

이미지 생성형 AI에 대한 기호학적 분석

- 기호의 자의성을 중심으로 -

문화콘텐츠학과 202127941 박정우

〈목차〉

1. 서론.....	1
2. 이론적 배경.....	2
2.1. 일반언어학 강의, 기호의 자의성	2
2.2. 이미지 생성형 AI 원리: CLIP 모델	3
3. 이미지 생성형 AI와 기호	5
3.1. 기표로서의 텍스트 프롬프트.....	5
3.2. 기의로서의 생성된 이미지.....	6
3.3. 기호로 보는 생성형 AI	6
4. 결론.....	9
5. 참고문헌.....	10

1. 서론

디지털 기술의 급격한 발전은 인간 사회 전반에 깊이 스며들고 있다. 그중에서도 ‘생성형 AI(인공지능)’는 디지털 혁신의 중심에 서 있다. 딥페이크, 영상과 이미지 및 음악 생성과 같은 생성형 AI 기술은 사회·정치 분야에서부터 문화·예술의 창작 영역에 이르기까지, 우리 일상의 전반에 걸쳐 광범위한 영향을 미친다. 이러한 생성형 AI를 작동하기 위해서는 사용자의 텍스트 입력이 우선 되어야 한다. 특히 ‘이미지 생성형 AI’는 텍스트 기반의 입력(프롬프트¹)을 통해 시각적 결과물(이미지)을 생성하는 기술로, 인간의 언어적 표현과 생성형 AI의 시각적 재현을 보여주는 대표적인 사례다. 이미지 생성형 AI가 텍스트를 이미지로 변환하는 과정에서, 생성형 AI는 동일한 입력 텍스트가 상황과 맥락 혹은 데이터셋의 확률적 선택에 따라 여러 이미지를 산출하는 특징을 보인다. 이는 《일반언어학강의》에서 말하는 언어 기호의 성격 중 ‘기호의 자의성’과 밀접하게 관련된다.

¹ 프롬프트: 생성형 AI에 입력하는 입력값을 의미.

소쉬르(Ferdinand de Saussure)는 그의 저서 《일반언어학강의》에서 기호란 기표와 기의의 조합으로 이루어진다고 설명하며, 이 둘의 관계는 필연적이지 않고 사회적 합의와 맥락에 따라 자의적으로 이루어진다고 설명한다². 이미지 생성형 AI 역시 입력된 텍스트를 학습 데이터에 기반한 확률적 모델로 해석하며, 이 과정에서 텍스트와 이미지 간의 연결은 고정적이지 않은 형태로 이루어진다. 이는 AI가 기표-기의 관계를 재구성하며 기호학적 자의성을 디지털 기술로 구현하고 있음을 보여준다고 할 수 있다. 즉 AI의 텍스트-이미지 변환 과정에서 동일한 텍스트 입력이 다양한 시각적 결과물을 생성하는 현상은, 디지털 미디어가 기호학의 원리를 실질적으로 구현하고 확장하고 있음을 시사한다.

따라서 본 연구는 이미지 생성형 AI의 작동 원리를 기호학적 관점에서 분석하며, 특히 텍스트 프롬프트와 이미지 생성 과정에서 나타나는 기호의 자의성을 탐구하는 데 목적이 있다. AI가 생성하는 다채로운 결과물이 기표와 기의의 자의성에서 비롯된다는 점을 확인함으로써, 디지털 미디어 기술이 단순한 기술적 도구이기 이전에 인문학적 원리를 내포하는 매개체로 이해될 수 있음을 논의한다. 이를 통해 AI와 같은 디지털 기술이 인문학적 관점에서 어떻게 해석되고 확장될 수 있는지에 대한 통찰을 제공하고자 한다.

2. 이론적 배경

본 연구를 진행하기에 앞서, 연구의 이론적 토대가 되는 소쉬르의 기호 이론을 기표와 기의 개념을 중심으로 정리한다. 아울러, 이미지 생성형 AI의 작동 원리를 제시하며 글의 이해를 돕고자 한다.

2.1. 일반언어학 강의, 기호의 자의성

《일반언어학 강의》의 제1부 ‘일반 원리’ 중, 제1장 ‘언어 기호의 성격’을 중점적으로 정리 및 요약하여 제시하고자 한다. 도서의 저자이자 언어학자인 소쉬르는 언어 기호가 ‘기표(signifier)’와 ‘기의(signified)’라는 두 요소로 구성된다고 보았다. 기표는 청각적 영상과 같은 감각적 표현이며, 기의는 개념적이고 추상적인 의미를 지닌다. 두 요소의 결합은 본질적으로 자의적이며, 특정 기표가 특정 기의를 지칭하는 것은 자연적 이유에서 비롯된 것이 아니라

² 페르디낭 드 소쉬르. (2012). 일반언어학 강의 (김현권 옮김). 지식을만드는지식. 138쪽.

사회적 합의에 의해 형성된 것이라고 설명한다. 또한, 한 번 확립된 기호는 개인이 임의로 변경할 수 없으며 이는 언어 체계가 집단적 규약에 따라 작동함을 보여준다.

이와 같은 기표와 기의의 관계로 인해, 하나의 기표는 다양한 기의를 가질 수 있게 된다. 이것은 여러 요인에 의해 해석된다. 먼저, 맥락 의존성이다. 기표는 문맥에 따라 서로 다른 기의를 가질 수 있다. 예를 들어 ‘은행’이라는 단어는 문맥에 따라 금융기관 ‘은행’을 지칭할 수도 있고, 나무의 종류 중 하나로 ‘은행’을 의미할 수도 있다. 또한, 문화적·사회적 차이도 기표와 기의 간의 관계에 영향을 미친다. 동일한 기표라 하더라도 사회적 규약이나 문화적 배경에 따라 그 기의가 달라질 수 있는 것이다. 예를 들어, ‘빨강’이라는 색은 서양에서는 열정이나 사랑을 내포하지만, 동아시아에서는 위험이나 행운의 상징으로 받아들여질 수 있다. 이와 함께, 기표가 지닌 본질적인 다의성으로 여러 기의를 가질 수 있는 경우도 존재한다. 예를 들어, ‘키’라는 단어는 문을 여는 도구, 사람의 신체적 키, 음악의 음계 등을 모두 지칭할 수 있다. 마지막으로, 기표와 기의의 관계는 사용자의 의도에 따라 달라질 수도 있다. ‘바람’이라는 단어는 자연현상을 의미할 수도 있고, 사람의 희망이나 욕망을 나타낼 수도 있다. 이러한 요소들은 기호 체계가 맥락적이고 유연하게 작동하며, 기표와 기의의 관계가 고정적이지 않고 자의적으로 연결되었음을 보여준다.

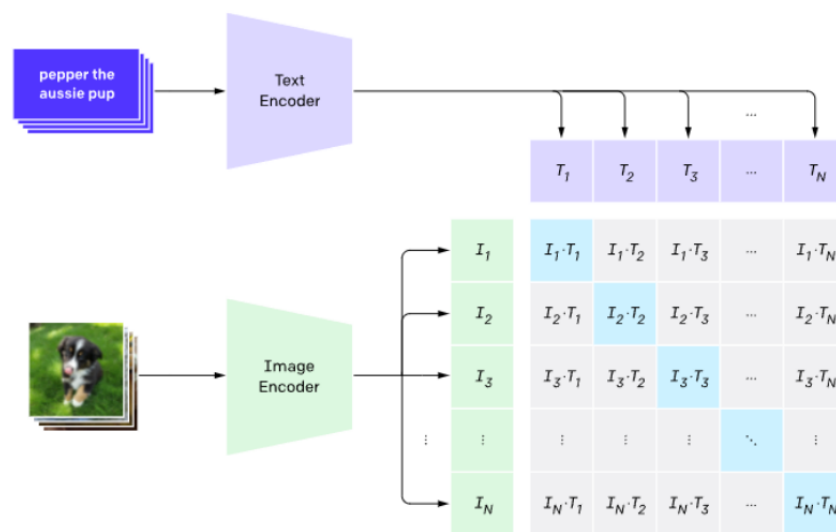
2.2. 이미지 생성형 AI 원리

이미지 생성형 AI는 텍스트 입력(프롬프트)을 기반으로 여러 이미지를 생성한다. 해당 과정에서는 GAN, VAE, 디퓨전 모델, CLIP 모델과 같은 다방면의 기술이 복합적으로 활용된다. 그중 ‘CLIP(Contrastive Language – Image Pretraining) 모델’은 텍스트와 이미지 간의 연관성을 대조 학습하여 최적의 결과물을 도출하는 데 중요한 역할을 하는 원리이다. 본 연구에서 다루는 기호학적 개념인 기표와 기의를, 생성형 AI의 작동 과정에서 각각 텍스트 프롬프트와 생성되는 이미지에 대응하여 분석하는 데 가장 직관적으로 표현할 수 있다고 판단했다. 따라서 이미지 생성형 AI의 기술 중 CLIP 모델을 중심으로 그 원리를 설명하고자 한다.

CLIP은 OpenAI에서 개발된 멀티모달 학습 모델³로, 텍스트를 시각적 결과물로 변환하는

³ 멀티모달 모델(Multimodal Model): 서로 다른 유형의 데이터(텍스트, 이미지, 음성 등)를 함께 고려하여 서로의 연관성을 학습하고 이해하도록 설계된 인공지능 모델.

이미지 생성형 AI의 작동 원리의 핵심이 된다. 기본적으로 CLIP은 텍스트와 이미지를 각각 처리하는데, 이를 컴퓨터가 이해할 수 있도록 숫자 데이터로 표현하는 텍스트 인코더와 이미지 인코더를 거친다. 텍스트 인코더는 입력된 텍스트를 수치적 벡터로 변환하여 잠재 공간에 위치시켜 언어적 의미를 나타낸다. 마찬가지로, 이미지 인코더는 이미지를 벡터로 변환해 동일한 잠재 공간에 위치시킨다. 이를 통해 CLIP은 텍스트와 이미지 간의 의미적 연관성을 학습하며, 어떤 이미지가 입력된 텍스트 프롬프트를 가장 잘 설명하는지 판별할 수 있는 능력을 갖춘다.



이미지 생성형 AI: CLIP 모델⁴

텍스트와 이미지의 임베딩⁵을 거친 후, 대조 학습(Contrastive Learning) 단계로 넘어간다. 대조 학습은 텍스트와 이미지 쌍이 맞는지 여부를 구분하며, 이를 통해 두 데이터 간의 의미적 일치도를 높인다. CLIP은 이러한 과정을 반복하며 텍스트와 시각적 데이터를 보다 정교하게 연결하는 능력을 발전시켜 최적의 결과물을 도출한다.

이와 같은 CLIP의 특징 중 하나는 텍스트 프롬프트를 다각적으로 해석해 동일한 입력에서도 색다른 이미지들을 생성한다는 점이다. 이는 CLIP이 텍스트의 언어적 의미를 다양한 잠재

⁴ Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In International Conference on Machine Learning (pp. 8748–8763). PMLR.

⁵ 임베딩(Embedding): 사람의 자연어를 기계가 이해하는 숫자의 나열인 벡터로 바꾼 결과 혹은 그 과정

공간에서 표현 가능하도록 학습한 결과다. 이를 통해, 이미지 생성형 AI는 하나의 텍스트 프롬프트에서도 다수의 창의적 결과물을 생성하게 된다.

3. 이미지 생성형 AI와 기호

CLIP 모델 기반의 이미지 생성형 AI는, 입력된 텍스트 프롬프트를 바탕으로 시각적 결과물을 생성한다. 이는 텍스트와 이미지 간의 의미적 연결을 통해 새로운 디지털 기호 체계를 형성하는 것이라 할 수 있다. 소쉬르의 기호 이론에서 기표는 감각적이고 물질적인 표현으로 특정 기의를 지칭하며, 기의는 이 기표가 지칭하는 개념적이고 추상적인 의미를 갖는다고 하였다. 이러한 기표와 기의의 관계가, 이미지 생성형 AI의 작동 과정에서 발견된다.

3.1. 기표로서의 텍스트 프롬프트

텍스트 프롬프트는 이미지 생성형 AI에서 입력값으로 사용되며, 기호학적 관점에서 기표의 역할을 수행하는 것으로 볼 수 있다. 기표는 특정 기의를 지칭하기 위한 물질적 표현으로, AI에서는 텍스트 프롬프트가 그 역할을 한다. AI는 입력된 텍스트 프롬프트를 수치화된 데이터로 변환하여 학습된 데이터와 알고리즘을 통해 이를 해석하는 과정을 거친다.

예를 들어, "A cat sitting on a chair"라는 텍스트 프롬프트는 단순히 문자적 의미로 고양이와 의자를 나타내는 것처럼 보이지만, 앞서 살핀 이미지 생성형 AI의 기술 중 CLIP 모델을 거쳐 이를 더 세부적으로 분석한다. 그렇게 텍스트 프롬프트는 AI가 학습한 데이터셋의 잠재 공간에서 벡터 표현으로 변환되며, 텍스트가 가진 언어적 의미를 수치적으로 나타낸다. CLIP 모델의 텍스트 인코더는 텍스트를 숫자 벡터로 변환하여, 이 데이터가 이미지와 연결될 수 있도록 한다.

해당 과정에서 CLIP 모델은 입력된 텍스트가 여러 가능성을 내포하고 있다고 간주하며, 동일한 텍스트 프롬프트에서도 폭넓은 시각적 해석이 가능하게 된다. 동일한 예시, "A cat sitting on a chair"라는 프롬프트는 현대적 스타일의 의자에 앉아 있는 고양이를 생성할 수도 있고 전통적 분위기의 고풍스러운 이미지를 생성할 수도 있는 것이다. 이는 기표가 맥락 의존적으로 여러 기의를 가질 수 있다는 소쉬르의 기호학적 원리와 일치한다.

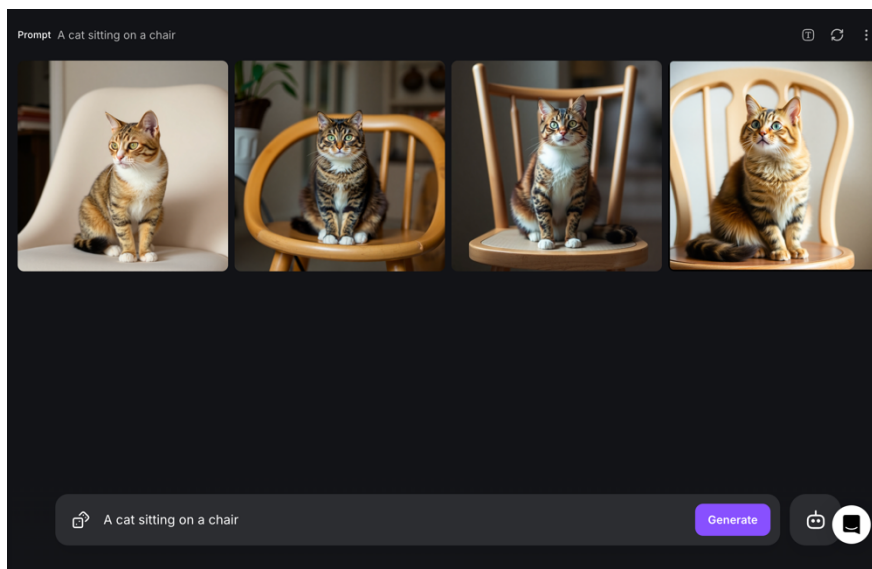
3.2. 기의로서의 생성된 이미지

텍스트 프롬프트를 입력받은 생성형 AI가 도출한 이미지는 기표의 의미를 구체화한 결과물로, 기호학적 관점에서 기의의 개념을 갖는다. 기의는 기표가 전달하려는 개념적이고 추상적인 의미를 나타내는데, CLIP은 학습된 대규모 데이터셋을 통해 텍스트(기표)와 대조되는 수많은 이미지(기의) 간의 의미적 연결을 확인하며, 텍스트가 지닌 의미를 다양한 시각적 형태로 완성한다.

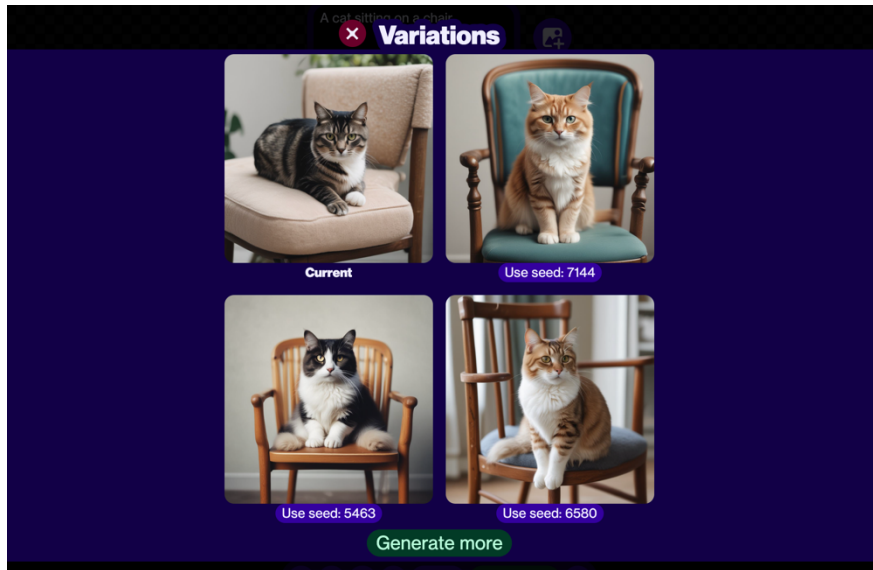
이는 CLIP의 이미지 인코더가 이미지를 벡터로 변환하여 텍스트와 동일한 잠재 공간에 위치시키고, 텍스트와 이미지 간의 일치도를 평가하는 것에서 파악할 수 있다. 예를 들어, "A cat sitting on a chair"라는 텍스트 프롬프트는 고양이와 의자의 이미지를 생성한다. 하지만, AI는 학습된 데이터와 알고리즘의 설정에 따라 각기 다른 형태와 스타일의 이미지를 만들어낸다. 이 과정은 AI가 텍스트 프롬프트의 기표적 특성을 다층적인 방식으로 해석하여 기의를 디지털 기술로 구현하는 과정을 보여준다고 할 수 있다.

3.3. 기호로 보는 생성형 AI

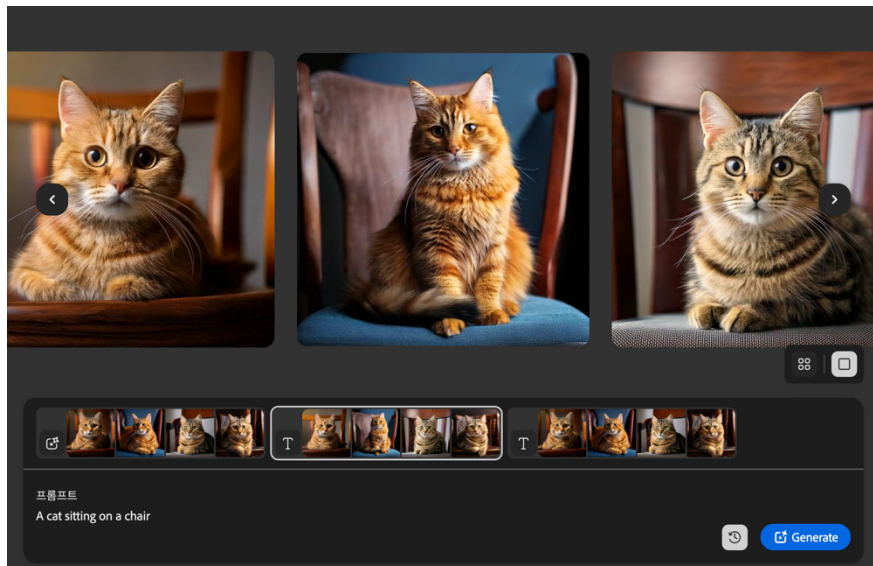
전술한 "A cat sitting on a chair"라는 예시를, 텍스트 프롬프트로 입력하여 3개의 이미지 생성형 AI 도구(Midjourney, Artbreeder, Adobe Firefly)를 통해 그 결과물을 확인해 보았다.



Midjourney



Artbreeder



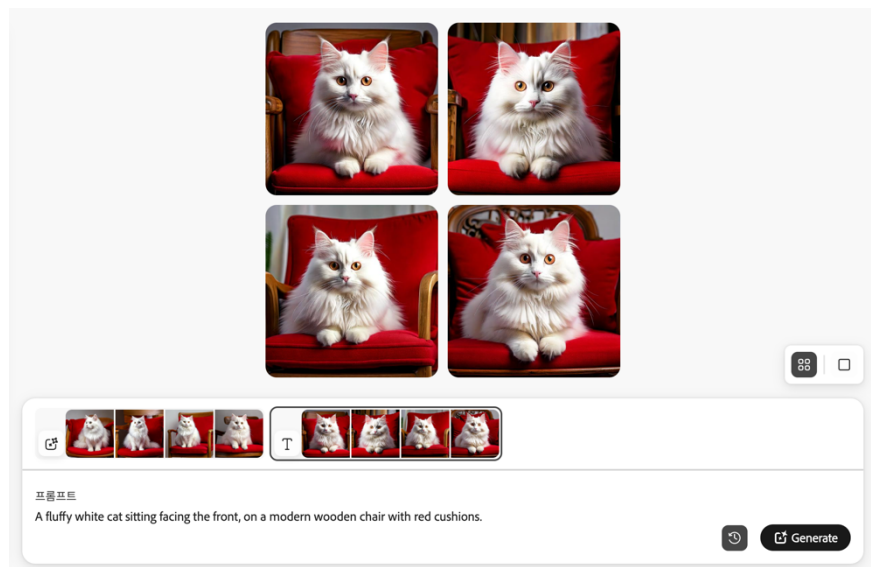
Adobe Firefly

세 가지의 이미지 생성형 AI에서 모두 동일한 텍스트 프롬프트를 입력했지만, 각기 다른 이미지가 다양하게 도출된 것을 알 수 있다. 프롬프트에 사용된 "A cat", "sitting", "a chair"라는 단어들은 AI가 학습한 데이터와 알고리즘적 해석에 따라 다양한 형태로 변형된다. "A chair"는 단순히 의자를 의미하지 않고, 목재, 금속, 현대적 디자인, 혹은 전통적인 의자 등 여러 스타일로 묘사될 수 있을 것이다. "A cat sitting" 또한 고양이의 생김새부터, 그의 자세, 방향, 또는 배경 맥락에 따라 다른 방식으로 표현될 가능성을 가진다. 이러한 결과는 언어적 기표가 특정 기의를 반드시 고정적으로 지칭하지 않는 기호학적 자의성을 반영한다.

이와 같은 현상은 사람 간의 의사소통, 즉 언어에서 보이는 특성과도 연관된다. 인간 사회

에서 언어로 특정 기표를 통해 의사를 전달할 때, 상대방이 떠올리는 기의는 화자가 의도한 것과 다를 수 있다. "의자"라는 기표를 사용하면 상대는 자신이 경험한 의자를 기반으로 다양한 기의를 떠올릴 수 있다. 화자가 생각하는 기의를 상대가 정확히 이해하려면, 단순히 "의자"라는 단어만 제시하는 것이 아니라, 그 의자의 특성이나 맥락에 대한 부연적인 설명을 덧붙이는 것이 필요하다. 이러한 방식으로 의사소통은 더 명확하고 매끄럽게 이루어진다.

이미지 생성형 AI와의 상호작용에서도 동일한 원리가 적용된다. 사용자가 생성형 AI에 텍스트 프롬프트를 입력하여 자신이 생각하는 특정 기의를 얻어내기 위해서는, 그에 대응하는 기표를 상세하고 구체적으로 작성해야 한다. 단순히 "A cat sitting on a chair"라는 문구를 입력하기보다는, "A fluffy white cat sitting facing the front, on a modern wooden chair with red cushions"처럼 상세한 맥락과 특성을 덧붙인 프롬프트를 입력하는 것이다.



Adobe Firefly

이로써 본인이 의도한 결과를 도출할 가능성이 높아진다. 결국, 인간의 언어적 소통에서 기표와 기의의 관계를 이해하고 명확히 하고자 하는 노력이, 이미지 생성형 AI와 상호작용하는 과정에서도 드러난다. 생성형 AI를 이용하는 사용자는 자신이 원하는 결과물을 얻기 위해, AI와의 소통을 인간과의 언어적 소통과 유사하게 접근하는 것이다. 이러한 점을 고려했을 때, 이미지 생성형 AI의 운용과 그 원리의 이해는 단순한 기술적 활용에서 그치지 않는다. 이는 기호학적 원리에 기반한 인문학적 해석과 실천을 요구하며, 디지털 기술을 더 효과적이고 창의적인 수행을 가능케 한다.

4. 결론

본 연구에서는 이미지 생성형 AI의 작동에서, 텍스트 프롬프트와 이미지 생성을 연계하는 과정을 소쉬르의 언어학적 기호로 전제하고 분석했다. 텍스트 프롬프트와 생성된 이미지의 관계를 기표와 기의 간의 연결점에서 해석하며, 이미지 생성형 AI가 다양한 결과물을 도출하는 과정이 기호학적 자의성에 기반하여 작용되고 있음을 밝혔다. 이미지 생성형 AI의 작동 방식이 인간 언어 체계의 기호적 특성과 유사한 방식으로 설명될 수 있음을 조명했다. 이로써, 인간과 AI 간의 소통 메커니즘을 기호학적으로 풀어냈다는 점에서 의미 있는 접근을 시도했다고 할 수 있다. 결론적으로, 디지털 기술을 단순한 도구로 바라보는 데서 나아가 이를 인문학적 이해와 해석의 대상으로 제시했다는 점에서 의의가 있다.

그러나 본 연구에서는 이미지 생성형 AI에서도 CLIP 모델에 한정하여 분석했기 때문에, 생성형 인공지능 전반에 대한 일반화는 어렵다. 더불어, CLIP 모델의 수식적 원리와 알고리즘의 세부적인 분석까지 미치지 못하여 기술을 상세히 논의하지 못한 점에 한계가 있다. 끝으로, 본 연구가 디지털 기술에 대한 인문학적 접근을 확장하고 기술과 인간의 관계를 더욱 심층적으로 이해하기 위한 후속 연구의 토대를 제공하길 기대한다.

5. 참고문헌

- [1] 김현옥. (2024). 해(解): 기호와 사유 강의노트. 아주대학교.
- [2] 박지호. (2021, August 9). Weekly NLP #42: Text-to-image AI models. Jiho-ML. <https://jiho-ml.com/weekly-nlp-42/>.
- [3] 박한하, & 한혜원. (2024). 생성형 인공지능의 프롬프트 글쓰기 분석. *한국콘텐츠학회 논문지*, 24(4), 59 – 67.
- [4] 유이경. (2022, May 20). CLIP: Connecting text and images. 고려대학교 DMQA Open Seminar. <http://dmqm.korea.ac.kr/activity/seminar/365>.
- [5] 페르디낭 드 소쉬르. (2012). *일반언어학 강의* (김현권 옮김). 지식을만드는지식.
- [6] Candywave. (2022, October 3). *AI 그림 두번째: 10분컷 AI 이미지 생성의 원리 - Deep learning은 들어봤는데... 달리? 미드저니?*. YouTube. <https://www.youtube.com/watch?v=LkrFw6OFptc&t=10s>.
- [7] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In International Conference on Machine Learning (pp. 8748–8763). PMLR.