대(=.88)와 노력기대(=.81), 유용성(=.88), 능력(=.88), 커뮤니케이션(=.88), 공동창작(=.81), 효과성(=.81), 혁신성(=.81), 그리고 상호작용품질(=.81)을 포함하는 9개 평가지표의 CVI가 최소만족 기준을 충족시키는 것으로 나타났다. 따라서 이들 9개의 평가지표를 최종분석에서 사용하였다.

### 3-3. 설문도구 설계

문헌 고찰과 델파이를 거쳐 도출된 9개의 ChatGPT UX 평가지표의 값을 측정하기 위하여 각 지표의 설문

19) Jacobs, C., Foote, G., & Williams, M. 'Evaluating user experience with immersive technology in simulation-based education: A modified Delphi study with qualitative analysis.' PLoS ONE, 2023, 08, Vol.18, No.8, pp.e0275766-e0275766

20) 이한진, 구현희 '메타버스 플랫폼 사용성 평가체계 구축에 관한 델파이연구: 로블록스, 제페토, 게더타운 사례를 중심으로.' 한국콘텐츠학회논문지, 2022, Vol.22, No.9, pp.179-193

문항을 설계하였다(표 2). 모든 문항들은 5점 리커트 척도(예., 1=전혀 그렇지 않다, 5=매우 그렇다)를 기반으로 설계되었다.

‘성과기대에 관한 본 연구에서의 조작적 정의는 ChatGPT와의 상호작용을 통해 사용자가 주어진 목표를 달성할 것이라는 기대 수준이다. 그리고 ‘노력기대’에 대한 정의는 ChatGPT를 사용하는 과정에서 요구되는 인지적 및 물리적 노력의 정도이다. 이들 2개의 지표 값을 측정하기 위하여 김현진(2020)<sup>21)</sup>이 설계한 각 4개의 문항을 본 연구의 목적과 맥락에 맞게 수정하였다. ‘유용성’에 관한 조작적 정의는 ChatGPT가 사용자의 필요를 충족시키고, 목표를 달성하거나 문제를 해결하는 데 효과적으로 기여하는지에 대한 체험 수준이다. Davis(1989)<sup>22)</sup>와 Wang(2024)<sup>23)</sup>이 사용한 6개의 문항을 기반으로 유용성에 관한 문항을 설계하였다. 본 연구는 ‘능력’에 대한 조작적 정의로서 ChatGPT와의 상호작용을 통하여 사용자의 질문이나 요청에 대해 적절한 응답을 제공할 수 있는 역량으로 설정하였다. Chiu et al.(2021)<sup>24)</sup>의 연구에서 사용된 11개의 문항을 본 연구의 맥락에 맞게 수정하여 능력의 수준을 측정하였다. ‘커뮤니케이션’은 ChatGPT와 사용자가 상호작용하는 과정에서 정보가 얼마나 명확하고 효과적으로 전달되는지에 대한 사용자 인식이다. 그리고 ‘공동창작’은 ChatGPT와 사용자가 협력하여 새로운 아이디어, 콘텐츠, 또는 해결책을 생성하는 과정에서 사용자가 느끼는 상호작용적 기여와 창의적 시너지의 정도를 의미한다. 이들 2개의 UX 평가지표 수준을 측정하기 위하여 Gao & Huang(2019)<sup>25)</sup>의 연구를

기반으로 각각 5개 및 3개의 문항을 설계하였다. 이와 함께, ‘효과성’에 대한 조작적 정의는 ChatGPT가 사용자 외의 상호작용에서 의도한 목표나 요청을 정확하고 신속하게 달성하는 능력에 대한 인식으로 설정하였다. 그리고 Kashive et al.(2020)<sup>26)</sup>이 사용한 설문도구를 기반으로 4개의 문항을 설계하였다. 이와 함께, ‘혁신성’은 ChatGPT가 기존의 도구나 서비스와 비교하여 얼마나 독창적이고 새로운 가치를 제공하며, 사용자에게 참신하고 차별화된 경험을 제공하는지에 대한 사용자 인식으로 정의하였다. 혁신성에 관한 10개의 설문 문항은 Al-Kfairy et al.(2024)<sup>27)</sup>의 연구를 기반으로 설계하였다. 상호작용 품질은 ChatGPT와의 대화 과정에서 사용자가 경험하는 응답의 정확성, 일관성, 그리고 대화의 자연스러움에 관한 전반적인 수준으로 조작적 정의를 설정하였다. Pelau et al.(2021)<sup>28)</sup>가 사용한 9개의 문항을 바탕으로 본 연구의 맥락을 반영하여 설계하였다.

**[표 2] ChatGPT 사용자 경험 평가지표와 설문도구 특징**

평가지표	조작적 정의	문항 수	연구 문헌
성과기대	ChatGPT와의 상호작용을 통해 사용자가 주어진 목표를 달성할 것이라는 기대 수준	4	김현진(2020);
노력기대	ChatGPT를 사용하는 과정에서 요구되는 인지적 및 물리적 노력의 정도	4	김현진(2020);
유용성	ChatGPT가 사용자의 필요를	6	Davis(1989);

21) 김현진. Untact vs. Contact: 럭셔리 브랜드 소비자의 대화형 에이전트 이용 의도에 관한 연구. 한양대학교 박사학위논문, 2024

22) Davis, F. D. ‘Perceived usefulness, perceived ease of use, and user acceptance of information technology.’ MIS Quarterly, 1989. 09, Vol.13, No.3, pp.319-340

23) Wang, J. 정보 특성과 개인적 특성이 생성형 AI 서비스 사용의도에 미치는 영향에 관한 실증적 연구: ChatGPT 서비스를 중심으로. 세종대학교 박사학위논문, 2024

24) Chiu, Y. T., Zhu, Y. Q., Corbett, J. ‘In the hearts and minds of employees: A model of pre-adoptive appraisal toward artificial intelligence in organizations.’ International Journal of Information Management, 2021. 10, Vol.60, p.102379

25) Gao, B. & Huang, L. ‘Understanding interactive user behavior in smart media content service: An

integration of TAM and smart service belief factors.’ Heliyon, 2019. 12, Vol.5, No.12, p.e02983

26) Kashive, N., Powale, L., & Kashive, K. ‘Understanding user perception toward artificial intelligence (AI) enabled e-learning.’ International Journal of Information and Learning Technology. 2021. 01, Vol.38, No.1, pp.1-19

27) Al-Kfairy, M., Mustafa, D., Al-Adaileh, A., Zriqat, S., & Sendaba, O. ‘User acceptance of AI voice assistants in Jordan’s telecom industry.’ Computers in Human Behavior Reports, 2024. 12, Vol.16, p.10

28) Pelau, C., Dabija, D-C., & Irina, E. ‘What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry.’ Computers in Human Behavior, 2021. 09, Vol.122, p.106855

	충족시키고, 목표를 달성하거나 문제를 해결하는 데 효과적으로 기여하는지에 대한 체험 수준		Wang(2024)
능력	ChatGPT와의 상호작용을 통하여 사용자의 질문이나 요청에 대해 적절한 응답을 제공할 수 있는 역량	11	Chiu et al.(2021)
커뮤니케이션	ChatGPT와 사용자가 상호작용하는 과정에서 정보가 얼마나 명확하고 효과적으로 전달되는지에 대한 사용자 인식	5	Gao & Huang(2019)
공동창작	ChatGPT와 사용자가 협력하여 새로운 아이디어, 콘텐츠, 또는 해결책을 생성하는 과정에서 사용자가 느끼는 상호작용적 기여와 창의적 시너지의 정도	3	Gao & Huang(2019)
효과성	ChatGPT가 사용자와의 상호작용에서 의도한 목표나 요청을 정확하고 신속하게 달성하는 능력에 대한 인식	4	Kashive et al.(2020)
혁신성	ChatGPT가 기존의 도구나 서비스와 비교하여 얼마나 독창적이고 새로운 가치를 제공하며, 사용자에게 참신하고 차별화된 경험을 제공하는지에 대한 사용자 인식	10	AL-Kfairy et al.(2024)
상호작용 품질	ChatGPT와의 대화 과정에서 사용자가 경험하는 응답의 정확성, 일관성, 대화의 자연스러움에 관한 전반적인 수준	9	Pelau et al.(2021)

### 3-4. 조사 대상자 및 자료의 수집

본 연구는 설문조사를 수행하여 ChatGPT UX의 평가지표 값을 수집하였다. 조사 대상자의 성별과 연령, 최종학력 수준 등에 관한 인구통계적 문항과 함께 56개의 ChatGPT UX 평가지표 문항이 포함된 설문도구를 온라인 설문지 폼(Naver Form)으로 설계하여 조사 대상자에게 배포하였다. 설문조사는 2024년 7월 30일부터 8월 15일까지 수행하였다. 3년 이상의 디자인 관련 실무경력을 가지는 212명이 설문조사에 참여하였다. 조사 대상자의 평균 연령은 35.1세(표준편차 =11.73)이다. 그리고 대상자 중에서 남성과 여성이 차지하는 비율은 각각 47.2%와 52.8%이다. 전체 대상자 중에서 69.48%는 4년제 대학교 졸업 이상의 최종학력 수준을 가진다고 응답하였다.

### 3-5. 자료의 분석 방법

본 연구가 수집한 설문자료의 통계적 분석을 수행하기 위하여 IBM SPSS 26.0과 IBM AMOS 26.0을 활용하였다. 통계적 유의수준은  $p=.05$ 로 설정하였다. 구체적으로, 동일한 평가지표가 포함하는 설문문항 값의 내적 일관성(internal consistency)을 평가하기 위하여 Cronbach's alpha 값을 계산하였다. 이와 함께, 측정 변수의 잠재변수(latent variable) 구조를 확인하기 위하여 탐색적 요인분석을 수행하였다. 측정모형의 구성 타당도와 이해타당도를 평가하기 위하여 확인적 요인분석을 수행하였다.

## 4. 연구 결과

### 4-1. 기술통계량

[표 3]은 56개의 설문문항 값에 대하여 계산된 평균과 표준편차, 왜도, 그리고 첨도 값을 나타내고 있다. 최대우도 추정법(maximum likelihood estimation) 기반의 AMOS를 활용하여 확인적 요인분석을 수행하기 위해서는 수집된 자료가 정규성(normality)의 특징을 가져야 한다.<sup>29)</sup> 이를 위하여 본 연구는 Kline(2011)<sup>30)</sup>이 제시한 최소만족 기준을 충족시키는지 확인하였는데, 이는 왜도와 첨도의 절대값이 각각 3.0 및 10.0 이하이다. 계산된 왜도와 첨도의 절대값을 확인한 결과, 왜도의 절대값은 각각 .292-1.989와 .798-4.286의 범위를 가지는 것으로 나타났다. 이들은 Kline(2011)의 가이드라인을 충족시키며, 따라서 본 연구가 수집한 설문문항 값은 정규성을 위배하지 않는 것으로 판단하였다.

[표 3] 기술통계량: 설문문항 값의 평균, 표준편차, 왜도, 첨도

설문문항	평균	표준편차	왜도	첨도
성과기대1	3.49	.926	.895	-.798
성과기대2	3.38	.923	.894	-.236

29) Li, C-H. 'Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables.' Behavior Research Methods, 2021. 10, Vol.53, No.5, pp.2191-2213

30) Kline, R. B. 『Principles and practice of structural equation modeling.』, 3rd ed, New York: Guilford Press, 2011

성과기대3	3.47	1.004	.573	-.761
성과기대4	3.51	.976	.622	-.751
노력기대1	3.11	.812	.913	2.070
노력기대2	3.04	.704	1.174	3.727
노력기대3	3.08	.687	1.400	3.926
노력기대4	3.13	.771	1.153	2.415
유용성1	3.23	.808	1.134	1.639
유용성2	3.27	.821	1.116	1.339
유용성3	3.38	.859	1.133	.192
유용성4	3.37	.885	1.028	.044
유용성5	3.34	.881	.986	.313
유용성6	3.25	.827	1.221	.984
능력1	3.11	.724	1.348	3.123
능력2	3.13	.647	1.989	4.286
능력3	3.13	.742	1.326	2.732
능력4	3.15	.730	1.389	2.936
능력5	3.23	.699	1.925	2.696
능력6	3.08	.759	1.111	2.616
능력7	3.18	.726	1.513	2.604
능력8	3.14	.784	1.293	1.864
능력9	3.20	.798	1.371	1.332
능력10	3.17	.764	1.447	1.868
능력11	3.17	.754	1.384	2.167
커뮤니케이션1	3.10	.796	1.007	2.193
커뮤니케이션2	3.15	.824	1.048	1.600
커뮤니케이션3	3.15	.798	1.028	2.116
커뮤니케이션4	2.98	1.009	.467	.548
커뮤니케이션5	3.08	.851	.787	1.627
공동창작1	3.16	.892	.700	1.208
공동창작2	3.16	.833	1.030	1.458
공동창작3	3.24	.851	1.009	1.073
효과성1	3.16	.859	.915	1.270
효과성2	3.25	.824	1.159	1.216
효과성3	3.17	.820	.926	1.904
효과성4	3.22	.826	1.002	1.587
혁신성1	3.11	.940	.680	.730
혁신성2	3.27	.914	.898	.254
혁신성3	3.40	.961	.687	-.263
혁신성4	3.43	1.003	.525	-.471
혁신성5	3.29	.896	.836	.622
혁신성6	3.25	.924	.780	.430
혁신성7	3.07	.852	.802	1.650
혁신성8	3.34	.977	.596	-.042
혁신성9	3.30	.866	.975	.754

혁신성10	3.13	.893	.865	.929
상호작용품질1	3.11	.679	1.416	3.896
상호작용품질2	3.18	.741	1.385	2.598
상호작용품질3	3.09	.706	1.419	3.152
상호작용품질4	3.06	.798	.965	2.151
상호작용품질5	3.06	.767	.912	2.770
상호작용품질6	2.86	.895	.473	1.430
상호작용품질7	3.01	.860	.614	1.747
상호작용품질8	2.68	.849	.292	1.511
상호작용품질9	3.20	.814	1.007	1.774

#### 4-2. 내적 일관성 평가 결과

동일한 ChatGPT UX 평가지표가 포함하는 다수의 설문문항이 동일한 개념 또는 속성을 일관되게 측정하였는지의 여부를 확인하기 위하여 Cronbach's alpha 계수 값을 계산하였다. [표 4]가 나타내는 바와 같이, 9개의 평가지표에 대해 계산된 Cronbach's alpha 계수 값은 최소값 .713과 최대값 .920의 범위를 가지며, 이는 최소만족 기준인 .7 이하를 충족시킨다.

[표 4] Cronbach's alpha 계산 결과

평가지표	문항 수	Cronbach's alpha
성과기대	4	.787
노력기대	4	.865
유용성	6	.802
능력	11	.917
커뮤니케이션	5	.876
공동창작	3	.920
효과성	4	.713
혁신성	10	.887
상호작용 품질	9	.866

#### 4-3. 탐색적 요인분석 수행 결과

ChatGPT UX에 관한 56개 측정변수가 가지는 잠재 요인 구조를 평가하기 위하여 탐색적 요인분석을 수행하였다. 요인분석의 적합성을 확인하기 위하여 비틀렛의 구형성 검정(Bartlett's test of sphericity)과 KMO 측도 값을 계산한 결과(Hoelzle & Meyer, 2013), 구형성( $\chi^2=7873.359(p=0.000<0.05)$ ), KMO 측도 =0.901로 나타났다. 이는 탐색적 요인분석의 수행을 위한 최소만족 기준을 충족시킨다.<sup>31)</sup>

31) Hair, Joseph F. 『Multivariate data analysis』 Upper Saddle River, N.J.: Pearson Prentice Hall, 2005

베리맥스(Varimax)와 주성분 분석(principal component analysis)을 각각 요인회전과 요인추출 방법으로 사용하여 탐색적 요인분석을 수행한 결과, 15 번의 반복 요인회전에 의해 총 13개의 성분이 추출되었다. 이 중에서 7개의 성분이 1.0 이상의 고유값(eigenvalue)을 가지는 것으로 나타났다(그림 1).

측정변수	1	2	3	4	5	6	7
성과기대1	.245	.154	.597	.037	.156	-.168	.202
성과기대2	.180	.091	.684	.057	.222	-.119	.044
성과기대3	.088	.223	.657	.217	.236	.055	.052
성과기대4	.029	.148	.741	.033	.169	.069	.098
노력기대1	.291	.095	.231	.092	.714	.182	.124
노력기대2	.226	.093	.125	.103	.754	.005	.027
노력기대3	.180	.176	.230	.048	.789	.111	.105
노력기대4	.183	.043	.186	.244	.722	-.050	.138
유용성1	.130	.107	.115	.208	.330	.031	.291
유용성2	.169	.289	.242	.133	.384	.055	.328
유용성3	.176	.341	.456	.017	.046	.062	.106
유용성4	.150	.257	.470	.207	.182	.066	.166
유용성5	.113	.194	.411	.109	.158	.049	.126
유용성6	.372	.150	.294	.233	.085	-.019	.039
능력1	.525	-.050	.202	.194	.160	.235	.073
능력2	.740	.059	.019	.129	.214	-.027	.100
능력3	.667	.010	.120	.083	.031	.138	.309
능력4	.661	.234	.061	.113	.099	.145	.250
능력5	.685	.179	.117	.117	.188	.116	.175
능력6	.720	.157	.191	.157	.252	.209	.119
능력7	.442	.191	.177	-.148	.247	.467	.084
능력8	.692	.115	.059	.188	.112	.142	.015
능력9	.558	.115	.112	.163	.223	.271	.131
능력10	.609	.136	.104	.422	.155	.004	.062
능력11	.599	.094	.184	.144	.045	.044	.197
커뮤니케이션1	.176	.253	.186	.600	.090	.110	.155
커뮤니케이션2	.232	.185	.116	.733	.105	.101	.251
커뮤니케이션3	.303	.200	.169	.585	.241	.058	.214
커뮤니케이션4	.219	.144	.082	.705	.122	.393	.009
커뮤니케이션5	.244	.084	.114	.714	.136	.230	.169
공동창작1	.232	.138	.114	.154	.092	.226	.812
공동창작2	.249	.175	.198	.155	.166	.130	.798
공동창작3	.195	.144	.090	.164	.113	.025	.815
효과성1	.170	.076	.111	.159	.240	.178	.155
효과성2	.334	.170	.279	.091	.195	.157	.331
효과성3	.160	.211	.298	.255	.032	.161	.254
효과성4	.082	.319	.279	.217	.062	.315	.017
혁신성1	.092	.296	-.058	.325	.236	.364	.060
혁신성2	.127	.612	.233	.150	.197	.071	.186
혁신성3	.084	.736	.211	.044	.118	.085	.101
혁신성4	.194	.568	.137	.119	.111	.110	.177
혁신성5	.168	.650	.249	-.183	-.017	.091	.125
혁신성6	.163	.503	.153	.150	.057	.099	-.026
혁신성7	.103	.568	.206	.040	.026	.436	.152
혁신성8	.053	.788	.058	.153	.157	.112	.012
혁신성9	.109	.432	.471	.170	.056	.201	.051
혁신성10	.155	.467	.144	.102	-.028	.336	.191
상호작용품질1	.152	.328	.034	.153	.010	.236	.125
상호작용품질2	.191	.200	.059	.050	.061	.130	.053
상호작용품질3	.084	.188	.148	.109	.096	.383	.155
상호작용품질4	.202	.161	.146	.154	.123	.528	.096
상호작용품질5	.200	.328	-.100	.125	.138	.503	.127
상호작용품질6	.204	.116	.109	.289	-.026	.689	.098
상호작용품질7	.229	.319	.021	.213	.106	.575	.228
상호작용품질8	.099	.070	-.037	.226	.018	.657	.064
상호작용품질9	.222	.105	.058	.036	.303	.110	.120
고유값	19.372	3.314	3.027	1.824	1.770	1.612	1.390
설명된 분산(%)	34.592	5.919	5.406	3.258	3.160	2.879	2.482
누적 분산(%)	34.592	40.511	45.917	49.175	52.334	55.213	57.695

[그림 1] 탐색적 요인분석 수행 결과: 회전된 성분행렬

요인적재량의 선정 기준을 .5 이상으로 설정하여 (Knehta et al., 2019) 요인분석으로부터 얻어진 결과를 관찰한 결과, ‘유용성’과 ‘효과성’에 관한 모든 측정 변수의 요인적재량은 최소만족 기준을 충족시키지 않는

것으로 나타났다. 따라서 이들 2개의 UX 지표를 최종 분석에서 제외하였다. 이와 함께, 8개의 측정변수인 ‘능력7’, ‘혁신성1’, ‘혁신성 9’, ‘혁신성 10’, ‘상호작용품질1’, ‘상호작용품질2’, ‘상호작용품질3’, 그리고 ‘상호작용품질9’에 대해 계산된 요인 적재량은 최소만족 기준인 .5 이상을 충족시키지 않는 것으로 나타났다. 따라서 이들을 제외한 38개의 측정변수와 값을 최종분석에 사용하였다.

#### 4-4. 확인적 요인분석 수행 결과

ChatGPT UX의 측정모형에 관한 구성타당도를 평가하기 위하여 확인적 요인분석을 수행하였다. 이를 위하여 본 연구는 측정모형의 수렴타당도와 판별타당도를 확인하였다.

측정모형의 수렴타당도를 확인하기 위하여 3가지 기준을 만족시키는지 확인하였다. 이는 (1) 측정변수에 대해 계산된 표준화된 요인적재량이 .5 이상(Wong, 2013), (2) 평균분산추출(average variance extracted; AVE) 값이 .5 이상(Fornell & Larcker, 1981), 그리고 (3) 복합신뢰도(composite reliability) 값이 .7 이상<sup>32)</sup>이다. 확인적 요인분석을 수행한 결과(표 5), 측정변수에 대해 계산된 표준화 요인 적재량은 .612-.956의 범위를 가지는 것으로 나타났다. 그리고 각 평가지표에 대해 계산된 AVE와 CR 값은 각각 .484-.798 및 .789-.922의 범위를 가진다. 2개의 평가지표인 성과기대와 혁신성에 대한 AVE 값은 각각 .484와 .489이며, 이들 값은 최소만족 기준을 엄격하게 충족시키지 않으나, 크게 위배하지 않는다고 판단하였다. 따라서 ChatGPT UX 측정모형의 수렴타당도의 수준은 적절한 것으로 판단하였다.

[표 5] 확인적 요인분석 수행 결과: 표준화 요인 적재량, AVE, CR

구성개념	측정변수	표준화 요인 적재량	AVE	CR
성과기대	성과기대1	0.625	.484	.789
	성과기대2	0.721		
	성과기대3	0.732		
	성과기대4	0.700		
노력기대	노력기대1	0.755	.626	.870
	노력기대2	0.779		
	노력기대3	0.848		

32) Hair, Joseph F. 『Multivariate data analysis』 Upper Saddle River, N.J.: Pearson Prentice Hall, 2005

	노력기대4	0.779		
능력	능력1	0.614	.523	.916
	능력2	0.704		
	능력3	0.669		
	능력4	0.731		
	능력5	0.754		
	능력6	0.802		
	능력8	0.709		
	능력9	0.732		
	능력10	0.764		
	능력11	0.735		
	커뮤니케이션	커뮤니케이션1		
커뮤니케이션2		0.800		
커뮤니케이션3		0.741		
커뮤니케이션4		0.794		
커뮤니케이션5		0.781		
공동 창작	공동창작1	0.905	.798	.922
	공동창작2	0.956		
	공동창작3	0.813		
혁신성	혁신성2	0.717	.489	.879
	혁신성3	0.731		
	혁신성4	0.666		
	혁신성5	0.774		
	혁신성6	0.612		
	혁신성7	0.683		
	혁신성8	0.701		
	상호작용 품질	상호작용품질4		
상호작용품질5		0.720		
상호작용품질6		0.748		
상호작용품질7		0.711		
상호작용품질8		0.651		

측정모형의 판별타당도를 확인하기 위하여 각 평가지표 간의 HTMT(Heterotrait-Monotrait) 값을 계산한 결과(Henseler et al., 2016), 최소값 .314와 최대값 .700의 범위를 가지는 것으로 나타났다. 이는 최소 만족 기준인 .9 이하를 충족시킨다.

이와 함께, 측정모형의 모형적합도를 평가하기 위하여 정규화 카이제곱(normed chi-square)과 표준화적합지수(normed fit index; NFI), 비교적합지수(comparative fit index; CFI), 증분적합지수(incremental fit index; IFI), 그리고 근사오차평균자승의 이중근(root mean square error of approximation; RMSEA)의 값을 계산하였다.([표 6]).

[표 6] 측정모형의 모형적합도 계산 결과

모형적합도 지수	$\chi^2(df)$	NFI	CFI	IFI	RMSEA
모형적합도	1.867	.879	.882	.883	.064
최소 만족기준	$\leq 3.0$	$\geq 0.8$	$\geq 0.8$	$\geq 0.9$	$\leq 0.08$

IFI를 제외한 모든 지수 값은 최소만족 기준을 충족시키는 것으로 나타났다. 그리고 계산된 IFI 값은 최소 만족 기준을 충족시키지 않으나 기준을 크게 위반하지 않는다고 본 연구는 판단하였다. 따라서 ChatGPT UX의 측정모형은 양호한 수준의 모형적합도를 가지는 것으로 나타났다.

2차 구성개념인 ChatGPT UX가 7개의 1차 구성개념인 성과기대, 노력기대, 능력, 커뮤니케이션, 공동 창작, 혁신성, 그리고 상호작용 품질에 의해서 이루어지는지를 확인하기 위하여 이해타당도를 평가하였다.<sup>33)</sup>

[표 7]이 나타내는 바와 같이, 7개의 1차 구성개념에 대해 계산된 표준화 요인 적재량은 최소값 .654와 최대값 .833의 범위를 가진다. 그리고 2차 구성개념인 ChatGPT UX에 대해 계산된 AVE와 CR 값은 각각 .546과 .893이며, 이는 최소만족 기준인 .5 이상 및 .7 이상을 충족시키는 것으로 나타났다.<sup>34)</sup> 따라서, 본 연구에서의 7개 UX 평가지표는 ChatGPT UX를 통계적 유의수준에서 타당하게 구성하는 것으로 판단하였다.

[표 7] 확인적 요인분석 수행 결과: 1차 구성개념과 2차 구성개념 간의 이해타당도

2차 구성개념	1차 구성개념	표준화 요인 적재량	AVE	CR
ChatGPT UX	성과기대	.654	.546	.893
	노력기대	.662		
	능력	.819		
	커뮤니케이션	.833		
	공동 창작	.687		
	혁신성	.739		
	상호작용 품질	.758		

## 5. 결론

AI 기술의 발전은 전 세계적으로 사용자 경험(UX) 디자인과 평가의 새로운 패러다임을 만들어내고 있다. 특히 ChatGPT와 같은 대화형 AI 서비스는 사용자와의

33) Bollen, Kenneth A. 『Structural equations with latent variables.』 New York: John Wiley & Sons, Inc. 1989

34) Hair, Joseph F. 『Multivariate data analysis.』 Upper Saddle River, N.J.: Pearson Prentice Hall, 2005

상호작용을 통해 기존 서비스와 차별화된 경험을 제공하며, 이를 효과적으로 평가할 수 있는 UX 평가지표의 필요성이 대두되고 있다. 본 연구는 이러한 배경에서 ChatGPT의 사용자 경험을 평가하기에 적합한 UX 평가지표를 개발하고 타당성을 검증하는 데 목적이 있다.

이를 위해 선행연구로서 문헌고찰을 진행하였는데, ChatGPT의 UX 평가지표 관련 문헌 개수가 매우 제한적이기 때문에 AI서비스의 UX평가지표로 범위를 확장하고 문헌을 탐색하였다. 문헌을 통해 총 36개의 초기 UX 평가지표를 도출하였으며, 이는 서비스를 사용하는 과정에서 경험할 수 있는 유용성, 용이성, 개인화 등과 더불어 서비스를 사용하기 전과 후에 경험할 수 있는 사회적 영향, 성과기대, 공동창작, 혁신성 등 까지 포함한다. 이후 디자인 전문가 16명이 참여한 델파이를 통해 내용타당도에 따라 총 9개의 ChatGPT UX평가지표를 선정하였고, 이후 212명의 설문조사에 대해 탐색적 요인분석을 수행한 결과, 최종적으로 총 7개의 ChatGPT UX 평가지표를 도출하였다.

최종 도출된 ChatGPT UX 평가지표는 성과기대, 노력기대, 능력, 커뮤니케이션, 공동 창작, 혁신성, 상호작용 품질이다. 이 7개의 평가지표에 따르면, ChatGPT는 사용자의 '성과기대와 '노력기대에 부응하며 '효율성을 극대화해야 한다. 또한, 사용자의 '능력을 강화하고 '커뮤니케이션을 원활하게 지원하며, '공동창작의 가능성을 확대하는 등 보다 인간 중심적인 설계가 요구된다. 또한 '상호작용 품질'과 '혁신성'을 중요한 평가 요소로 다루어 진다는 점에서 ChatGPT의 UX 설계 시 다른 디지털 서비스와 차별화 되는 주요 가치로 인식하고 개발될 필요가 있다.

흥미롭게도 본 연구에서는 일반적인 UX 평가지표에서 가장 보편적으로 쓰이는 유용성과 효과성이 탐색적 요인분석으로부터 제거되어 최종 지표로 채택되지 않았다. 이는 ChatGPT와 같은 AI 서비스의 맥락에서 유용성과 효과성의 정의가 불명확하고, 다른 지표와 의미가 희석되어 독립적인 평가 항목으로서의 타당성이 약한 것으로 판단된다.

본 연구의 주요 의의는 ChatGPT와 같은 AI 특성을 반영한 UX 평가지표를 새롭게 도출했다는 점에 있다. 이는 향후 ChatGPT뿐 아니라 다양한 AI 기반 서비스의 UX 평가와 디자인을 위한 가이드라인으로 활용될 가능성을 열어줄 것으로 기대된다.

## 참고문헌

1. 안무정, 강태임. '디지털 트랜스포메이션 경영을 위한 챗GPT 사용자 경험(UX) 디자인 평가: 오픈 AI 챗GPT와 마이크로소프트 Bing 챗GPT 교차활용을 중심으로'. 한국디자인문화학회지, 2023
2. 이한진, 구현희 '메타버스 플랫폼 사용성 평가체계 구축에 관한 델파이연구: 로블록스, 제페토, 게더타운 사례를 중심으로.' 한국콘텐츠학회논문지, 2022
3. Abbas, A., Imran, K., & Ting, C-Y. 'User experience design using machine learning: A systematic review.' IEEE Access, 2022
4. Al-Kfairy, M., Mustafa, D., Al-Adaileh, A., Zriqat, S., & Sendaba, O. 'User acceptance of AI voice assistants in Jordan's telecom industry.' Computers in Human Behavior Reports, 2024
5. Andrade-Arenas, L. & Yactayo-Arias, C. 'Chatbot with ChatGPT technology for mental wellbeing and emotional management.' IAES International Journal of Artificial Intelligence, 2024
6. Ashfaq, Muhammad, et al. 'I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents.' Telematics and Informatics, 2020
7. Bozkurt, Aras. 'Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift.' Asian Journal of Distance Education. 2023
8. Chiu, Y. T., Zhu, Y. Q., Corbett, J. 'In the hearts and minds of employees: A model of pre-adoptive appraisal toward artificial intelligence in organizations.' International Journal of Information Management, 2021
9. Davis, F. D. 'Perceived usefulness, perceived ease of use, and user acceptance of information technology.' MIS Quarterly, 1989

10. Gao, B. & Huang, L. 'Understanding interactive user behavior in smart media content service: An integration of TAM and smart service belief factors.' *Heliyon*, 2019
11. Gupta, B., Mufti, T., Sohail, S. S., & Madsen, D. Ø. 'ChatGPT: A brief narrative review. *Cogent Business & Management*' 2023
12. HO, Brittany, et al. 'ChatReview: A ChatGPT-enabled natural language processing framework to study domain-specific user reviews.' *Machine Learning with Applications*, 2024
13. Hosseini, M., & Horbach, S. P. 'Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review.' *Research Integrity and Peer Review*, 2023
14. Jacobs, C., Foote, G., & Williams, M. 'Evaluating user experience with immersive technology in simulation-based education: A modified Delphi study with qualitative analysis.' *PLoS ONE*, 2023
15. Kashive, N., Powale, L., & Kashive, K. 'Understanding user perception toward artificial intelligence (AI) enabled e-learning.' *International Journal of Information and Learning Technology*. 2021
16. Kim, Jeong Soo, Minseong Kim, and Tae Hyun Baek. "Enhancing User Experience With a Generative AI Chatbot." *International Journal of Human-Computer Interaction*, 2024
17. Li, C-H. 'Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables.' *Behavior Research Methods*, 2021
18. Ma, X. & Huo, Y. 'Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework.' *Technology in Society*, 2023
19. Manzoor, S. R., Ullah, R., Khattak, A., Ullah, M., & Han, H. 'Exploring tourist perceptions of artificial intelligence devices in the hotel industry: Impact of industry 4.0.' *Journal of Travel & Tourism Marketing*, 2024
20. Meidute-Kavaliauskiene, I., Çiğdem, S., Yıldız, B., & Davidavicius, S. 'The effect of perceptions on service robot usage intention: A survey study in the service sector'. *Sustainability*, 2021
21. Meyer-Waarden, L. & Cloarec, J. "Baby, you can drive my car": Psychological antecedents that drive consumers' adoption of AI-powered autonomous vehicles.' *Technovation*, 2022
22. Mohr, S. & Kühl, R. 'Acceptance of artificial intelligence in German agriculture: an application of the technology acceptance model and the theory of planned behavior.' *Precision Agriculture*, 2021
23. Niu, Ben, and Gustave Florentin Nkoulou Mvondo. 'I Am ChatGPT, the ultimate AI Chatbot! Investigating the determinants of users' loyalty and ethical usage concerns of ChatGPT.' *Journal of Retailing and Consumer Services*, 2024
24. Park, A. & Lee, S. B. 'Examining AI and systemic factors for improved chatbot sustainability.' *Journal of Computer Information Systems*, 2024
25. Pelau, C., Dabija, D-C., & Irina, E. 'What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry.' *Computers in Human Behavior*, 2021
26. Pillai, R., Sivathanu, B., & Dwivedi, Y. K. 'Shopping intention at AI-powered automated retail stores (AIPARS).' *Journal of Retailing and Consumer Services*, 2020

27. Prakash, A. V. & Das, S. 'Medical practitioner's adoption of intelligent clinical diagnostic decision support systems: A mixed-methods study.' *Information & Management*, 2021
28. Rahimi, F., & Abadi, A. T. B. 'ChatGPT and publication ethics'. *Archives of Medical Research*. 2023
29. Ribeiro, M. A., Gursoy, D., & Chi, O. H. 'Customer acceptance of autonomous vehicles in travel and tourism.' *Journal of Travel Research*. 2022
30. Roy, P., Ramaprasad, B. S., Chakraborty, M., Prabhu, N., & Rao, S. 'Customer Acceptance of Use of Artificial Intelligence in Hospitality Services: An Indian Hospitality Sector Perspective.' *Global Business Review*, 2024
31. Sohn, K. & Kwon, O. 'Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products.' *Telematics and Informatics*, 2020
32. Song, S. Y., & Kim, Y.-K. 'Factors influencing consumers' intention to adopt fashion robot advisors: Psychological network analysis.' *Clothing and Textiles Research Journal*, 2022
33. Srinivasan, R. & Gonzalez, B. S. M. 'The role of empathy for artificial intelligence accountability.' *Journal of Responsible Technology*. 2022
34. Seyitoğlu, F. & Ivanov, S. 'Robots and emotional intelligence: A thematic analysis.' *Technology in Society*, 2024
35. Upadhyay, N., Upadhyay, S., & Dwivedi, Y. K. 'Theorizing artificial intelligence acceptance and digital entrepreneurship model.' *International Journal of Entrepreneurial Behavior & Research*, 2022
36. Vu, Hong Tien ; Lim, Jeongsub, 'Effects of country and individual factors on public acceptance of artificial intelligence and robotics technologies: a multilevel SEM analysis of 28-country survey data. *Behaviour & Information Technology*, 2022
37. 김현진. Untact vs. Contact: 럭셔리 브랜드 소비자의 대화형 에이전트 이용 의도에 관한 연구. *한양대학교 박사학위논문*, 2024
38. Wang, J. 정보 특성과 개인적 특성이 생성형 AI 서비스 사용의도에 미치는 영향에 관한 실증적 연구: ChatGPT 서비스를 중심으로. *세종대학교 박사학위논문*. 2024
39. Bollen, Kenneth A. 『Structural equations with latent variables.』 New York: John Wiley & Sons, Inc. 1989
40. Hair, Joseph F. 『Multivariate data analysis.』 Upper Saddle River, N.J.: Pearson Prentice Hall, 2005
41. Kline, R. B. 『Principles and practice of structural equation modeling.』, 3rd ed, New York: Guilford Press, 2011
42. kjuve, Marita, Asbjørn Følstad, and Petter Bae Brandtzaeg. 'The user experience of ChatGPT: findings from a questionnaire study of early users.' *Proceedings of the 5th international conference on conversational user interfaces*. 2023
43. [www.economist.com](http://www.economist.com)
44. [www.nature.com](http://www.nature.com)

## 부록

45. 설문지 링크 : <https://naver.me/x3jyZkHX>