

초보 사용자와 숙련 사용자를 위한 이미지 생성형 AI의 설명가능성: 정적 및 동적 설명 비교를 중심으로

양 나 은¹ · 한 은 빈¹ · 최 준 호^{2*}

¹연세대학교 정보대학원 UX트랙 석사과정

²연세대학교 정보대학원 UX트랙 교수

Explainability of Image Generative AI for Novice and Expert Users: A Comparative Study of Static and Dynamic Explanations

Na-Eun Yang¹ · Eun-Bin Han¹ · Jun-Ho Choi^{2*}

¹Master's Course, UX Track, Graduate school of Information, Yonsei University, Seoul 03722, Korea

²Professor, UX Track, Graduate school of Information, Yonsei University, Seoul 03722, Korea

[요 약]

최근 생성형 AI 기술이 급속도로 발전하며 이미지 생성형 AI 도구의 활용 범위도 크게 확장되고 있다. 그러나 이미지 생성형 AI의 복잡성으로 초보 사용자(novice)가 활용 방식에 어려움을 느끼는 등 사용성 개선이 필요한 상황이다. 이 연구는 설명가능한 AI(XAI)의 접근법 중 하나인 상호작용 설명 방식이 경험의 수준이 다른 사용자들이 인지하는 설명가능성에 어떤 영향을 미치는지 조사하였다. 특히, 정적 설명과 동적 설명을 비교하여 설명가능성, 이해성, 정보성, 신뢰성, 만족도에 미치는 영향을 분석하였다. 연구 결과, 동적 설명이 정적 설명에 비해 설명가능성과 이해성을 향상시키며, 특히 초보 사용자에게 더 효과적임을 확인했다. 이를 통해 맞춤형 설명 전략의 중요성을 강조하며, 보다 사용자 친화적이고 접근 가능한 생성형 AI 시스템 디자인에 기여하고자 한다.

[Abstract]

As generative AI technology continues to advance, the utilization of image generative AI tools has significantly increased. However, the complexity of these systems presents challenges for both novice and expert users, highlighting a need for improvements in usability. This study explores the impact of interactive explanation methods, a branch of Explainable AI (XAI), on perceived explainability among users with varying experience levels. Specifically, it compares dynamic and static explanations, analyzing their effects on explainability, understandability, informativeness, trustworthiness, and user satisfaction. The findings indicate that dynamic explanations enhance both explainability and understandability, particularly for novice users. These results emphasize the importance of tailored explanation strategies in designing more user-friendly and accessible generative AI systems.

색인어 : 이미지 생성형 AI, 설명가능한 AI, 동적 설명, 상호작용 설명, 사용자 경험

Keyword : Image Generative AI, Explainable AI (XAI), Dynamic Explanation, Interactive Explanation, User Experience

<http://dx.doi.org/10.9728/dcs.2024.25.8.2261>



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 non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 06 July 2024; Revised 23 August 2024

Accepted 24 August 2024

*Corresponding Author, Jun-Ho Choi

E-mail: uxlab.junhochoi@gmail.com

I. 서론

AI(Artificial Intelligence) 기술이 발전함에 따라 생성형 AI(Generative AI)는 텍스트를 넘어 이미지, 음성, 영상 등 일상 속 활용범위가 다양해지고 있다. 그중 학습된 데이터를 바탕으로 이미지 생성 및 편집을 도와주는 이미지 생성형 AI에 관한 관심이 높아지며, 최근 구글, 메타 등 여러 기업들이 이미지 생성형 AI 시장에 뛰어들었다. 그러나 나날이 발전하는 이미지 생성 기술에 비해, 창작자의 사용자 경험 측면에 대한 논의는 아직 초기 단계이다. 현재의 생성형 AI는 뿌리 깊은 사용성 문제를 가지고 있다고 지적되며, 더 나은 사용성을 가지는 것이 경쟁력에서 상당한 우위를 점할 수 있을 것으로 예측된다[1].

이미지 생성형 AI 분야에서는 숙련된 사용자뿐 아니라 AI를 통한 이미지 생성 경험이 적거나 없는 사용자들의 탐색과 같이 생성 경험 여부에 따라 세분화된 사용자 경험에 관한 검토가 필요한 상황이다. 디자인 실무자를 대상으로 이미지 생성형 AI 서비스의 사용성을 조사한 결과, 사용성, 감정 유발, 생성된 결과 만족도에서 상당한 개선이 필요한 것으로 조사되었다[2]. 또한 Adobe에서 전 세계 크리에이터 전문가들을 대상으로 실시한 설문(2024)에 따르면, 83%가 업무에서 생성형 AI를 활용해본 경험이 있으며, 그중 가장 많이 활용된 분야는 이미지 제작으로 나타났다. 이들이 생성형 AI 도구를 선택할 때 중요하게 고려하는 요소로는 사용 편의성(35%)이 가장 높았다. 반면 사용해 보지 않은 가장 큰 이유로는 정확한 사용법을 모르는 것(31%)으로, 아직 AI 생성 경험의 진입 장벽이 존재함을 볼 수 있다[3].

설명가능한 AI(Explainable AI, XAI)는 생성형 AI가 대중적으로 활용됨에 따라 사용자 경험을 향상시킬 수 있는 해결책으로 주목받고 있다[4]. 다만 이미지 생성형 AI 사용 경험과 XAI를 분석한 선행 연구로는 대부분 XAI 필요성을 발굴하거나 프롬프트 작성에서의 어려움에만 집중하여 프롬프트 작성 시 필요한 사항이나 가이드라인을 제안하는 연구가 대부분이었다[5]. 자연어 처리(NLP, Natural Language Processing)와 대규모 언어 모델(LLM, Large Language Models)의 급속한 발전에 따라 대화를 통한 상호작용이 XAI의 새로운 접근법으로 제시되고 있으며, 사용자와 맥락에 따라 맞춤형 설명을 제공하는 정적(dynamic) 설명 방식은 초기 사용자의 진입장벽을 높이는 이미지 생성형 AI의 복잡성을 완화할 수 있을 것으로 기대되고 있다[10].

이 연구에서는 사용자의 이미지 생성형 AI 사용 경험의 여부에 따라 어떤 설명 제공 방식이 적합한지 파악하고자 했다. 특히 시스템과 상호작용하며 맞춤형 설명을 제공하는 동적 설명 접근법이 초보 사용자가 인지하는 설명가능성에 어떤 영향을 미치는지 알아보았다.

II. 이론적 배경

2-1 이미지 생성형 AI와 설명가능성

이미지 생성형 AI(Image Generative AI)는 텍스트 설명이나 기타 입력 데이터를 기반으로 새로운 시각적 콘텐츠를 생성하는 생성형 AI의 하위 분야이다. 달리2(DALL·E2), 미드저니(Midjourney) 등의 서비스를 통해 사용자는 간단한 텍스트 프롬프트로부터 매우 상세하고 맥락에 맞는 이미지를 생성할 수 있다[6],[7]. 창작자들은 자동화(automation), 아이디어 확장(exploration), 커뮤니케이션 중재(mediation) 등 다양한 창작 상황에서 생성형 AI 기술을 활용하고 있다[8]. 그러나, 급격한 기술 발전으로 인해 많은 시각 예술가들이 이미지 생성형 AI를 보다 적극적으로 활용하는 데 어려움을 겪고 있다[8]. 사용자가 고도화된 생성형 AI 기술을 활용하기 위해서는 시스템의 작동방식과 결과물의 생성 이유에 대해 이해하는 것이 필요하기 때문이다.

설명가능한 AI(Explainable AI, XAI)는 이러한 문제에 대응하여 인간이 AI 시스템의 결과를 더 이해하기 쉽게 만드는 것을 목표로 하는 연구 분야이다[9]. XAI는 인간이 AI의 출력 결과물이나 모델 전체를 이해할 수 있도록 도움을 주는 설명을 제공하는 것을 의미한다[30].

설명가능한 AI의 중요성은 생성형 AI 시대 이전과 이후로 나뉘며 최근 더욱 조명을 받고 있는데, 생성형 AI의 생성물과 시스템이 더 복잡해지면서 설명해야 할 방법과 내용의 선택지가 크게 늘어났기 때문이다[10]. 또한 생성형 AI가 주목을 받으며 XAI 연구의 방향성은 알고리즘 중심에서 인간 중심 XAI(human-centered XAI)으로 확대되었다. 초기에는 AI 알고리즘 원리를 이해하고 신뢰할 수 있는 설명 모델 개발에 관한 연구가 많이 이루어졌다[30]. 그러나 생성형 AI가 대중화됨에 따라 최근 인간 중심의 XAI에 관한 연구가 활발하게 이루어지고 있다. 이전의 설명가능한 AI는 개발자들이 모델의 기술적 이해를 돕는 개념 중심이었다면, 최근에는 최종적으로 시스템을 사용하게 되는 사용자(end user)의 맥락, 환경 등에 주목해 인터페이스 디자인과 심리적 분석을 다룬 연구를 볼 수 있다[41].

설명가능한 이미지 생성형 AI 연구 역시 주요 이슈로 떠오르고 있다. 참 혹은 거짓으로 결과가 출력되는 이진 분류 시스템에서는 설명해야 할 정보가 하나뿐이지만, 이미지를 생성하는 시스템의 경우 설명이 필요한 정보의 수가 수백만에 이른다는 특징이 있다[10]. 따라서 사용자가 출력물의 모든 가능한 세부 사항의 이유를 이해하는 것은 거의 불가능하다. 이는 사용자가 시스템과 상호작용하여 설명을 얻고 시스템이 사용자의 요구에 맞춘 적절한 설명을 제공해야 할 필요성을 시사한다.

이미지 생성형 AI에 관한 선행 연구에서도 인간 중심 XAI 연구를 다수 확인할 수 있다. 하지만 대부분의 연구는 텍스트

-이미지 모델에서 필요한 사용자의 어려움 분석[42], 텍스트 프롬프트 자동화와 같은 시스템 설계[5] 등 텍스트 프롬프트 작성에 관련되어 있다. 즉, 사용자에 대한 배경, 목표 세밀한 접근의 연구는 부족하다. 이 연구에서는 XAI 접근법을 기반으로 이미지 생성형 AI 사용자의 배경, 특히 사용 경험 여부에 따라 어떠한 설명방식이 적절한지 탐구하고자 했다.

2-2 정적 설명과 동적 설명

XAI의 설명 방식은 선행연구를 바탕으로 크게 다섯 가지로 분류할 수 있으며, 표 1에서 확인할 수 있다. 이러한 설명 방식은 상호보완적으로 한 시스템 내에서 활용될 수 있다. 생성형 AI의 발전에 따라 다양한 XAI 접근 방법 중에서도 상호작용 설명(Interactive explanation)의 중요성이 강조되고 있다. 상호작용 설명을 구현하는 대표적인 방식은 XAI를 대화로 재구성하는 것이다[11]. 한 연구에서는 사용자의 자연어 요청(query)을 기존에 정의된 작업 명령 중 하나로 변경하여 적절한 설명을 제공하고자 했다[12]. 이를 통해 사용자는 생성물에 가장 중요하게 영향을 미친 특징을 알아내거나 이를 변경할 수 있다. 또한, 대규모 언어 모델(Large Language Model, LLM)을 통해 상호작용이 가능한 XAI 추천 시스템이 제안되기도 했다[13].

표 1. 설명가능한 AI(XAI)의 설명 방식

Table 1. Explanation approach of Explainable AI (XAI)

Approach	Definition	Related works
Feature Importance and Attribution	Explain which features or inputs are most influential in the AI system's decision-making process.	[14],[15]
Visual Explanations	Use graphical representations to make AI decisions more understandable.	[16],[17]
Example-Based Explanations	Provide users with instances from the dataset that are similar to the input data or that have led to similar outputs.	[18],[19]
Rule-Based Explanations	Involve generating human-readable rules that approximate the behavior of the AI model.	[20],[21]
Interactive Explanations	engage users in a dialogue, allowing them to explore and understand the AI model's behavior dynamically.	[22],[23]

상호작용을 통한 XAI의 주요 개념으로 정적(static) 설명과 동적(dynamic) 설명이 있다. 정적 설명과 동적 설명은 AI 모델을 통해 만들어진 생성물의 어떤 속성(property)이 어떻게 설명되는지에 따라 구분된다[10]. 정적 설명에서는 모든 생성물에 대해 동일한 속성을 동일한 방식으로 설명한다. 예를 들어, 픽셀 속성 지도(Pixel Attribution Map)를 이용해 생성물을 설명할 경우, 어떤 이미지이든 동일하게 각 픽셀에

대한 중요도 값을 시각적으로 나타낸다. 반면, 동적 설명에서는 하나의 AI 모델에서 생성된 여러 생성물에 대해 서로 다른 속성을 서로 다른 방식으로 설명할 수 있다. 텍스트 자가 설명(Textual self-explanation)의 예시를 살펴보면, 생성된 이미지에 대해 ‘움직이는 빨간 자동차가 이미지에 생동감을 주고 시청자를 끌어들이기 위해 만들어졌습니다’라는 설명은 이미지 내 한 부분의 특정한 속성에 대해 설명하고 있다[10]. 생성된 이미지에서 설명되는 부분, 속성, 방법이 생성물이나 사용자 등 다양한 요소에 따라 달라진다는 점에서 이러한 설명 방식은 동적(dynamic)이라고 할 수 있다. 동적 설명의 목표는 시스템이 적절한 설명을 실시간으로 선별하여 제공하거나 사용자가 설명을 자유롭게 선택하면서도 사용자의 목적 달성을 극대화하는 것이다[10].

생성형 AI는 텍스트나 이미지를 생성하는 과정에서 톤(tone), 스타일, 의미론 등 다양한 측면에 대해 수많은 결정을 내리기 때문에 사용자는 생성 결과가 특정 속성을 나타내는 이유에 대해 수많은 질문이 가능하다[24]. 따라서 동적 설명은 복잡한 생성형 AI 모델의 설명가능성을 향상시키는 데 적합하다. 이 연구에서는 이미지 생성형 AI에서의 동적 설명이 시스템의 설명가능성에 긍정적인 영향을 미치는지 살펴보고자 한다.

2-3 사용자 숙련도와 설명가능한 AI

설명가능한 AI를 시스템에 적용하는 데 있어 사용자를 고려하는 것은 중요하다. 최종 사용자와 사용 맥락을 이해하지 않고는 XAI 방법을 개발할 수 없다는 인식이 커짐에 따라 [25], 연구자들은 사람들의 역할[26]-[28], 전문 지식[19], 보다 세분화된 지식 및 사용 목적[29]을 기반으로 XAI 필요성을 특징짓기 위한 개념적 프레임워크를 제안했다[30]. XAI에 대한 니즈는 사용자의 배경과 관심 수준에 따라 다양한 것으로 밝혀지고 있으나, 사용자의 특성에 따른 맞춤형 XAI 접근법에 대한 연구는 부족하며 최종 사용자(end-user)의 니즈를 충족하지 못하는 상황이다[30].

사용자의 다양한 특성 중에서도 AI 시스템 사용 경험의 여부는 맞춤형 설명을 제공하는 데 있어 중요한 변인이다 [29],[31]. 이미지 생성형 AI 사용 경험이 부족한 10명의 비전문가를 대상으로 한 연구에 따르면, 절반의 참가자가 주어진 작업을 해결하기 위해 텍스트 프롬프트 작성을 시작하는 것에 어려움을 겪었다[32]. 또한, 복잡하고 반복적 수정이 필요한 프롬프팅 방식으로 인해 초보 사용자(novice)에게는 이미지 생성형 AI 사용이 여전히 어려운 과제인 것으로 나타났다[5]. 이 연구에서는 숙련 사용자(advanced)를 ‘목적 지향적으로 이미지 생성형 AI를 주기적으로 활용하는 사용자’로, 초보 사용자(novice)를 ‘이미지 생성형 AI를 사용해보지 않았거나 일회적 체험으로 시스템에 대한 이해가 부족한 사용자’로 조작적 정의했다.

동적 설명 방식은 이미지 생성 과정에서 사용자가 자신의

필요에 따라 모델에 대해 대화를 나눌 수 있으며 맥락에 맞는 맞춤형 설명을 얻을 수 있다는 점에서 초보 사용자(novice)의 사용 경험을 향상시킬 것으로 기대된다. 특히 시스템이 사용자에게 풍부한 설명 피드백을 제공하여 반복적으로 프롬프트 작성 능력을 향상시킬 수 있는 기능이 필요하다는 관점에서[33], 시스템과의 상호작용을 통한 동적 설명은 이를 충족할 수 있다. 또한, 개발 및 디자인 도구 등 복잡한 시스템을 처음 사용하는 과정에서 자연어 쿼리(query)의 사용은 사용자가 더 빠르게 시스템을 학습할 수 있게 돕는다는 점에서[34], 시스템과의 대화를 통한 동적 설명은 초보 사용자의 이해와 학습을 도울 것으로 기대된다.

이 연구에서는 이미지 생성형 AI의 사용 경험 여부에 따른 설명 방식의 영향을 알아보려고 했다. 특히, 초보 사용자에게 있어 동적 설명 방식이 이미지 생성형 AI에 대한 설명가능성을 향상시키는지 알아보았다.

III. 연구 방법

3-1 실험 설계

1) 연구 질문과 독립 변인

이 연구는 이미지 생성형 AI의 설명 방식이 초보 사용자(novice)와 숙련 사용자(advanced)가 인지하는 설명가능성(explainability)에 미치는 영향을 검증하고자 했다. 특히, 동적 설명이 초보 사용자가 인지하는 이미지 생성형 AI에 대한 설명가능성에 긍정적인 영향을 미치는지 알아보았다.

실험은 2(초보 사용자와 숙련 사용자)X2(정적 설명과 동적 설명)의 이원 혼합 설계(two-way mixed design)로 구성했다. 첫 번째 독립 변수인 이미지 생성형 AI ‘경험 여부’는 참가자 간 설계로 두 집단(초보 사용자와 숙련 사용자)으로 나누었다. ‘경험 여부’ 독립 변인의 집단은 선행연구[5],[30],[35]를 기반으로 설정했다. 초보 사용자(novice) 집단은 이미지 생성형 AI를 사용한 경험이 없거나 일회적인 사용 경험이 있는 사용자로 설정했다. 숙련 사용자(advanced)는 이미지 생성형 AI를 주기적으로 월 1회 이상 사용하는 사용자로 설정했다.

두 번째 독립 변수인 ‘설명 방식’은 정적 설명과 동적 설명의 두 조건으로 나누었으며, 참가자 내 설계로 반복 측정했다. 정리하면, 참가자는 사전 설문을 통해 사용 경험에 따른 두 집단으로 나뉘었으며, 두 가지 설명 조건에 따른 실험 처치물을 무작위 순서(counterbalancing)로 모두 할당받아 평가하였다. 이에 대한 통계 분석으로는 이원 반복측정 분산분석을 실시하였다. 연구 질문과 연구 가설은 다음과 같다.

연구 질문. 동적 설명은 초보 사용자가 인지하는 이미지 생성형 AI에 대한 설명가능성에 긍정적인 영향을 미치는가

연구 가설1. 이미지 생성형 AI ‘경험 여부’에 따라 설명가능성 평가에 차이가 있다.

연구 가설2. 이미지 생성형 AI의 ‘설명 유형’에 따라 설명가능성 평가에 차이가 있다.

연구 가설3. 이미지 생성형 AI의 ‘설명 유형’에 따른 설명가능성 평가는 사용자의 ‘경험 여부’에 따라 차이가 있다.

2) 실험 처치물: ‘그려줘AI’

이 연구에서는 이미지 생성형 AI의 경험 여부와 설명 방식이 사용자가 인지하는 설명가능성에 미치는 효과를 알아보기 위해 두 가지 설명 방식에 따른 이미지 생성형 AI 프로토타입 ‘그려줘AI’를 디자인하였으며, 사용 과정을 영상물로 제작하여 실험에 활용하였다. 그려줘AI는 대표적인 이미지 생성형 AI 서비스인 DALL-E, Midjourney, Stable Diffusion 사례 분석을 통해 개별 속성을 기반으로 자연어 프롬프팅과 그래픽 사용자 인터페이스를 통합한 형태로 디자인하였다(그림 1, 2). 프로토타입은 제공되는 설명 방식에 따라 정적 설명을 제공하는 ‘그려줘AI1’과 동적 설명을 제공하는 ‘그려줘AI2’로 나누었으며, 참가자는 두 가지 프로토타입 사용 영상을 무작위 순서(counterbalancing)로 시청하고 평가하였다.

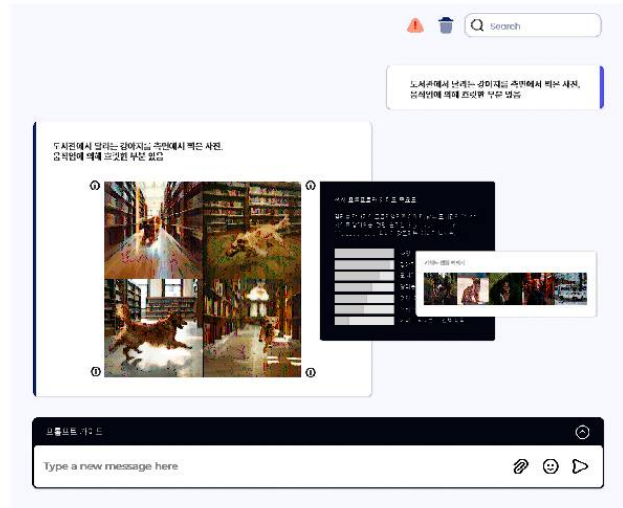
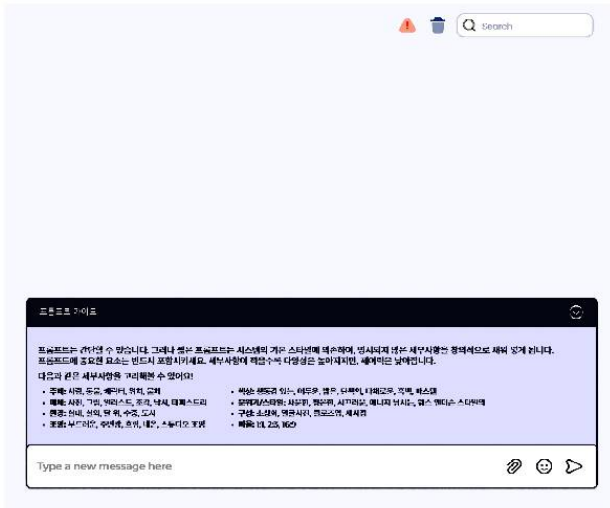
정적 설명을 제공하는 ‘그려줘AI1’은 이미지 생성 과정에서 매번 동일한 설명을 얻을 수 있도록 설계되었다(그림 1). 모델에 대한 설명 및 프롬프트 작성 가이드는 메시지 박스 위에 배치하여 프롬프팅 과정에서 활용할 수 있도록 제시되었다. 또한, 생성된 이미지에 대해서는 매번 동일하게 특징 중요도 설명과 사례 기반 설명을 시각화하여 제시했다.

동적 설명을 제공하는 ‘그려줘AI2’는 대화를 통해 상황에 맞는 설명을 얻을 수 있도록 설계되었다(그림 2). 따라서 사용자는 메시지 박스 위에 배치되는 맥락별 프롬프트 추천을 눌러 설명을 확인하거나 직접 궁금한 점을 입력하여 원하는 설명을 얻을 수 있다. 이때 제공되는 설명은 내용과 맥락에 따라 적절한 방식으로 선별되어 제공되었다.

두 가지 ‘그려줘AI’에 대해 설명 방식 이외의 변수를 통제하기 위해서 두 프로토타입의 인터페이스와 영상 속 사용자가 입력하는 프롬프트 및 생성되는 이미지를 모두 동일하게 구성했다. 또한, 동적 설명 조건에서 제시되는 설명은 정적 설명에서 제시되는 특징 중요도 설명과 사례 기반 설명을 맥락에 따라 선별하여 제공하도록 하여 두 조건에서 제시되는 정보가 유사하도록 조절하였다.

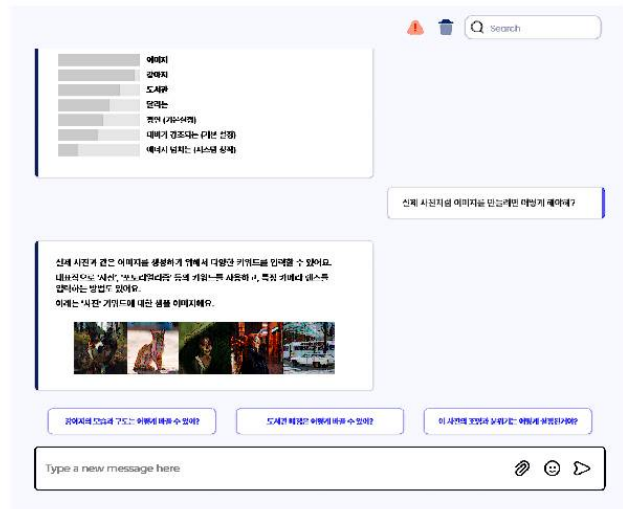
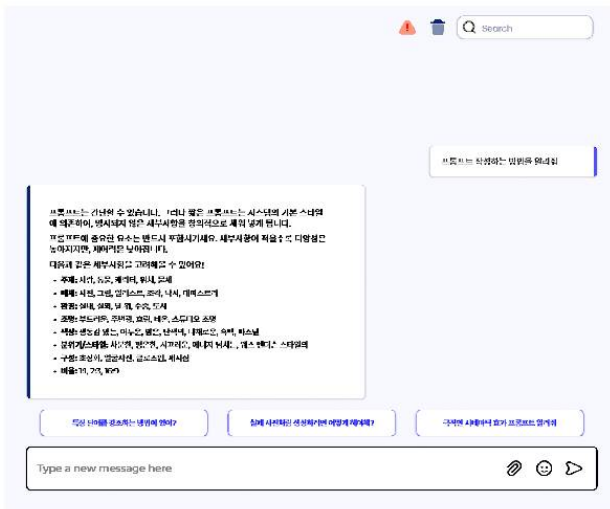
3) 실험 시나리오

참가자가 시청한 영상물의 실험 시나리오는 이미지 생성형 AI의 사용 목적과 사용 행동에 대한 선행 연구를 참고하여 설계하였다. 이미지 생성형 AI의 사용 목적은 아이디어이션을 위한 탐색(exploration), 생성물과 인간의 작업물을 통합하는 합성(composites), 생각하는 정확한 이미지를 도출하기 위한 명확한 목표(targeted goals)로 나눌 수 있다[36]. 명확한 목표로 이미지를 생성하는 경우 시스템에 대한 설명가



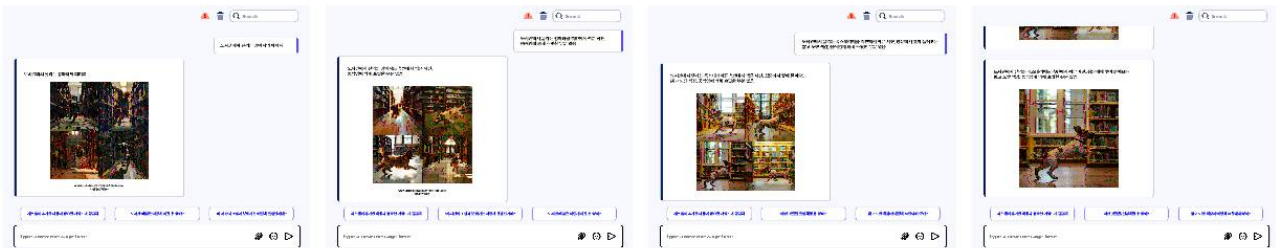
*This is an image of experimental treatments designed for Korean participants.

그림 1. 정적 설명을 제공하는 '그려줘AI'
Fig. 1. Static explanation of 'DrawAI'



*This is an image of experimental treatments designed for Korean participants.

그림 2. 동적 설명을 제공하는 '그려줘AI'
Fig. 2. Dynamic explanation of 'DrawAI'



*This is an image of experimental treatments designed for Korean participants.

그림 3. 이미지 생성형 AI 사용 시나리오
Fig. 3. Image generative AI usage scenario

능성의 필요가 가장 높다는 선행 연구[36]에 따라, 이 연구에서도 생성하고자 하는 이미지가 명확한 상황을 시나리오로 설정했다. 한편, 이미지 생성형 AI의 창작 과정은 크게 초기 프롬프트 구조화 과정(structure)과 이후의 수정 과정(refinement)으로 나눌 수 있다[36],[37]. 따라서 이 연구에서는 초기 구조화 과정과 두 번의 수정 과정을 포함하여 사용자가 총 세 번 프롬프트를 입력하도록 시나리오를 설계했다.

시나리오에서 사용자는 그림 3의 ‘도서관에서 달리는 강아지’의 이미지 생성을 명확한 목표로 하여 프롬프트 구성과 수정 단계를 통해 이미지를 생성하게 된다. 참가자는 동일한 시나리오에 대해 정적 설명과 동적 설명을 제공하는 두 가지 버전의 ‘그려줘AI’ 사용 영상을 무작위 순서(counterbalancing)로 시청하고 각각에 대해 평가하였다.

4) 종속 변인

사용 경험과 설명 방식에 따른 이미지 생성형 AI의 설명가능성을 측정하기 위해 선행 연구에 기반하여 설명가능성, 이해성, 정보성, 신뢰성, 만족도 지표의 7점 리커트 척도 평가와 참가자의 주관적 의견을 수집하였다. XAI 시스템에 대한 평가에 있어 인간 중심적(human-centered) 프레임워크[38]와 이미지 생성형 AI의 설명가능성에 대한 실무적 평가 프레임워크[39]를 종합하여 본 연구에 적용 가능한 다섯 가지 지표를 종속변인으로 설정하였다.

설명가능성(Explainability)은 시스템의 작동 원리와 작동 결과에 대해 인간이 이해할 수 있는 방식으로 설명을 제공하는 시스템의 속성으로 정의하였다[39],[40]. 이해성(Understandability)은 모델의 내부 구조나 알고리즘적 처리 방식을 설명할 필요 없이 모델이 어떻게 작동하는지 인간이 이해할 수 있는 특성으로 정의된다[38],[40]. 정보성(Informativeness)은 시스템의 작동 방식과 결과 및 문제 해결에 유용한 정보를 제공하여 의사 결정을 지원하는 특성으로 정의된다[40]. 신뢰성(Trustworthiness)은 사용자가 시스템이 제공하는 설명과 생성물에 대해 믿을 수 있는 정도로 정의된다[39]. 만족도(Satisfaction)는 시스템이 제공하는 설명을 바탕으로 시스템 사용 목적을 달성함으로써 발생하는 만족감으로 정의하였다[28],[31]. 각 지표를 평가하기 위한 하위 설문 문항은 선행 연구[28],[29],[31]를 참고하여 작성 및 사용하였으며, Cronbach’s α 를 확인하여 신뢰도를 검증했다.

3-2 실험 절차

온라인을 통해 실험 참가자를 모집하였으며, 약 15~20분이 소요되는 비대면 온라인 실험을 진행했다. 사전 설문 단계에서는 참가자의 인구통계학적 정보와 이미지 생성형 AI 사용 경험에 대한 설문을 진행하였다. 이후 본 실험 단계에서 참가자는 ‘그려줘 AI’를 사용하여 ‘도서관에서 달리는 강아지’ 이미지를 생성 및 수정하는 실험 영상을 각각 제공하였다. 이

때, 참가자는 설명 방식에 따라 ‘그려줘AI’의 두 가지 버전(정적 설명, 동적 설명)을 확인하였으며, 두 조건의 제시 순서는 무작위(counterbalancing)로 설정되었다.

참가자는 설명 방식에 따른 한 조건의 영상을 시청한 후 해당 영상을 시청했는지 확인을 위한 객관식 질문에 답변했으며, 이후 설명가능성을 평가하기 위해 주관식 문항 및 리커트 7점 척도로 설명가능성, 이해도, 정보성, 신뢰성, 만족도 관련 문항에 응답하도록 했다. 이후 참가자들은 다른 설명 유형으로도 동일한 절차의 영상 시청, 확인 질문, 평가를 진행했으며, 마지막으로 ‘그려줘 AI’에 대한 주관식 의견을 응답하며 설문을 종료했다.

3-3 참가자

이 연구에서는 2024년 6월 5일에서 6월 30일까지 온라인 실험을 실시하여 총 104명이 실험 참가했다. 그 중 실험 처리물 영상을 시청하지 않은 불성실 응답자 27명이 선별되었다. 불성실 응답자를 제외한 총 77명의 데이터를 수집하여 분석하였다. 사용 경험 조건에 따라서는 이미지 생성형 AI의 초보 사용자(novice) 46명, 숙련 사용자(advanced) 31명의 데이터를 수집했다. 참가자의 인구통계학적 정보는 표 2와 같다. 모든 실험 참여자들은 사전에 실험 내용과 절차에 관해 안내를 받았으며, 실험 종료 이후 3,000원 상당의 보상을 지급받았다.

표 2. 인구통계학 정보

Table 2. Survey demographics results

	class	Novice(%)	Advanced(%)
Gender	Male	18 (39.1)	10 (32.3)
	Female	28 (60.9)	21 (67.7)
Age	10 to under 20	1 (2.2)	1 (3.2)
	20 to under 30	40 (87.0)	21 (67.7)
	30 to under 40	3 (6.5)	9 (29)
	40 and over	2 (4.3)	0 (0)
Academic Background	Liberal arts	22 (47.8)	7 (22.6)
	Science	16 (34.8)	11 (35.5)
	Et cetra	8 (17.4)	13 (41.9)
Total		46 (100)	31 (100)

IV. 연구 결과

4-1 설문 문항의 신뢰도 검증

설명가능성(Explainability), 만족도(Satisfaction), 이해성(Understandability), 정보성(Informativeness), 신뢰성(Trustworthiness) 설문 문항의 신뢰도를 검증하기 위해 Cronbach’s α 계수를 통한 문항 신뢰도 분석을 진행하였다. 그 결과, 설명가능성($\alpha = .70$), 만족도($\alpha = .66$), 이해성($\alpha =$

.81), 정보성($\alpha = .74$), 신뢰성($\alpha = .72$)으로 나타났다. 만족도는 0.7보다 작은 것으로 나타났으나 0.6 이상이면 신뢰도가 수용 가능한 것으로 판단되고, 이외 항목에 대하여 Cronbach's α 값이 0.7 이상으로 분석되어 종속 변인에 대한 신뢰도가 검증되었다. 각 변인의 왜도는 절대값 3미만, 첨도는 절대값 8을 초과하지 않아, 정규성에 위배되지 않는 것으로 확인되었다.

4-2 종속 변인의 측정 결과

설명 방식과 사용 경험 여부가 사용자가 느끼는 시스템의 설명가능성에 미치는 영향을 알아보기 위해 이원배치 반복측정 분산분석(Two-way Repeated Measures ANOVA)을 실시하였다. 변인 별 평균과 표준편차의 값은 표 3, 분산분석 결과는 표 4와 같다. 종속변수 별로 각 요인의 상호작용 효과를 검증했으며, 결과는 다음과 같다.

표 3. 모든 종속변인의 평균(표준편차)
Table 3. Mean(SD) of all dependent variables

Variables	Explanation	Novice	Advanced
Explainability	Static	5.20 (0.86)	5.41 (0.81)
	Dynamic	5.59 (0.81)	5.48 (0.89)
Understanbability	Static	4.70 (1.1)	5.29 (0.90)
	Dynamic	5.05 (0.89)	5.46 (0.61)
Informativeness	Static	4.80 (0.88)	5.54 (0.51)
	Dynamic	5.17 (0.92)	5.43 (0.84)
Trustworthiness	Static	4.96 (0.92)	5.45 (0.84)
	Dynamic	5.08 (1.00)	5.25 (0.63)
Satisfaction	Static	4.84 (1.1)	5.66 (0.69)
	Dynamic	5.16 (0.96)	5.37 (0.69)

표 4. 이원 반복측정 분산분석 결과
Table 4. Results of a two-way repeated measures ANOVA

Variables	Explanation	MS	F	p
Explainability	Explanation(E)	2.08	4.24*	.04
	Use Experience(UE)	.98	.11	.75
	E X UE	.97	1.97	.16
Understanbability	Explanation(E)	2.57	4.89*	.03
	Use Experience(UE)	9.38	8.21**	.00
	E X UE	.31	.59	.45
Informativeness	Explanation(E)	.55	1.43	.24
	Use Experience(UE)	9.4	10.868***	.00
	E X UE	2.14	5.45*	.02
Trustworthiness	Explanation(E)	.07	.15	.70
	Use Experience(UE)	3.98	3.70	.05
	E X UE	.95	1.99	.16
Satisfaction	Explanation(E)	.01	.02	.88
	Use Experience(UE)	9.87	9.43**	.00
	E X UE	3.52	6.32*	.01

*** $p < .001$, ** $p < .01$, * $p < .05$

1) 설명가능성

설명가능성은 설명 방식에 따라 유의한 차이가 있는 것으로 나타났다[F(1, 75) = 4.24, $p < .05$]. 반면, 경험 여부에 따른 유의한 차이는 발견되지 않았다[F(1, 75) = .11, $p = .75$]. 또한, 설명 방식과 경험 여부에 따른 상호작용 효과가 유의하지 않은 것으로 나타났다[F(1, 75) = 1.97, $p = .16$].

2) 이해성

이해성은 설명 방식에 따라 유의한 차이가 있는 것으로 나타났다[F(1, 75) = 4.89, $p < .05$]. 또한, 경험 여부에 따라 유의한 차이가 발견되었다[F(1, 75) = 8.21, $p < .01$]. 그러나, 설명 방식과 경험 여부에 따른 상호작용 효과는 유의하지 않은 것으로 나타났다[F(1, 75) = .59, $p = .45$].

3) 정보성

정보성에 있어 설명 방식에 따른 유의한 차이는 발견되지 않았다[F(1, 75) = 1.43, $p = .24$]. 그러나, 경험 여부에 따라 유의한 차이가 발견되었다[F(1, 75) = 10.868, $p < .01$]. 또한, 설명 방식과 경험 여부에 따른 상호작용 효과가 유의한 것으로 나타났다[F(1, 75) = 5.45, $p < .05$]. 이를 구체적으로 살펴보기 위해 Bonferroni 사후 검정을 통하여 단순 주 효과 분석(Simple main effects)을 실시한 결과, 정적 설명을 제공한 경우에 미사용자(95%, CI[.80, .19], $p < .001$)가 기사사용자(95%, CI[.19, .80], $p < .001$)에 비해 정보성을 더 낮게 인지하는 것으로 나타났다.

4) 신뢰성

신뢰성에 있어서는 설명 방식에 따른 유의한 차이가 발견되지 않았다[F(1, 75) = .15, $p = .70$]. 또한, 경험 여부에 따라 유의한 차이가 나타나지 않았다[F(1, 75) = 3.70, $p = .05$]. 설명 방식과 경험 여부에 따른 상호작용 효과도 유의하지 않은 것으로 나타났다[F(1, 75) = 1.99, $p = .16$].

5) 만족도

만족도에 있어 설명 방식에 따른 유의한 차이는 발견되지 않았다[F(1, 75) = .02, $p = .88$]. 반면, 경험 여부에 따라 만족도에 유의한 차이가 있는 것으로 나타났다[F(1, 75) = 9.43, $p < .01$]. 또한, 설명 방식과 경험 여부에 따른 상호작용 효과가 유의한 것으로 나타났다[F(1, 75) = 6.32, $p < .05$]. 이를 구체적으로 살펴보기 위해 Bonferroni 사후 검정을 통하여 단순 주 효과 분석(Simple main effects)을 실시한 결과, 정적 설명을 제공한 경우에 미사용자(95%, CI[.85, .18], $p < .001$)가 기사사용자(95%, CI[.18, .85], $p < .001$)에 비해 만족도가 더 낮은 것으로 나타났다.

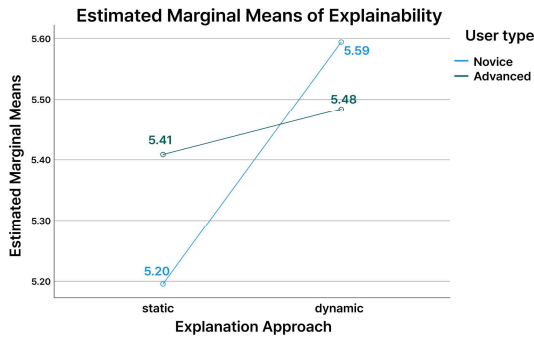


그림 4. 설명가능성의 기술통계 그래프
 Fig. 4. Statistics graph of explainability

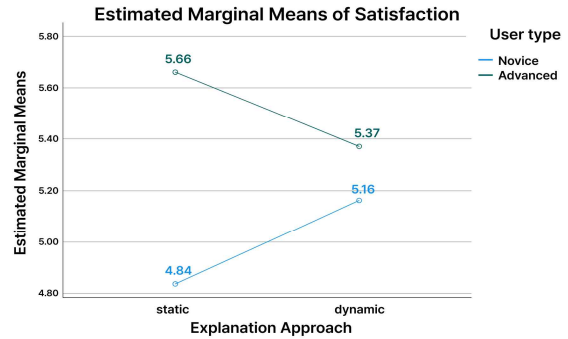


그림 8. 만족도의 기술통계 그래프
 Fig. 8. Statistics graph of satisfaction

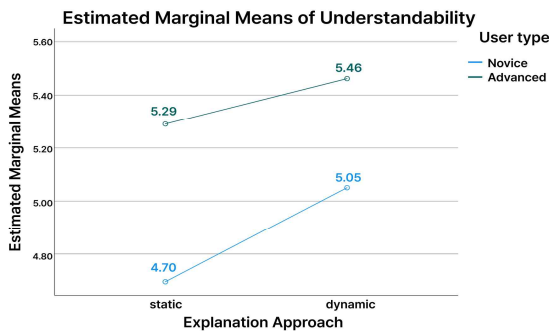


그림 5. 이해성의 기술통계 그래프
 Fig. 5. Statistics graph of understandability

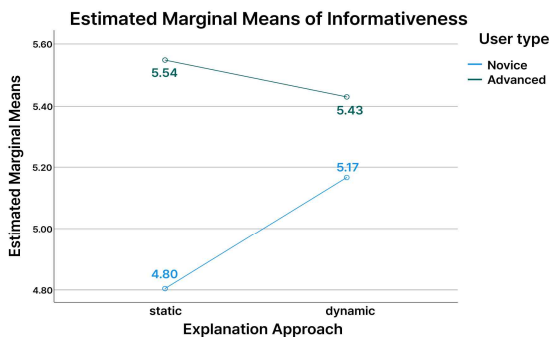


그림 6. 정보성의 기술통계 그래프
 Fig. 6. Statistics graph of informativeness

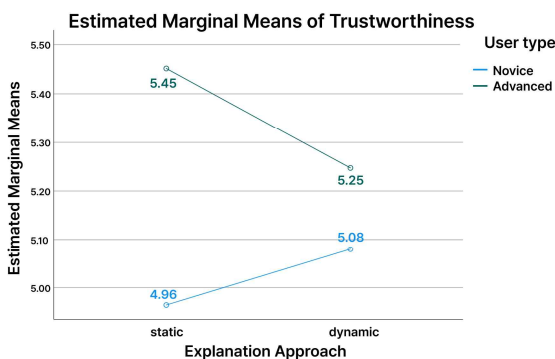


그림 7. 신뢰성의 기술통계 그래프
 Fig. 7. Statistics graph of trustworthiness

V. 결론 및 논의

이 연구는 이미지 생성형 AI의 설명 방식이 초보 사용자 (novice)와 숙련 사용자(advanced)가 인지하는 설명가능성 (explainability)에 미치는 영향을 실증적으로 분석하였다. 특히, 동적(dynamic) 설명이 초보 사용자(novice)가 인지하는 이미지 생성형 AI의 설명가능성에 긍정적인 영향을 미치는지 검증하고자 했다.

연구 결과, 시스템과 상호작용하며 사용자의 필요와 맥락에 맞는 설명을 제공하는 동적(dynamic) 설명은 정적(static) 설명에 비해 초보 사용자가 인지하는 이미지 생성형 AI의 설명가능성을 향상시키는 것으로 나타났다. 첫 번째, 동적 설명이 정적 설명에 비해 설명가능성과 이해성 지표에서 더 높은 평균값을 나타냈고, 두 설명 방식 간의 차이가 통계적으로 유의미했다. 두 번째, 정보성과 만족도 지표에서 설명 방식과 사용 경험 간의 상호작용 효과가 유의미한 것으로 확인되었다. 이에 대한 주효과 분석 결과, 정적 설명 조건에서 초보 사용자가 인지하는 정보성과 만족도가 숙련 사용자에 비해 유의미하게 낮은 것으로 나타났다.

이는 초보 사용자 집단의 주관적 의견과도 부합하는데, 시스템과 질문을 주고 받으면서 사용자 수준과 필요에 맞는 설명을 얻을 수 있다는 점에서 동적 설명을 더 긍정적으로 평가하였다. P32는 ‘(동적 설명 조건이) 더 원활하게 소통할 수 있다고 느꼈고, 내가 궁금한 것을 적극적으로 반영해준다는 느낌을 받았다’고 보고했으며, P46은 ‘대화의 형식으로 풀어서 설명하는 것이 이해하기에 더 도움이 된다고 느끼며, 각 상황마다 사용자가 할 만한 질문 등을 제시하는 것이 프로그램에 폭넓게 적응하도록 돕는다’고 언급하였다.

나아가 이미지 생성형 AI의 동적 설명은 숙련 사용자에 비해 초보 사용자가 인지하는 설명가능성에 더 긍정적 영향을 미치는 것으로 나타났다. 특히, 숙련 사용자 집단이 인지하는 설명가능성은 동적 설명과 정적 설명에 따라 큰 차이가 없었으며, 정적 설명의 정보성과 만족도에 대해서는 초보 사용자보다 유의미하게 더 높게 평가한 것으로 나타났다. 참가자의

주관적 의견에 따르면, '(동적 설명은) 궁금한 점에 따라 이미지가 어떤 원리로 생성되었는지 더 명확히 알 수 있어 좋으나 매번 질문해야 하는 번거로움이 있다'(P24)고 언급하였다. 이는 설명가능한 AI 디자인에 있어서 초보 혹은 숙련 사용자 등 사용자 특성을 반영하여 각 집단의 요구를 충족할 필요성을 시사한다.

정리하면, 이미지 생성형 AI의 동적 설명은 초보 사용자가 인지하는 설명가능성에 긍정적인 영향을 미치는 것으로 나타났다. 숙련 사용자의 경우, 동적 설명과 정적 설명의 차이가 두드러지지 않았다. 참가자의 주관적 의견과 종합했을 때, 동적 설명의 맞춤형 상호작용 피드백은 초보 사용자가 이미지 생성형 AI에 적응하고 시스템을 이해하는 것을 돕는다. 따라서 실무적으로는 초보 사용자의 성공적인 온보딩과 락인(Lock-in)을 위해 개별 사용자의 필요와 매번 새롭게 생성되는 결과물의 특성을 반영하는 동적 설명을 제공할 필요성이 있다.

연구 결과를 바탕으로, 생성형 AI 시스템에서 초보 사용자를 위한 동적 설명의 디자인 고려 사항을 제안한다. 첫 번째, 초보 사용자를 위한 동적 설명은 맥락과 목적에 따라 맞춤형 피드백이 가능하도록 자연어를 통한 대화 형태로 설계할 수 있다[10]. 특히, 이 연구에서는 초보 사용자가 설명을 얻기 위해 어떤 질문을 할 수 있는지 예시 질문을 생성 및 제안하여 초보 사용자의 대화를 돕고자 했다. 두 번째, 초보 사용자와 숙련 사용자 집단을 모두 고려하고, 초보 사용자의 학습 효과 또한 고려하여, 동적 설명과 정적 설명을 선택할 수 있는 시스템 옵션을 제공하여 사용자 자유도(autonomy)를 높이고 상이한 사용자 집단의 요구를 충족할 수 있다.

이 연구는 신규 사용자의 진입 장벽이 높은 이미지 생성형 AI의 문제점에 대응하여 초보 사용자의 시스템 설명가능성을 높이기 위한 방안으로 동적 설명의 효과를 확인했다는 점에서 의의가 있다. 또한, 다양한 사용자 집단의 상이한 요구에 반응하는 설명가능성 접근이 필요하다는 점을 강조하고, 초보 사용자를 위한 동적 설명의 디자인 고려 사항을 제안했다.

이러한 기여에도 불구하고 이 연구에는 몇 가지 한계가 있다. 첫째, 실험 환경에서 정해진 시나리오를 기반으로 연구가 이루어져 실제 사용 환경과의 차이가 있을 수 있다. 특히, 이미지 생성 AI는 구체적 결과물을 위한 이미지 생성과 아이디어이션을 위한 사용 등 다양한 목적으로 활용되기 때문에 후속 연구에서는 이러한 목적에 따른 효과를 살펴볼 필요가 있다. 둘째, 이 연구에서는 설명 방식의 조건별로 단일 시나리오를 제시하여 일회적인 경험을 측정했다는 점에서 한계가 있다. 따라서 이 연구에서 설계한 프로토타입을 구현 후 다이어리 방법론 등으로 사용자가 보다 다양한 상황에서 설명 방식을 체험할 수 있는 장기적인 경험에 대해 연구할 필요가 있다.

감사의 글

이 연구는 2024년도 연세대학교 연구비의 지원을 받아 수행되었습니다.

참고문헌

- [1] Nielsen Norman Group. AI: First New UI Paradigm in 60 Years [Internet]. Available: <https://www.nngroup.com/articles/ai-paradigm/>.
- [2] J. Casteleiro-Pitrez, "Generative Artificial Intelligence Image Tools among Future Designers: A Usability, User Experience, and Emotional Analysis," *Digital*, Vol. 4, No. 2, pp. 316-332, April 2024. <https://doi.org/10.3390/digital4020016>
- [3] Adobe Blog. Creative Pros Are Leveraging Generative AI to Do More and Better Work [Internet]. Available: <https://blog.adobe.com/en/publish/2024/02/02/creative-pros-generative-ai-usage>.
- [4] H. Yu, Y. Dong, and Q. Wu, "User-Centric AIGC Products: Explainable Artificial Intelligence and AIGC Products," arXiv:2308.09877, August 2023. <https://doi.org/10.48550/arXiv.2308.09877>
- [5] Z. Wang, Y. Huang, D. Song, L. Ma, and T. Zhang, "PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement," in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, Honolulu: HI, 185, May 2024. <https://doi.org/10.1145/3613904.3642803>
- [6] D. M. Reddy, S. M. Basha, M. C. Hari, and N. Penchalaiah, "Dall-E: Creating Images from Text," *UGC Care Group I Journal*, Vol. 8, No. 14, pp. 71-75, 2021.
- [7] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv:1312.6114, December 2013. <https://doi.org/10.48550/arXiv.1312.6114>
- [8] H.-K. Ko, G. Park, H. Jeon, J. Jo, J. Kim, and J. Seo, "Large-Scale Text-to-Image Generation Models for Visual Artists' Creative Works," in *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*, Sydney, Australia, pp. 919-933, March 2023. <https://doi.org/10.1145/3581641.3584078>
- [9] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, Vol. 6, pp. 52138-52160, September 2018. <https://doi.org/10.1109/access.2018.2870052>
- [10] J. Schneider, "Explainable Generative AI (GenXAI): A Survey, Conceptualization, and Research Agenda,"

- arXiv:2404.09554, April 2024. <https://doi.org/10.48550/arXiv.2404.09554>
- [11] C. Singh, J. P. Inala, M. Galley, R. Caruana, and J. Gao, “Rethinking Interpretability in the Era of Large Language Models,” arXiv:2402.01761, January 2024. <https://doi.org/10.48550/arXiv.2402.01761>
- [12] D. Slack, S. Krishna, H. Lakkaraju, and S. Singh, “Explaining Machine Learning Models with Interactive Natural Language Conversations Using TalkToModel,” *Nature Machine Intelligence*, Vol. 5, No. 8, pp. 873-883, August 2023. <https://doi.org/10.1038/s42256-023-00692-8>
- [13] Y. Gao, T. Sheng, Y. Xiang, Y. Xiong, H. Wang, and J. Zhang, “Chat-REC: Towards Interactive and Explainable LLMs-Augmented Recommender System,” arXiv:2303.14524, April 2023. <https://doi.org/10.48550/arXiv.2303.14524>
- [14] M. T. Ribeiro, S. Singh, and C. Guestrin, ““Why Should I Trust You?”: Explaining the Predictions of Any Classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, San Francisco: CA, pp. 1135-1144, August 2016. <https://doi.org/10.1145/2939672.2939778>
- [15] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*, Long Beach: CA, pp. 4768-4777, December 2017.
- [16] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps,” arXiv:1312.6034, December 2013. <https://doi.org/10.48550/arXiv.1312.6034>
- [17] D. Bahdanau, K. H. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” arXiv:1409.0473, September 2014. <https://doi.org/10.48550/arXiv.1409.0473>
- [18] S. Wachter, B. Mittelstadt, and C. Russell, “Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR,” *Harvard Journal of Law & Technology*, Vol. 31, No. 2, pp. 841-887, 2018. <https://doi.org/10.2139/ssrn.3063289>
- [19] A. Aamodt and E. Plaza, “Case-Based Reasoning: Foundational Issues, Methodological Variations, And System Approaches,” *AI Communications*, Vol. 7, No. 1, pp. 39-59, 1994. <https://doi.org/10.3233/AIC-1994-7104>
- [20] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, New York, NY: Chapman and Hall/CRC, 1984. <https://doi.org/10.1201/9781315139470>
- [21] R. Andrews, J. Diederich, and A. B. Tickle, “Survey and Critique of Techniques for Extracting Rules from Trained Artificial Neural Networks,” *Knowledge-Based Systems*, Vol. 8, No. 6, pp. 373-389, December 1995. [https://doi.org/10.1016/0950-7051\(96\)81920-4](https://doi.org/10.1016/0950-7051(96)81920-4)
- [22] A. Bunt, M. Lount, and C. Lauzon, “Are Explanations Always Important?: A Study of Deployed, Low-Cost Intelligent Interactive Systems,” in *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces (IUI '12)*, Lisbon, Portugal, pp. 169-178, February 2012. <https://doi.org/10.1145/2166966.2166996>
- [23] J. Krause, A. Perer, and K. Ng, “Interacting with Predictions: Visual Inspection of Black-Box Machine Learning Models,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, San Jose: CA, pp. 5686-5697, May 2016. <https://doi.org/10.1145/2858036.2858529>
- [24] K. Yin and G. Neubig, “Interpreting Language Models with Contrastive Explanations,” in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, Abu Dhabi, UAE, pp. 184-198, December 2022. <https://doi.org/10.18653/v1/2022.emnlp-main.14>
- [25] U. Ehsan and M. O. Riedl, “Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach,” in *Proceedings of the 22nd HCI International Conference (HCII 2020)*, Copenhagen, Denmark, pp. 449-466, July 2020. https://doi.org/10.1007/978-3-030-60117-1_33
- [26] M. Langer, D. Oster, T. Speith, H. Hermanns, L. Kästner, E. Schmidt, ... and K. Baum, “What Do We Want From Explainable Artificial Intelligence (XAI)? - A Stakeholder Perspective on XAI and a Conceptual Model Guiding Interdisciplinary XAI Research,” *Artificial Intelligence*, Vol. 296, 103473, July 2021. <https://doi.org/10.1016/j.artint.2021.103473>
- [27] A. Preece, D. Harborne, D. Braines, R. Tomsett, and S. Chakraborty, “Stakeholders in Explainable AI,” arXiv:1810.00184, September 2018. <https://doi.org/10.48550/arXiv.1810.00184>
- [28] R. Tomsett, D. Braines, D. Harborne, A. Preece, and S. Chakraborty, “Interpretable to Whom? A Role-Based Model for Analyzing Interpretable Machine Learning Systems,” arXiv:1806.07552, June 2018. <https://doi.org/10.48550/arXiv.1806.07552>
- [29] H. Suresh, S. R. Gomez, K. K. Nam, and A. Satyanarayan, “Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and Their Needs,” in *Proceedings of the 2021*

- CHI Conference on Human Factors in Computing Systems (CHI '21)*, Yokohama, Japan, 74, May 2021. <https://doi.org/10.1145/3411764.3445088>
- [30] S. S. Y. Kim, E. A. Watkins, O. Russakovsky, R. Fong, and A. Monroy-Hernández, ““Help Me Help the AI”: Understanding How Explainability Can Support Human-AI Interaction,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, Hamburg, Germany, 250, April 2023. <https://doi.org/10.1145/3544548.3581001>
- [31] S. Mohseni, N. Zarei, and E. D. Ragan, “A Survey of Evaluation Methods and Measures for Interpretable Machine Learning,” arXiv:1811.11839, November 2018. <https://doi.org/10.48550/arXiv.1811.11839>
- [32] J. D. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, and Q. Yang, “Why Johnny Can’t Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, Hamburg Germany, 437, April 2023. <https://doi.org/10.1145/354454.8.3581388>
- [33] H. Strobel, A. Webson, V. Sanh, B. Hoover, J. Beyer, H. Pfister, and A. M. Rush, “Interactive and Visual Prompt Engineering for Ad-Hoc Task Adaptation with Large Language Models,” *IEEE Transactions on Visualization and Computer Graphics*, Vol. 29, No. 1, pp. 1146-1156, January 2023. <https://doi.org/10.1109/TVCG.2022.3209479>
- [34] T. S. Kim, D. E. Choi, Y. Choi, and J. Kim, “Stylette: Styling the Web with Natural Language,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, New Orleans: LA, 5, April-May 2022. <https://doi.org/10.1145/3491102.3501931>
- [35] J.-K. Ryou and Y.-H. Pan, “A Study on the User Experience Design of Blockchain Wallet Considering Users’ Technology Knowledge Level,” *Journal of Digital Contents Society*, Vol. 21, No. 12, pp. 2073-2081, December 2020. <https://doi.org/10.9728/dcs.2020.21.12.2073>
- [36] A. Mahdavi Goloujeh, A. Sullivan, and B. Magerko, “Is It AI or Is It Me? Understanding Users’ Prompt Journey with Text-to-Image Generative AI Tools,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, Honolulu: HI, 183, May 2024. <https://doi.org/10.1145/3613904.3642861>
- [37] D. Han, D. Choi, and C. Oh, “A Study on User Experience through Analysis of the Creative Process of Using Image Generative AI: Focusing on User Agency in Creativity,” *The Journal of the Convergence on Culture Technology*, Vol. 9, No. 4, pp. 667-679, July 2023. <https://doi.org/10.17703/JCCT.2023.9.4.667>
- [38] P. Lopes, E. Silva, C. Braga, T. Oliveira, and L. Rosado, “XAI Systems Evaluation: A Review of Human and Computer-Centred Methods,” *Applied Sciences*, Vol. 12, No. 19, 9423, September 2022. <https://doi.org/10.3390/app12199423>
- [39] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, ... and F. Herrera, “Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI,” *Information Fusion*, Vol. 58, pp. 82-115, June 2020. <https://doi.org/10.1016/j.inffus.2019.12.012>
- [40] Z. Ren, X. Qin, and B. Wang, “Unraveling the Impact of Explainability of Artificial Intelligence-Generated Content(AIGC) on Design Style Transfer Effects,” in *Proceedings of the 2023 9th International Conference on Communication and Information Processing (ICCIP '23)*, Lingshui, China, pp. 171-185, December 2023. <https://doi.org/10.1145/3638884.3638910>
- [41] G. H. Park, User-centered Design for Explainable AI (XAI): Focused on Smart Home Health Care Domain, Master’s Thesis, Sungshin Women’s University, Seoul, February 2023.
- [42] V. Liu and L. B. Chilton, “Design Guidelines for Prompt Engineering Text-To-Image Generative Models,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, New Orleans: LA, 384, April-May 2022. <https://doi.org/10.1145/3491102.3501825>



양나은 (Na-Eun Yang)

2023년 : 연세대학교 심리학 (학사)

2023년~현 재: 연세대학교 정보대학원 UX트랙 석사과정

※ 관심분야 : HCI(Human Computer Interaction), UX, Generative AI, Conversational AI, Explainable AI



한은빈 (Eun-Bin Han)

2023년 : 연세대학교 국어국문학과 (학사)

2024년~현 재: 연세대학교 정보대학원 UX트랙 석사과정

※ 관심분야 : UX, 생성형 AI(Generative AI), 대화형 에이전트, 메타버스(Metaverse)



최준호 (Jun-Ho Choi)

1995년 : 연세대학교 신문방송학 (석사)

2002년 : 뉴욕주립대학 버팔로 커뮤니케이션학과 (박사)

2002년~2006년: Rensselaer Polytechnic Institute. HCI 석사과정 조교수

2006년~2009년: 광운대학교 미디어영상학부 디지털미디어트랙 부교수

2009년~현 재: 연세대학교 정보대학원 UX트랙 교수

※ 관심분야 : 사용자 경험(UX), 금융, 모빌리티, AI, 로봇UX