JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 25, No. 2, pp. 341-354, Feb. 2024

# 이미지 생성 AI의 프롬프트 요소와 적용 사례 연구: 미드저니, 스테이블 디퓨전, 파이어플라이, 달·이를 중심으로

권 동 현*
*경기대학교 애니메이션학과 초빙교수

# Analysis of Prompt Elements and Use Cases in Image-Generating AI: Focusing on Midjourney, Stable Diffusion, Firefly, DALL·E

**Dong-Hyun Kwon***
*Professor, Department of Animation, Kyonggi University, Seoul 03746, Korea

**[요 약]**

오늘날, 전 세계는 이미지 생성 AI에 주목하면서, 다양한 분야에 적용하고자 노력함과 동시에 윤리 문제, 저작권 문제, 직업 위기에 대응하면서, 알 수 없는 매개변수들로 인한 위험성도 걱정하고 있다. 그럼에도 생성 AI는 기존 디지털 도구들을 능가하는 유용성과 효율을 보일 것으로 예상되므로 창작자들은 AI와 협업할 수 있는 소양과 함께 프롬프트에 관한 기본 지식을 갖추어야 할 것이다. 이를 위하여 본 연구는 이미지 생성 AI의 프롬프트를 인식과 기술 측면으로 나누고 3단계로 구분하여 요소들을 알아보았으며 미드저니, 스테이블 디퓨전, 파이어플라이, 달·이 에 적용할 때 주의할 점들과 이미지가 어떻게 나타나는지 사례를 정리하였다. 또한 가장 최신 서비스인 달·이 3의 프롬프트 생성 기능에서 AI가 인간의 프롬프트를 이해하고 생성한 프롬프트와 본 연구의 구성요소를 비교하여 확인하였다. 프롬프트는 멀티모달로 구현하는 인공지능을 가장 쉽게 사용할 수 있는 인터페이스이므로 이미지 창작자들도 사고의 훈련과 글쓰기 역량을 키워 앞으로 등장할 다양한 생성형 AI를 잘 활용하여야 할 것이다.

**[Abstract]**

Today, the world is focusing on artificial intelligence (AI)-based image generation, striving to apply it across various fields while simultaneously dealing with ethical issues, copyright issues, and job crises, in addition to worrying about the risks posed by unknown parameters. Despite these challenges, generative AI is expected to surpass the usefulness and efficiency of existing digital tools; therefore, creators must equip themselves with the competence to collaborate with AI as well as attain a basic understanding of prompts. To facilitate this requirement, this paper divides the prompts in image-generative AI based on their cognitive and technical aspects and categorizes them into three stages to explore their elements. It also presents cases based on the different aspects that must be considered and the variations in images when applied to four state-of-the-art image generation models: Midjourney, Stable Diffusion, Firefly, and DALL·E. Moreover, it examines how AI understands human prompts and compares the generated prompts with the components of this study, particularly, with the latest generative model, DALL·E 3. As prompts are the easiest interface for utilizing multimodal AI, image creators must also develop their thinking and writing skills to effectively use the various generative AI systems that will emerge in the future.

**색인어 :** 창작자를 위한 AI 교육, 멀티모달 인터페이스, 프롬프트 기술, 프롬프트 인식, AI 협업
**Keyword :** AI Education for Creators, Multimodal Interface, Prompt Engineering, Prompt Recognition, AI Collaboration

**http://dx.doi.org/10.9728/dcs.2024.25.2.341**

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(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** 21 January 2024; **Revised** 02 February 2024
**Accepted** 13 February 2024

***Corresponding Author; Dong-Hyun Kwon**
**Tel:** ※ 개인정보 표시제한
**E-mail:** msz009@naver.com

## Ⅰ. 서 론

### 1-1 연구 배경 및 목적

2022년 11월 출시된 챗GPT 서비스는 컴퓨터가 아닌 마치 사람과 대화하는 것 같은 경험으로 이전과 비교할 수 없는 관심을 불러왔으며 지금도 놀라운 속도로 사용자가 증가하고 있다[1]. 이러한 생성 AI는 텍스트, 이미지, 코딩, 음성, 영상, 사운드, 3D 등 다양한 분야로 확장하고 있으며[2] 특히 이미지 생성 AI는 인간이 그린 것보다 뛰어난 기법의 그림은 물론 사진과 같은 높은 수준의 이미지를 생성하여 사회적, 경제적 큰 영향을 미치고 있다. 또한 서비스 개수 면에서도 수많은 생명이 폭발적으로 나타난 캄브리아기로 비교될 만큼 출시되고 있으며 기술 발전 속도도 매우 빠르다[3].

생성 AI의 등장은 사용자와 창작자 모두에게 중요한 변화를 불러오고 있으며, 제작 방식의 변화뿐만 아니라 창작의 본질을 재정립하는 새로운 패러다임을 요구하고 있다. 이미 산업계에서는 많은 변화가 일어나고 있으며, 현재 교육계에서는 영향이 크지 않지만, 앞으로의 변화를 예측하고 적극적으로 교육에 도입해야 할 필요가 있기에 연구자는 이미지, 영상 창작자들이 생성 AI를 경쟁 상대로 보지 않고 유용한 도구이자 파트너로 함께 할 수 있도록 이미지 생성 AI의 원리를 사용자 수준에서 이해하고 교육 현장에 적용하기 위한 선행연구를 진행하였다[4].

연구자는 선행연구에서 더 나아가 생성형 AI를 활용하기 위한 첫 단계이자 텍스트 인터페이스의 역할을 하는 '프롬프트'의 구성요소를 정리하고 이미지 생성 AI에 적용하는 사례와 함께 23년 10월 출시한 최신 기술인 챗GPT 4의 이미지 생성 AI 프롬프트 작성 기능을 활용하여 인간이 작성한 프롬프트를 AI가 어떻게 이해하는지도 알아보고자 후속 연구로서 본 연구를 진행하였다. 본 연구 내용이 창작자들이 AI 협업을 하기 위한 기본 소양을 갖추는 데 도움이 되고 프롬프트 기본 지식을 교육할 수 있는 자료로서 역할을 할 수 있기를 기대한다.

### 1-2 선행 연구 및 연구 방법

AI 기술의 빠른 발전 속도를 고려하여 선행연구 이후 프롬프트와 관련한 새로운 연구를 조사하기 위하여 ‘AI, 인공지능, 프롬프트’라는 키워드로 검색해 본 결과 총 38편이 검색되었다. 게재 시기를 살펴보면 학위 논문 1편을 제외하고 모두 2023년에 게재되었으며, 이는 챗GPT 출시일이 2022년 11월인 것을 고려해 볼 때 인공지능과 관련한 연구 중 ‘프롬프트’ 키워드가 포함된 연구가 개발자 측면이 아니라 사용자 측면에서 연구되는 대상임을 유추해 볼 수 있다.

검색된 38편의 연구들은 대부분 챗GPT에 관한 내용이거나 프롬프트 자체를 직접 연구한 내용은 아니었으며 본 연구의 목적인 이미지 창작자의 관점에서 프롬프트를 사용하는 방안에 관한 연구 내용에 부합하는 것으로 총 6편을 검토하였다. 이중 이명주, 최에스더의 네일아트 이미지 생성 프롬프트 연구[5], 박서유, 김선철의 아이디어 전개 과정에서의 이미지 생성 AI 활용 만족도 연구[6], 이수환, 송기상의 교사가 입력한 프롬프트로 생성된 이미지의 수업 자료 가능성 2편의 연구[7],[8], 이렇게 총 4편은 선행연구에서 이미 분석한 것으로 생성 AI의 기능을 연구하면서 프롬프트를 언급하지만, 이미지 창작자가 필요로 하는 내용은 아니었다. 새롭게 검색된 연구는 1편으로, 한홍비는 원하는 이미지를 생성하기 위하여 텍스트 프롬프트와 함께 추가 이미지를 학습하는 인스턴스 프롬프트를 적용한 과정을 연구하였지만[9], 텍스트 프롬프트 내용 자체를 직접 연구한 것은 아니었다. 연구자의 선행연구[10]는 애니메이션 제작 교육에 가장 적합한 이미지 생성 AI를 선정하는 연구로서 먼저 사용자 수준에서 이해할 수 있도록 발전 과정과 구현 원리를 설명하고 저작권, 직업 선택권, 윤리 문제를 언급하였다. 그리고 프롬프트 교육을 위한 글쓰기의 중요성과 함께 앞으로 이미지 창작자는 그리는 행위보다 AI가 생성한 이미지 선택의 과정이 중요하므로 이를 ‘창의 선택’이라고 명명하였다. 마지막으로 생성 AI의 결과물과 사용 환경, 비용, 접속 방식 등을 고려하여 어도비 파이어플라이를 교육 현장에 가장 적절한 이미지 생성 AI로 선정하였다.

본 연구는 선행연구에서 정리한 프롬프트 교육의 당위성과 포괄적인 프롬프트 교육 방법론에서 더 나아가 실제 프롬프트의 구성요소를 감각적, 기술적 측면으로 나누고 3단계로 구분하여 정리하고자 한다. 그리고 중요한 이미지 생성 AI 서비스 4종류를 선정하고 프롬프트를 사용하는 방법을 설명한 후 프롬프트를 실제 적용한 사례를 정리할 것이다. 마지막으로 23년 10월부터 챗GPT 4와 DALL·E(이하 달·이) 3이 통합되면서 AI가 프롬프트를 작성할 수 있게 되었으므로 사람이 입력한 프롬프트를 챗GPT 4가 이해하고 바꾼 프롬프트를 살펴봄으로써 연구자의 프롬프트 구성요소와 비교해 보고자 한다.

## Ⅱ. 프롬프트 연구

### 2-1 멀티모달과 텍스트 명령의 의미

생성 AI는 대규모 학습데이터가 필요하며 이를 처리하고, 패턴을 구조화하고, 예측하고 결정을 내리는 알고리즘을 모델(Model)이라고 한다. 일반적으로 널리 알려진 딥러닝도 최신 모델 중 하나다. 그리고 비슷한 단어로서 모달(Modal)이 있는데 이는 시각, 청각, 미각 등 각각의 감각을 뜻하는 Modality의 줄임말로써 멀티모달이란 여러 개의 감각이 상호작용하는 것을 의미한다. 이미지 생성 AI는 대량의 이미지와 함께 이를 설명한 텍스트도 함께 학습한다. 이렇게 텍스트와 이미지를 동시에 학습하는 딥러닝 모델을 멀티모달 모델이라 하며 이미지, 텍스트뿐만 아니라 소리의 파형과 텍스트를 함께 학습하는 음성인식 모델도 멀티모달 모델이라 할 수 있다. 챗GPT를 공개한 오픈AI는 텍스트에 대한 방대한 학습데이터

를 사용하여 텍스트로부터 이미지를 생성하는 멀티모달을 만들었으며 이미지 생성 AI 서비스인 달·이를 가장 먼저 출시할 수 있었다[10].

모든 감각과 텍스트를 연결하는 멀티모달 모델은 또 다른 의미가 있다. 컴퓨터 발전 과정에서의 큰 이슈를 살펴보면, 거대한 연구실 컴퓨터에서 개인용 컴퓨터로, 텍스트 입력에서 그래픽 인터페이스 마우스 입력으로, 개별 공간에서 인터넷을 통한 전 세계 연결로, 누구나 어디서든지 손가락만으로 쉽게 접근할 수 있는 모바일 컴퓨터로 발전하였다. 이 과정은 접근이 쉬운 하드웨어와 그것을 제어하는 인터페이스 혁신의 연속이었으며 인간의 언어를 이해하는 멀티모달은 인터페이스 혁신의 최종 형태라 할 수 있다. 예를 들어 복잡한 프로그램 코딩은 전문가의 영역이었으나 앞으로는 말로 명령하여 코딩할 수 있는 시대가 올 것이다. 자연어는 인간에게 가장 친숙한 인터페이스로서 멀티모달은 누구나 특별한 교육을 받지 않아도 높은 수준의 컴퓨터 제어가 가능한 시대로 이끄는 핵심기술이라 할 수 있다.

### 2-2 생각을 정리하고 텍스트로 설명하기

황요안은 생성 AI 시대에 프롬프트의 중요성을 인식하고 프롬프트 리터러시가 필요한 시대가 도래할 것이라 연구하였다. 그리고 프롬프트 작성에 필요한 소양으로, PROMPT의 앞 글자로 6가지를 제안하였는데 문제해결을 위한 개척자(Problem- Solving Pioneer), 탄력적 자원 동원가(Resilient Resourcer), 개방적 기획자(Open-Minded Organizer), 다분야적 사고가(Multi disciplinary Thinker), 개인화된 정보 분별사(Personalized scriminator), 시대의 흐름을 읽은 인재(Trend Seeker)로 구분하였다[11]. 이유미는 AI 리터러시를 연구하면서 가르치고 배우는 고정적인 역할이 사라진 가치 혼란의 시대에 개인이 가져야 하는 능력은 스스로 학습할 수

**표 1.** 프롬프트 구성요소와 프롬프트 예제
**Table 1.** Prompt components and prompt examples

| Classification | | | Example of a Prompt | |
|---|---|---|---|---|
| 1 | 2 | 3 | | |
| perceptual aspects | story composition | person | appearance | gender, posture, tall, short, slim, muscular, curvy, brunette, bald, tanned, pale, freckled, wrinkled, youthful, athletic, chubby, bearded, clea5n-shaven, pierced, tattooed, eye color, skin tone, hair color, upper & lower clothing style, number of people… |
| | | | accessories | necklace, earrings, bracelet, ring, watch, tie, scarf, sunglasses, belt, hat, airband, glove, brooch, cuff links, anklet, nose ring, bangle, lapel pin, tiara, arm band, bowtie, choker, eyeglasses, ear cuff, hijab… |
| | | space | indoors, outdoors, urban, rural, forest, desert, lake, mountain, ocean, beach, cityscape, countryside, park, island, town, waterfall, tunnel, skyscraper, street, alley, bridge, valley, cave, garden… | |
| | | environment | | nature, time, region, weather, season, ecosystem, habitat, landscape, mountainous, geology, flora, atmosphere, temperature, humidity, rainforest, coastal, tropical… |
| | concept | image style | famous style | famous director, writer, movie, magazine name / steampunk, dieselpunk, cyberpunk, biopunk, solar punk, atompunk, raypunk, nanopunk, decopunk, rococopunk, mythpunk, spacepunk, elfpunk, now punk, splatterpunk, new weird… |
| | | | movement | famous artist nsme in history / gothic, renaissance, deco, surrealism, romanesque, impressionism, rococo, expressionism, cubism, baroque, realism, bauhaus, neo-classicism, art nouveau, dadaism, art fauvism, futurism, constructivism, modernism… |
| | | color | vibrant, muted, bright, monochromatic, old film, matte, look,tinted, colorful, black and white, pastel, vivid, pastel tone, high contrast, sepia, desaturated, neon, duotone, saturated, warm, cool tone, retro, bleached bypass… | |
| | | atmosphere | dreamy, sedate, calm, raucous, energetic, serene, vib rant, tranquil, lively, mellow, chaotic, peaceful, tense, jubilant, somber, melancholic, exuberant, intense, relax ing, festive, gloomy, mystical, whimsical, soothing… | |
| technology aspects | image composition | layout | extreme close-up, close-up, medium close-upbust shot, head shot, knee shot, medium shot, waist shot, full shot, long shot, extreme long shot, two shot, three shot, low high angle, over-the-shoulder, bird's eye view, portrait… | |
| | | camera | lens caliber / standard, wide angle, telephoto, zoom, macro, fish-eye, tilt-shift, mirror, ultra-wide, long-focus, short-focus, aspherical, portrait, cine, anamorphic, soft focus, super telephoto lens, isometric, bird view … | |
| | | light | fill, top, practical, hard, soft, ambient, pin, natural, flood, motivated, bounce, diffused, spot, butterfly, rim, volume, silhouette, day for night, fairy lighting, cinematic lighting, soft, overcast, neon, studio… | |
| | image format | texture | 2d drawing, 3d rendering, pixel art, stained glass, mural, oil painting, watercolor, acrylic, charcoal, pastel, collage, gouache, fresco, graffiti, sketch, engraving, ink drawing, woodcut, lithography, embossing, etching, mosaic, air brush, tapestry, sculpture, doodle, photograph… | |
| | | ratio | | different inputs depending on service type |
| | | quality up | | high resolution, realistic shadow, high details, realistic lighting … |

있는 능력과 이를 통해 미래를 예측해 낼 수 있는 능력이 필요함을 강조하였다[12]. 이를 간추려 보면 충분히 열린 사고로 세상의 다양한 정보들을 찾아 분석하고 그것을 자기만의 관점으로 종합할 수 있어야 함을 설명하고 있다. 학생들이 스스로 좋아하고 만들고 싶은 자기만의 세계관을 분명히 하고 그것을 표현할 수 있도록 지도하는 것은 AI가 없던 이전 시대에서도 필요한 내용이다. 하지만 프롬프트로 다양한 생성 AI를 잘 사용해야 하는 오늘날에는 텍스트 작성 능력이 더욱 중요한 소양이 되었다고 할 수 있다.

창작자로서 작품을 만드는 과정을 돌이켜 볼 때 직관적 생각 가운데 아이디어를 떠올렸다고 할지라도 의식의 흐름을 유추해 보면 어떠한 외부적 요인이나 내부적 동기에 의해 생기는 감정이나 느낌을 그동안의 경험과 기억, 그리고 학습한 정보를 바탕으로 다양한 제작 기법을 고려하여 도출해 내는 과정이라 할 수 있다. 그러한 생각의 흐름 가운데 있는 단어를 나열하고 그중에서 중요한 핵심 단어들을 선별하는 과정의 결과물을 프롬프트라 하겠다.

### 2-3 프롬프트 구성요소

연구자는 그동안 이미지 생성 AI를 직접 사용한 경험과 함께 생성 AI 작가로 활동 중인 KEEPKWAN(킵콴)[13]의 작업을 연구하여 프롬프트 구성요소를 표 1과 같이 1단계로 인식과 기술 측면으로 나눈 다음 2단계, 3단계로 세분화하여 정리하였다.

다양한 프롬프트를 작성하고 적용해 본 결과 최대한 구체적일수록 정확한 이미지를 얻을 수 있었으며 표 1과 같이 프롬프트를 단계별 개념으로 이해하고 적었을 때 비교적 시행착오를 줄이면서 원하는 결과를 얻을 수 있었다. 이러한 구분은 사용자의 취향에 따라 다른 기준으로 설정할 수도 있으며 연구자도 지각 측면과 기술 측면으로 1단계를 나누었지만, 사용할 때는 지각 측면의 색감과 기술 측면의 조명을 붙이거나 지각 측면의 미술 사조와 기술 측면의 질감을 붙이기도 하는 등 연관 있는 단어들을 나열하는 방식으로 사용한다. 지각 측면의 2단계 이야기 구성요소로 사람, 공간, 환경으로 나누었지만, 개인의 경험과 상상력에 따라 예시의 내용과 다른 개념적 구분도 가능할 수 있으며 최대한 구체적으로 설명할 수 있는 방법론으로 개념을 찾아가야 할 것이다. 통상적으로 먼저 입력한 텍스트가 중요하지만, 꼭 그 순서를 따르지는 않는다. 성능이 좋은 이미지 생성 AI는 챗GPT 수준의 텍스트 인식 능력이 있으므로 문장으로 입력할 수 있으며 은유적인 묘사가 들어간 시를 프롬프트에 적으면 가끔 인간의 상상을 넘어서는 흥미로운 이미지를 생성하기도 한다. 연구자의 구분 중 기술 측면의 마지막에 있는 퀄러티 향상 프롬프트는 이미지 생성 AI 종류에 따라 사용하지 않아도 디테일이 높은 이미지를 기본적으로 생성하기도 한다.

### 2-4 테스트 프롬프트 작성

본 연구를 위하여 표 2와 같이 3종류의 프롬프트를 작성하였다. 이후 본문에서는 일반 명사로서의 프롬프트와 구별하여 이미지 생성에 쓰인 명령이라는 뜻으로의 프롬프트는 영문으로 하였다. Prompt 1은 3D 일러스트와 아이소메트릭 뷰라는 형식만 설정하고 내용 면에서는 인간의 삶이라는 많은 의미를 지니면서도 세부적인 설명을 생략하여 생성 AI의 자율성과 창의성을 알아보고자 하였다. Prompt 2는 좀 더 자세한 설명을 더 하여서 인상파 화가의 스타일과 화려한 색, 일반적이지 않은 형태의 집이라는 내용과 인상파 화가 스타일이라는 형식을 지정하여 생성되는 이미지를 살펴보고자 한다. Prompt 3은 앞선 Prompt 들과 다르게 표 1에서 분석한 구성요소를 적용하여 프롬프트를 작성하였다. 시간, 장소, 색감, 이미지 스타일, 카메라 구도, 조명, 건물 디자인과 배경 설명 등을 단어로 나열하였으며 서비스별로 어떤 이미지를 생성하는지 비교해 보고자 한다. 모두 이미지 비율은 지정하지 않았으므로 1대1로 생성된다.

**표 2.** 테스트 프롬프트 예제
**Table 2.** Test prompt sample

| | |
|---|---|
| Prompt 1 | Isometric, 3D illustration, human life |
| Prompt 2 | Impressionist Painter's painting, colorful and unusually shaped house |
| Prompt 3 | outdoors, night, structure in a huge rock mountain, gothic structure, H.R. Giger style, extraterrestrial civilization, complex structures, colorful nebula sky, pale colored buildings, wide angle, low angle, blue volume light glowing from the floor, high resolution, high details, realistic photography, realistic lighting |

## Ⅲ. 이미지 생성 AI 선정 및 프롬프트 적용

현재 생성 AI에 대한 사회의 관심은 매우 크지만, 업계의 상황은 새로운 수익원을 탐색하는 초기 단계다. 연구자는 24년 1월 현재 출시된 서비스들을 최대한 나열하고 연구에 적합한 생성 AI를 선별하고자 하였으나 산업 규모나 응용 분야와 같은 통계 자료가 없어 객관적 지표를 구할 수 없었다. 또한 지금도 이미지 생성 AI는 다양한 분야에 특화된 서비스들이 계속 나타나고 있으며 앞으로도 기술이 발전하면서 더욱 다양한 형태의 이미지 생성 AI가 출시될 것으로 예상된다. 그러므로 본 연구에서 검토할 대상으로서 선정한 기준은 공개된 정보들을 바탕으로 역사적, 기술적, 산업적 측면과 함께 대규모 자본을 배경으로 하거나 빠른 기술개발로 시장을 선점하여 앞으로도 지속적인 서비스가 가능할 것으로 예상되는 서비스들이다. 이 기준에 맞추어 연구자는 다음과 같이 서비스 4종류를 선정하였으며 각각의 서비스 특성과 앞서 연구한 프롬프트를 적용할 때의 여러 사례를 연구하고자 한다.

### 3-1 미드저니(Midjourney)

#### 1) 선정 이유

이미지 생성 AI가 본격화될 당시 저작권 문제로 주목받았던 스페이스 오페라 극장(Theatre D`opera Spatial, 2022)과 새벽의 자라야(Zarya of the Dawn, 2022) 웹툰 이미지 그림 1을 생성한 AI로써 프롬프트 사용에 익숙하지 않아도 높은 수준의 이미지를 생성할 수 있다. 모든 이미지 생성 AI는 전 세계에 몇 개 없는 대규모의 텍스트 라벨링을 마친 이미지 데이터를 학습하는 과정을 거치는데 그럼에도 초보 사용자의 프롬프트에도 다른 이미지 생성 AI보다 높은 수준의 이미지를 생성하는 것은 데이터 학습 결과인 매개변수를 효과적으로 미세 조정했기 때문이다. 누구나 쉽게 높은 수준의 이미지를 생성하므로 전 세계 많은 사용자가 있으며 그만큼 사용과 관련한 공유지식도 쉽게 찾을 수 있다. 이에 비교 대상으로 선정하였다.

**그림 1.** 새벽의 자라야, 2022 / 스페이스 오페라 극장, 2022
**Fig. 1.** Zarya of the Dawn, 2022 / Theatre D`opera Spatial, 2022

#### 2) 특징

게시판 형식의 채팅 플랫폼인 디스코드에서 명령을 내리고 이미지를 생성하는 인터페이스를 사용하므로 디스코드 가입과 채팅 방식이라는 불편함을 감수해야만 한다. 채팅으로 내린 명령으로 생성되는 이미지는 다음 이미지에 밀려 계속 위로 올라가므로 이전 그림을 찾기 위해 위아래로 움직이며 게시글을 찾아야 한다. 디스코드는 게임을 이용하면서 함께 사용하는 텍스트, 음성 채팅 플랫폼으로써 게임 사용자들에겐 익숙한 곳이다. 또한 디스코드에는 채팅봇이라는 부가 서비스를 개발하여 운영할 수도 있는데 미드저니 개발자들은 게임과 채팅이라는 익숙한 디스코드에서 최초 개발 협의와 테스트를 진행하였고 결국 최종 서비스까지 채팅봇 서비스 형태로 디스코드에서 운영하기로 하여 그림 3과 같은 게시판 형식의 인터페이스를 가지게 되었다. 디스코드를 처음 접하는 사용자는 채팅 명령형식의 결제 과정부터 프롬프트 입력법, 자기 서버 구성 등 구체적인 사용법 모두 새롭게 배우는 불편함을 거쳐야 하지만, 프롬프트에 익숙하지 않은 초보자들도 우수한 퀄러티의 이미지를 생성할 수 있어 사용자들이 꾸준히 늘고 있다. 사용료는 서비스 초기에는 생성 이미지를 경험해 볼 수 있도록 무료 횟수를 부여하였는데 현재는 무료로 사용할 수는 없으며 개인은 월 30달러를 결제해야 횟수에 상관없이 편하게 사용할 수 있다.

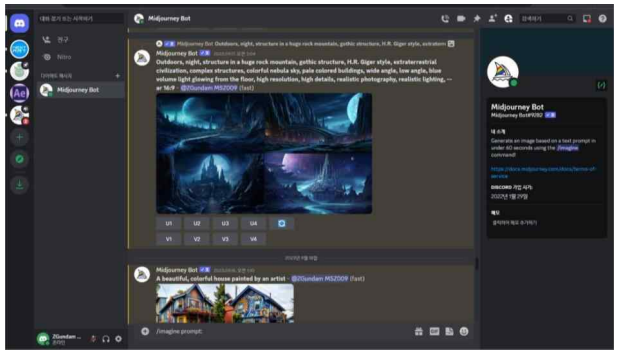

**그림 2.** 디스코드에서 실행되는 미드저니 게시판 형식 인터페이스
**Fig. 2.** Midjourney bulletin board-style interface running on Discord

#### 3) 인터페이스와 프롬프트 사용법

프롬프트 뒷부분에 '--'와 함께 추가적인 파라미터를 사용할 수 있다. 예를 들어 '--ar 16:9'는 16대9의 화면비율로 이미지를 생성하라는 의미이며 '--w 1920 --h 1080'과 같이 직접 생성 이미지 해상도를 지정할 수 있다. '--seed 숫자'로 생성 시드값을 지정하는 등 다양한 옵션을 사용할 수 있으나 대부분 전문적인 영역으로서 가로세로 비율 지정 외에는 잘 사용하지 않는다. 또한 '--no' 프롬프트는 특정 요소를 제거할 때 사용한다. 그러나 똑같은 프롬프트에서 하늘의 행성을 제거하고자 '--no planet' 옵션을 사용한 결과, 그림 3과 같은 전혀 다른 이미지가 생성되었음을 볼 때 '--no' 프롬프트는 사용에 주의해야 한다.

**그림 3.** 미드저니 '--no' 프롬프트 사용 시 생성 이미지 차이
**Fig. 3.** Differences when using Midjourney '--no' prompt.

디스코드에서 게시판 형식으로 운영되는 미드저니 인터페이스의 특성상 버튼이나 메뉴를 따로 둘 수 없으므로 이미지가 생성되면 그림 4에서와 같이 아래에 제어 아이콘이 생긴다. 기본적으로 하나의 프롬프트에 4개의 이미지를 생성하며 그중 한 개의 이미지를 선택하고 그것을 바탕으로 다시 4장의 이미지를 생성하거나 생성 이미지에서 확장하는 명령, 생성된 이미지를 확대하거나 내려받는 것 등, 아이콘을 선택해서 필수 기능 명령을 내릴 수 있다.

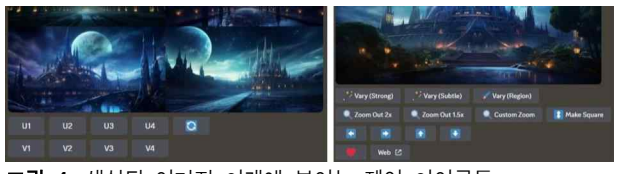


**그림 4.** 생성된 이미지 아래에 보이는 제어 아이콘들
**Fig. 4.** Control icons displayed below the generated image

#### 4) 테스트 프롬프트 적용

앞서 프롬프트 구성요소를 정리한 후 생성 이미지를 테스트하기 위해 만든 표 2의 Prompt 3종류를 미드저니에 적용하였을 때 생성된 이미지는 그림 5와 같다. 한 번에 4개의 이미지를 생성하지만, 제한된 지면 속에서 적절한 이미지 크기를 맞추기 위해 3개만 나열하였다.

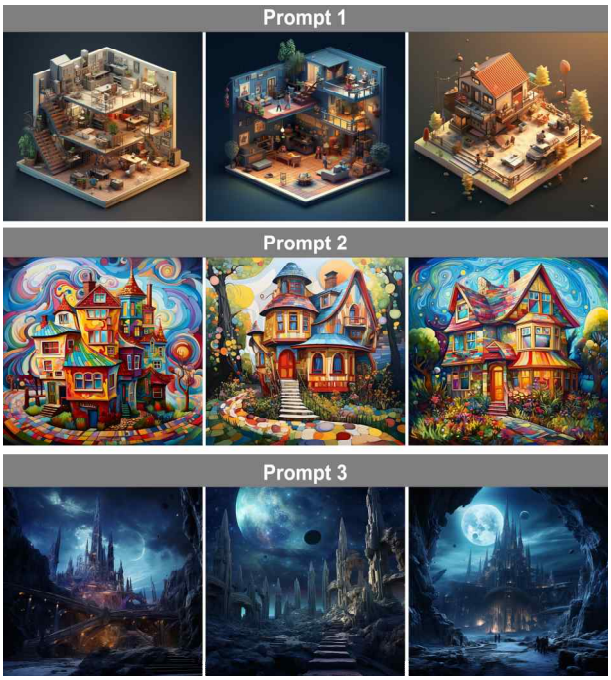


**그림 5.** 테스트 프롬프트를 적용한 미드저니 이미지 생성 결과
**Fig. 5.** Test prompt results of Midjourney image generation

미드저니는 연구자의 선행연구에 이어 본 연구에서도 프롬프트의 구성요소에 맞는 단어 유무에 상관없이 안정되고 디테일이 좋은 이미지를 생성하는 것을 확인할 수 있다. Prompt 1을 해석하면서 디테일이나 생성된 내용, 색감, 구조적인 형태의 완성도 등 모든 면에서 높은 퀄러티를 보여주었다. Prompt 2에서는 약간의 상상력을 더하였지만 대체로 적절한 구조의 집을 세밀한 붓 터치로 표현하였다. Prompt 3에서는 의외로 우주 성운 표현에서 명령하지 않은 커다란 행성을 표현한 점이 특이하다.

### 3-2 스테이블 디퓨전(Stable Diffusion)

#### 1) 선정 이유

현존하는 이미지 생성 AI 중에 유일하게 서비스에 접속하여 사용하지 않고 개인 컴퓨터에 설치하여 사용할 수 있다. 독일 뮌헨 대학서 여러 기업의 지원을 받아 개발하였으며, 오픈AI의 CLIP 멀티모달 모델에 어떤 학습모델을 추가하였는지 개발 과정은 물론 학습데이터와 매개변수 파일까지 모두 공개하였다. 학습데이터 역시 비영리 단체에서 오랫동안 구축한 공개데이터를 사용한 점도 흥미롭다. 공익적 의도로 만들어진 회사인 오픈AI도 초기에는 이미지 생성 AI의 핵심인 CLIP모델을 공개한 다음부터는 핵심기술은 공개하지 않고 있다. 그만큼 경쟁이 심한 상황임에도 스테이블 디퓨전은 모든 내용을 공개하였고, 특히 매개변수 파일을 공개함으로써 다른 소규모 벤처 기업들이 기본 매개변수를 미세조정 하는 것만으로도 차별적인 이미지를 생성할 수 있게 되었다. 이것으로 학습을 위해 대규모 서버에서 오랫동안 연산할 필요가 없게 됨에 따란 수많은 이미지 생성 AI 서비스가 출시되는 결과를 낳았다. 이러한 의미에서 비교 대상으로 선정하였다.

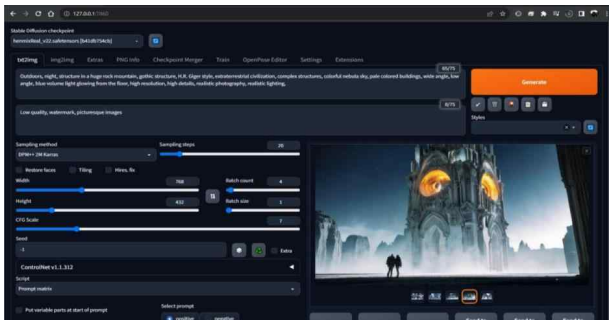

**그림 6.** 스테이블 디퓨전 웹브라우저 인터페이스
**Fig. 6.** Stable diffusion web browser interface

#### 2) 특징

전 세계 개발자들이 모이는 커뮤니티이자 소스 코드 공유 플랫폼인 깃허브에서 누구나 내려받아서 설치할 수 있다. 프롬프트 입력과 함께 세부적으로 제어할 수 있는 조절 값이 가장 많은 점은 고급 사용자에게는 장점이지만 초보자들에게는 단점이라 할 수 있다. 오픈소스인 만큼 초기에는 개발자 중심으로 매우 불편하였지만, 최근에는 전 세계 개발자들이 참여하여 설치 과정의 간소화는 물론 그림 6과 같이 HTML로 컴퓨터 내부 IP에 웹브라우저로 접속하여 수치나 프롬프트를 입력하는 인터페이스를 개발하여 일반인들도 쉽게 사용할 수 있도록 발전하였다. 무엇보다 공개된 학습데이터를 미세 조정한 또 다른 많은 학습데이터가 공유되고 있는 점은 가장 큰 장점이다. 예를 들어 미야자키 하야오 스타일, 실사 사진 스타일, 건담 로봇 스타일과 같은 학습데이터들을 쉽게 구할 수 있다[14].

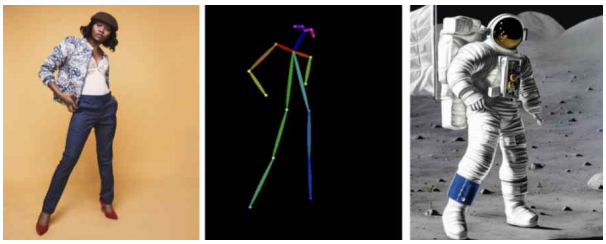

**그림 7.** 자세 추출 및 제어용 컨트롤넷 예시
**Fig. 7.** Example of ControlNet for posture extraction and control

또한 컨트롤 넷(Control Net)이라는 일종의 플러그인을 설치하여 그림 7처럼 사람의 자세를 추출하고 제어하거나 간단한 스케치를 화려한 그림으로 만들 수도 있고 다른 사진의 구도를 빌려 새로운 이미지로 생성하거나 외곽선을 추출하여 새로운 색으로 칠하기도 하는 등, 여타 이미지 생성 AI가 웹 사이트에 접속하여 서버에서 주어진 기능만 사용하는 것에 비해 다양한 기능을 추가하여 사용할 수 있다. 무엇보다 개인 컴퓨터에 설치하여 사용하므로 서버의 관리를 벗어나 프롬프트 사용에 제한이 없는 점은 불법적으로 사용할 수 있으므로 우려스러운 점이면서도 가장 많은 사용자가 있는 이유이기도 하다. 최근에 백악관 폭발[15]과 같은 조작된 사진으로 생긴 문제들은 스테이블 디퓨전으로 생성된 것들이다.

사용료는 공개된 소스 코드를 개인 컴퓨터 설치하는 형태이므로 무료이지만 컴퓨터 성능에 따라 이미지 생성 속도가 달라지는데 쾌적한 작업을 위해서는 100만 원 이상의 NVIDIA사의 고사양 그래픽카드를 설치해야 한다. 핵심 개발사인 스테이빌리티(Stability) AI에서 공식적으로 배포하는 모델데이터는 전 세계 사용자들이 미세 조정해서 만드는 모델데이터의 토대가 되는데, 23년 7월에 기존 모델보다 뛰어난 SDXL을 발표하면서 유일한 설치형 생성 AI로서의 위치를 계속 유지하고 있다.

### 3) 인터페이스와 프롬프트 사용법

유일하게 PC에 개별 설치하여 사용할 수 있으며 HTML로 만들어진 인터페이스를 사용하여 가장 복잡하면서도 세부적인 제어가 가능하다. 대표적인 특징으로 일반 프롬프트와 네거티브 프롬프트를 따로 적을 수 있으며 옵션이 많은 만큼 고유의 프롬프트 작성 문법도 있다. 기본적으로 앞에 단어가 중요하지만, 어떤 위치의 단어라도 ( )를 붙이면 중요도가 10% 더해지며, (((단어)))와 같이 최대 30%의 가중치를 부여할 수 있다. 반대로 [ ]를 붙이면 10% 가중치를 덜어낸다. 역시 30%까지 가능하다. 그림 8처럼 해상도는 물론 모든 제어 값은 슬라이더나 수치 입력으로 세밀하게 제어할 수 있다. 한

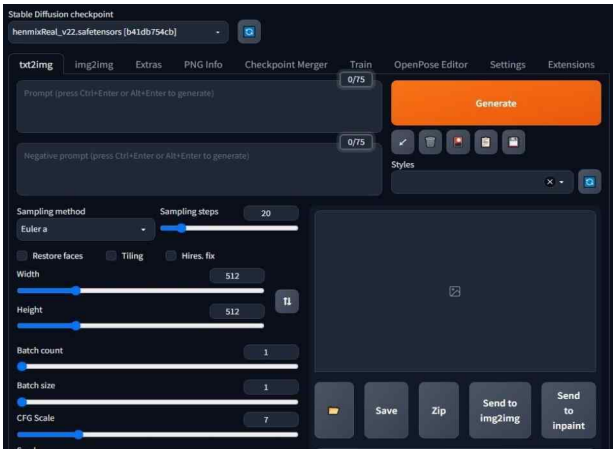


**그림 8.** 세부 제어 인터페이스
**Fig. 8.** Detailed control interface

번에 생성할 수 있는 이미지 개수나 이미지 변화 정도, 컨트롤넷 제어 등, 처음 사용자는 오히려 너무 많은 옵션으로 어렵게 느낄 수도 있으나 디폴트 값으로 두고 일반 프롬프트와 네거티브 프롬프트, 해상도와 생성 시드 값 정도만 제어하여 쉽게 이미지를 생성할 수 있다.

### 4) 테스트 프롬프트 적용

스테이블 디퓨전은 미세조정 학습데이터를 교체할 수 있으며 그에 따라 전혀 다른 느낌의 이미지가 생성되는 것이 가장 큰 장점이다. 그러므로 다른 이미지 생성 AI와 프롬프트 사용 방법의 차이를 비교하는 것이 의미가 있으나 생성한 이미지를 비교하는 것은 참고 사항이라 할 수 있다. 표 2의 3종류 Prompt를 똑같이 적용하였으며 스테이블 디퓨전만의 특징인 네거티브 프롬프트는 사용하지 않았다.

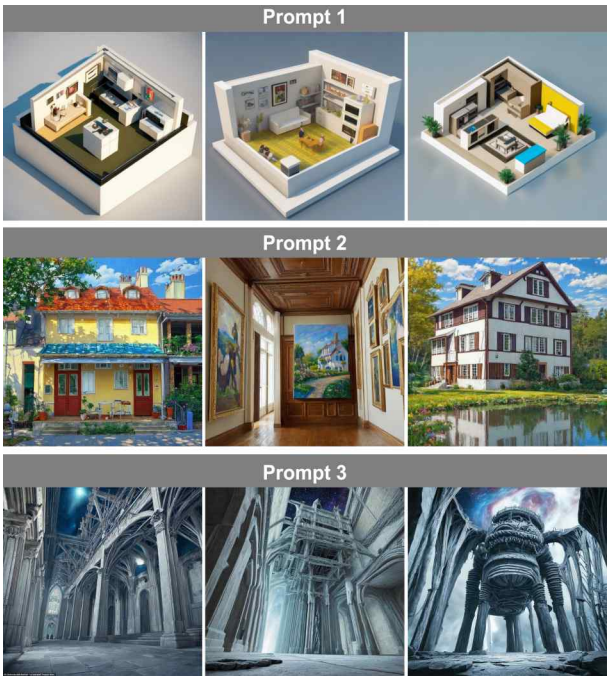


**그림 9.** 스테이블 디퓨전 테스트 프롬프트 이미지 생성 결과
**Fig. 9.** Stable diffusion test prompt image creation results

학습데이터는 모델 다운로드 사이트에서 가장 추천 수가 많은 Henmix_real_V2.2[16]을 설치하였으며 모델 특성은 이름과 같이 실제 사진과 같은 이미지를 생성하기에 적합하고 비현실적인 상상의 이미지를 생성하는 것은 부족할 수 있다.

생성된 이미지는 그림 9와 같으며 구체적으로 살펴보면 Prompt 1은 인간의 삶이라는 단어를 넣었지만, 사람은 없고 사물만 표현하였으며 사물들은 학습데이터 특성에 맞게 정확한 형태를 생성하였다. Prompt 2도 화려하고 특이한 형태의 집이라고 명령하였지만 전혀 특이하지 않은 집을 생성하였으며 집 내부를 표현하기도 하였다. 또한 직접 그린 듯한 붓 터

치는 표현하지 못하고 그림 같은 색감을 표현하였다. Prompt 3에서는 H. R 기거 이미지 스타일, 고딕 양식, 두 단어를 연구자가 기대한 느낌으로 가장 잘 표현하였다. 그러나 야외라는 단어가 있었음에도 주로 건물 내부를 표현한 점은 아쉽다.

### 3-3 파이어플라이(Fire fly)

#### 1) 선정 이유

그래픽 분야에서 가장 크고 영향력 있는 기업인 어도비는 이미 2016년부터 '어도비 센세이(Adobe Sensei)'라는 비교적 낮은 단계의 AI 서비스를 그래픽 프로그램들에 연동하고 있었으나 2021년부터 생성형 AI를 어도비 센세이에 도입하면서 그래픽 프로그램들의 기능을 극대화하고 있다. 그림 10처럼 파이어플라이는 어도비 웹에 접속하여 사용할 수 있지만 이보다는 기존 어도비 프로그램 속에서 구현하는 방식이 더 기대된다.

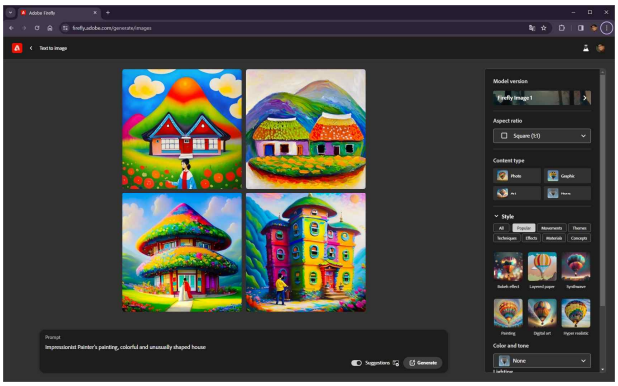

**그림 10.** 어도비 웹의 파이어플라이 인터페이스
**Fig. 10.** Firefly interface on Adobe Web

24년 1월 현재 2.0 버전이 나왔으며 아직 개발 중인 기능들도 있지만 충분한 시험을 거친 완성된 기능들은 포토샵과 같은 기존 구독제 사용자들이 프로그램 안에서 생성 AI 기능을 사용할 수 있도록 기능 업데이트에 바로 반영하고 있다. 생성되는 이미지의 퀄러티는 미드저니보다는 다소 부족하지만, 대기업의 개발력과 어도비가 보유한 다양한 그래픽 프로그램과의 연동 가능성을 고려할 때 앞으로의 발전과 확장성이 기대되는 서비스다. 이러한 의미에서 비교 대상으로 선정하였다.

#### 2) 특징

어도비는 교육기관에서 가장 많이 사용하는 프로그램 패키지를 제공하는 곳으로서 대부분 교육기관이 단체 사용권 계약을 맺고 있다. 미드저니나 달·이를 사용하기 위해 개인별로 비용을 지급해야 하거나 스테이블 디퓨전을 운용하기 위해 고사양의 컴퓨터를 구매해야 하는 점을 고려할 때 이미 기관별로 사용권 계약을 맺고 있는 어도비의 이미지 생성 AI 서비스는 매우 유용한 대안이 될 수 있다.

또한 해결되지 않은 학습데이터 저작권 문제는 앞으로 창작자가 이미지 생성 AI를 사용할 때 발생할 수 있는 문제로서 어도비는 자사가 저작권을 소유한 이미지뱅크 서비스인 어도비 스톡의 데이터를 학습하는 것으로 생성된 이미지에 대한 저작권 문제를 원천적으로 방지하였다. 또한 어도비 스톡에 사용자가 이미지를 업로드하여 학습데이터로 제공하면 이미지 생성 기여도에 따른 보상 체계까지 마련하여 선순환의 생태계를 구축하였다[17]. 여기에 더하여 어도비는 기업 고객에게는 자사의 생성 이미지로 인하여 법적 분쟁이 발생 시 법률지원 서비스까지 지원한다[18].

파이어플라이는 어도비 패키지 중 하나의 소프트웨어가 아니라 어도비의 생성형 AI 기능 전체를 일컫는 단어다. 그림 11처럼 23년 9월 업데이트된 포토샵 v25에서는 그동안 어도비 웹에서 베타테스트 하였던 생성형 채우기 기능과 프레임 확장 기능을 포토샵의 인터페이스 속에서 명령할 수 있도록 기능이 추가되어 특별한 재교육 없이 바로 사용할 수 있다. 포토샵에서 유용하게 사용하던 Content-Aware Fill 기능이 에프터이펙츠에도 적용되어 이미지를 합성에서 불필요한 부분을 지우는 기능으로 유용하게 사용했던 경우를 볼 때 지금의 이미지 생성 AI의 기능도 에프터이펙츠나 프리미어에 충분히 적용될 수 있을 것이라 기대한다. 어도비 프로그램들의 데이터 호환성은 매우 강력하여 포토샵의 비트맵 데이터와 일러스트레이터의 벡터 데이터 상호 호환은 물론 프리미어 편집 소스나 에프터이펙츠 특수효과 영상 소스로서 데이터의 특성을 온전하게 유지하면서 주고받을 수 있다. 여기에 생성 AI 기능 추가에 따른 다양한 시너지 효과가 기대된다.

**그림 11.** 어도비 포토샵의 파이어플라이 인터페이스
**Fig. 11.** Firefly interface on Adobe Photoshop

사용료를 살펴보면 23년 11월부터 이월되지 않는 월간 크레딧 제도가 도입되었다. 크레딧은 생성 개수를 말하며 기존 어도비 사용 요금제에 따라 1000 크레딧부터 25 크레딧까지, 다양하게 구성되어 있다[19]. 베타테스트 기간에는 무료로 사용하였지만, 이미지 생성에는 서버의 연산 자원을 사용하므로 전 세계 사용자들을 고려해 볼 때 유료화는 필수 과정이라 할 수 있다. 교육기관에서 주로 사용하는 라이선스인 SDL (Shared Device License)에서는 이미지 생성 AI 기능을 사

용할 수 없으므로 해당 관리자와 추가 비용에 대한 논의가 필요하다.

#### 3) 인터페이스와 프롬프트 사용법

포토샵이나 일러스트레이터와 같은 프로그램에서 직접 사용할 수도 있으나 다른 생성 AI와 비교를 위해 웹사이트에 접속하여 사용하는 기준으로 설명한다. 여타 이미지 생성 AI에서는 화면비율이나 이미지 스타일 옵션을 프롬프트로 작성하지만, 파이어플라이처럼은 인터페이스 우측에 풀다운 메뉴로 선택할 수 있도록 제공하여 사용자가 쉽게 생성 이미지를 세부적으로 제어할 수 있도록 하였다. 실제 옵션은 효과, 기법, 색상 및 톤 구성 등, 다양한 제어 명령을 인터페이스의 풀다운 메뉴를 최대한 활용하도록 유도하고 프롬프트 자체 문법은 없으므로 단어 나열로 쉽게 프롬프트 작성이 가능하다.

생성한 이미지 퀄러티는 미드저니보다 다소 부족하지만 사용자 인터페이스 편의성은 가장 뛰어나다고 할 수 있으며, 어도비 프로그램 간의 호환성은 물론 어도비 스톡 데이터 학습으로 생성 이미지의 저작권 문제도 없으므로 거대 기업의 지속적인 연구 지원으로 생성 이미지 퀄러티를 높인다면 가장 강력한 이미지 생의 AI가 될 것으로 예상한다.

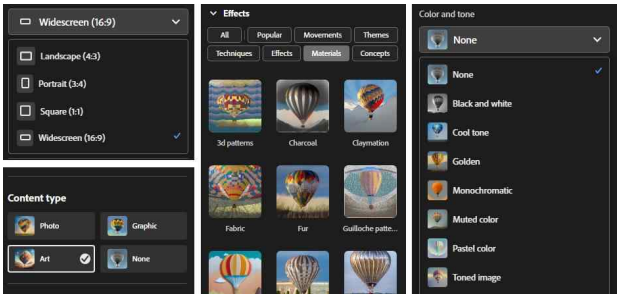


**그림 12.** 파이어플라이 풀다운 선택 메뉴 일부
**Fig. 12.** Part of the Firefly pull-down selection menu

#### 4) 테스트 프롬프트 적용

앞서와 같이 표 2의 Prompt 3종류를 적용하였으며 생성된 이미지는 그림 13과 같다. 어도비 웹 인터페이스도 한 번에 4개의 이미지를 생성하지만 3개만 나열하였다.

생성된 이미지를 구체적으로 살펴보면 Prompt 1은 인간의 삶이라는 프롬프트를 가장 강하게 반영하여 사람을 포함하여 생성한 점이 특이하였으며 3D라고 하였음에도 2D 일러스트 형식으로 생성한 것은 아쉽다. Prompt 2는 인상파 화가들의 붓 터치 표현이 부족하지만, 화려한 색의 일반적이지 않은 집이라는 단어는 비교적 잘 생성하였다. Prompt 3에서는 사진과 같은 디테일을 지정하였음에도 회화적인 느낌으로 생성하였다. 화려한 우주 성운은 잘 반영되었으나 'H. R 기거 이미지 스타일, 고딕 양식'이라는 단어는 잘 반영하지 못하였는데 이는 학습 데이터로 사용된 어도비 스톡의 특성상 개인이 촬영한 사진이 많고 유명 작가의 작품이 적은 것을 이유로 유추할 수 있다.

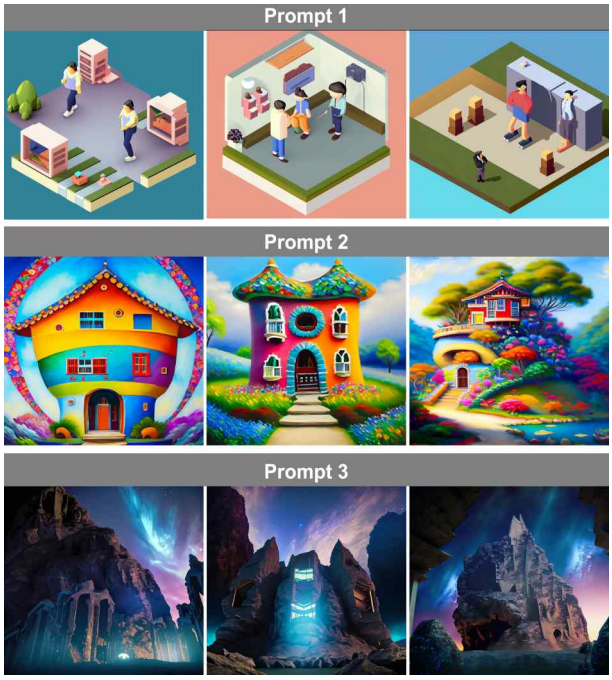


**그림 13.** 테스트 프롬프트를 적용한 파이어플라이 이미지 생성 결과
**Fig. 13.** Test prompt results of Firefly image generation

### 3-4 달·이(DALL·E)

#### 1) 선정 이유

챗GPT로 유명한 오픈AI에서 21년 1월 텍스트를 이미지로 변환하는 핵심기술인 멀티모달 모델 CLIP(Contrastive Language -Image Pre-training)을 공개하면서 동시에 전 세계에서 가장 먼저 일반 사용자들이 사용할 수 있는 이미지 생성 AI를 발표하였다. 이후의 모든 이미지 생성 AI는 달·이의 멀티모달 기술로부터 발전했다고 할 수 있다. AI의 불법적인 사용을 막기 위해 기술을 공개하자는 의도로 만들어진 오픈AI는 챗GPT와 함께 다양한 생성 AI 분야에서 가장 선도적인 기술과 서비스를 출시하고 있으며 마이크로소프트가 100억 달러를 투자하면서 마이크로소프트 오피스 프로그램과 유기적으로 연동하는 AI 서비스 개발에 집중하고 있다. 그러나 달·이는 처음 관심과 달리 현재는 미드저니에 비해 언급이 거의 되지 않고 있다. 생성한 이미지 퀄러티는 후발 개발사인 미드저니에 비하여 부족하지만, 생성 AI 기술을 선도하는 회사이면서 가장 먼저 이미지 생성 AI 서비스를 오픈하였기에 비교 대상으로 선정하였다.

#### 2) 특징

오픈AI 웹사이트의 서브 페이지 접속하여 사용할 수 있으며 서브 페이지로 운영 중인 것을 볼 때 서비스의 중요성을 높지 않게 보고 있음을 유추할 수 있다. 그림 14는 달·이 서비스에 접속하면 볼 수 있는 인터페이스로서 단순하고 직관적이다.

달·이라는 이름은 픽사 애니메이션의 주인공 로봇 WALL·E와 초현실주의 화가 살바도르 달리 이름을 합성한 것이다. 사용해 보면 세부적인 프롬프트가 아니면 다소 이미지 퀄러티가 떨어져 보일 수도 있지만 이름처럼 오히려 작가와 같은 개성 있는 작품처럼 보이는 이미지를 생성하는 경우가 많았다.

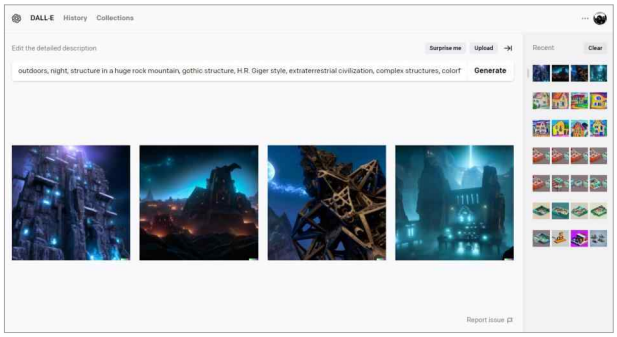


**그림 14.** 달·이의 간결한 웹페이지 인터페이스
**Fig. 14.** DALL·E's simple web-page interface

특히 오픈AI는 달·이 사용법을 상세하게 설명한 PDF 문서를 배포 중인데[20] 달·이와 더불어 다른 이미지 생성 AI를 사용하면서도 활용할 수 있는 좋은 지침서로서 AI 기술을 선도하는 연구소로서의 면모를 확인할 수 있다. 요금은 초기에는 무료로 경험할 수 있었지만, 현재 미드저니와 마찬가지로 무료 사용은 없으며 월 15달러를 결제하면 115번 생성할 수 있는 크레딧을 받을 수 있다.

### 3) 인터페이스와 프롬프트 사용법

가장 간단한 인터페이스를 가지고 있으며 프롬프트 사용법에서도 특별한 문법 없이 간단하다. 그림 15는 달·이의 특정 인터페이스 부분을 캡처한 것이다. 최초 생성 이미지는 1:1 비율로 1024×1024로 정해져 있으며 16:9 비율로 만들기 위해서는 원하는 이미지를 선택하면 보이는 서브 아이콘의 Edit로 들어가 어림짐작으로 16:9 정도 비율로 확장하는 과정으로 만들어야 한다. 그 외 Edit에는 우측과 같이 특정 이미지 부분만 지우고 새로 생성하는 것, 움직이고 저장하는 것, 총 4종류의 아이콘만 있다. 비교 대상 4개 중에서는 가장 직관적이고 쉬우므로 초급자에게는 좋지만, 전문적으로 사용하기에는 이미지 퀄러티 면이나 인터페이스 편의성 면에서 상대적으로 부족하다.

**그림 15.** 좌: 간단한 메뉴 버튼 / 우: Edit 모드의 기능 아이콘
**Fig. 15.** L: Simple menu button / R: Function icons in edit mode

### 4) 테스트 프롬프트 적용

달·이도 같은 방법으로 표 2 Prompt 3종류를 적용하였으며 생성된 이미지는 그림 16과 같다. 달·이 역시 한 번에 4개의 이미지를 생성하지만 3개만 나열하였다.

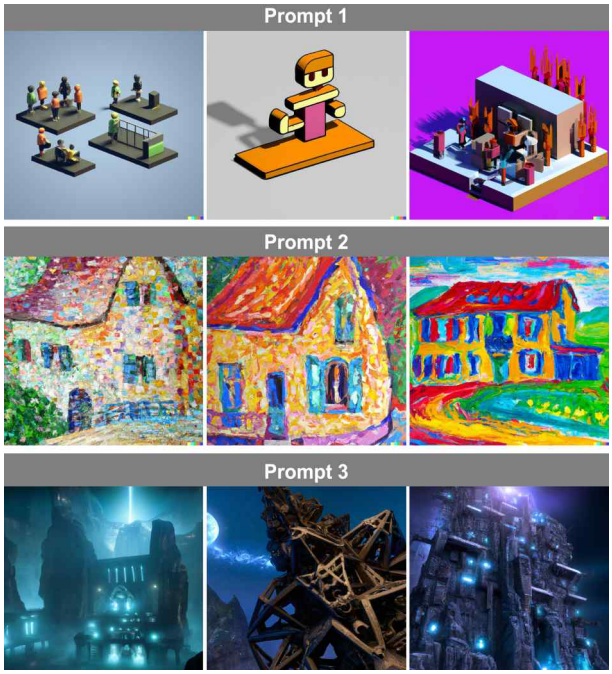


**그림 16.** 테스트 프롬프트를 적용한 달·이의 이미지 생성 결과
**Fig. 16.** Test prompt Results of DALL·E image generation

구체적으로 살펴보면 Prompt 1은 생성된 이미지의 편차가 가장 심했으며 복잡하게 보이는 이미지도 자세히 보면 구조 면에서 틀린 형태의 이미지를 생성하였다. Prompt 2는 달·이라는 서비스 이름에 걸맞게 인간이 그린 듯한 페인팅 느낌을 생성하였다. 그러나 일반적이지 않은 집이라는 명령의 표현은 왜곡되기는 했으나 특이한 형태는 아니었다. Prompt 3은 많은 단어를 사용하여 세부적으로 명령한 만큼 의도에 근접한 이미지를 생성하였으나 'H. R 기거 이미지 스타일, 고딕 양식'이라는 단어표현은 조금 부족해 보인다. 또한 실제와 같은 표현을 요구하였으나, 이미지를 자세히 보면 형태를 명확히 들어내지 않고 생략하거나 붓 터치로 그린 듯한 질감 등, 다소 부족하였다.

## Ⅳ. AI가 생성하는 프롬프트 의미와 분석 결과

### 4-1 이미지 생성 AI 사용에서의 어려움

이미지 생성 AI는 단어나 문장을 이해하고 이미지를 생성하지만, 작가가 의도한 이미지를 한 번에 생성하기는 불가능하므로 생성된 이미지 중에서 선택하고 그것을 기준으로 다

시 이미지를 생성하는 과정을 반복한다. 대부분의 이미지 생성 AI가 하나의 명령에 4장을 생성하도록 구성된 것도 선택의 폭을 넓혀주기 위함이다. 앞으로 작가와 생성 AI의 협업에는 이러한 선택의 과정이 중요할 것이며 연구자는 선행연구에서 이 과정에 의미를 더하여 '창의 선택'이라 정의하였다. 그러나 작가의 의도에 적합한 이미지를 생성하기 위해서는 반복해서 생성하는 것은 물론 다른 이미지를 참고하여 생성하거나 이미지 일부를 지정해서 수정하도록 하는 등 여러 기법을 사용해야 한다.

일반인이 텍스트 프롬프트를 입력하고 AI가 이미지를 생성하는 것은 매우 흥미로운 경험이 될 수 있지만 전문 작가나 프로덕션에서 생성 AI를 활용하여 정확히 의도된 이미지를 생성하기 위해서는 프롬프트에 따른 이미지 생성의 방향성을 경험적으로 숙지해야 하며 인터페이스에 따라 사용하는 방법과 특정 값에 대한 변화의 정도도 익혀야 하는 등, 숙련의 과정을 거쳐야만 한다. 이에 연구자는 선행연구에서 앞으로의 창작자들도 생성 AI를 보다 빨리 사용하면서 경험을 쌓는 것이 중요한 경쟁력이 될 것으로 예상한다. 그리고 23년 10월 챗GPT 환경에서 운영되는 달·이 3이 서비스를 시작하면서 프롬프트 작성에 효과적인 도움을 받을 수 있는 또 다른 방안을 제시하였기에 본 연구의 프롬프트 구성요소로 살펴보았다.

### 4-2 프롬프트 명령이 아닌 대화형 인터페이스

연구자는 멀티모달 설명에서 컴퓨터를 제어하는 최후의 인터페이스는 인간의 언어라 정의하였다. 텍스트로 이미지를 생성하는 AI는 모두 멀티모달이라 할 수 있지만 원하는 결과를 얻기 위한 과정을 살펴보면 프롬프트 단어를 더하고 빼는 반복 과정의 시행착오를 통해 원하는 이미지를 생성하는 비효율적인 방식으로서 진행하고 있었다. 이러한 답답한 대화 형식에서 달·이 3부터는 챗GPT와 통합하여서 사람과 온전히 대화하듯이 이미지를 생성할 수 있게 되었다. 이러한 협력 서비스는 챗GPT와 달·이를 개발하고 생성 AI 시장을 선도하는 오픈AI의 전략이라 할 수 있으며 그림 17과 같이 챗GPT와 같은 인터페이스 속에서 달·이 3이 그대로 구현되는 것을 확인할 수 있다.

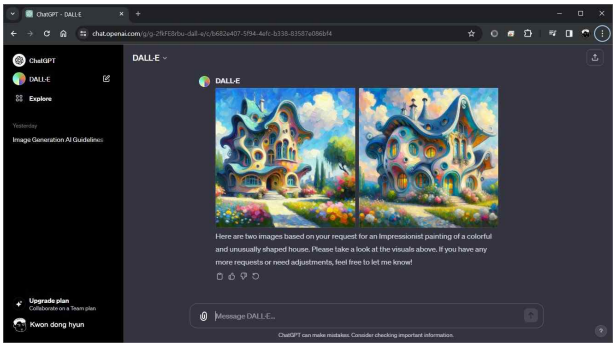


**그림 17.** 챗GPT와 달·이 3이 함께 작동하는 사용자 인터페이스
**Fig. 17.** User interface where ChatGPT & DALL·E 3 work together

챗GPT와 인터페이스가 통합되면서 생성한 이미지를 두고 AI와 이야기를 주고받으면서 추가적인 작업을 진행할 수 있게 되었다. 예를 들어 이전 달·이의 인터페이스에서는 정확한 비율로 이미지를 생성하기 힘들었으나 달·이 3에서는 말하듯이 16:9나 2.35:1로 그려달라고 하면 다시 생성한다. 연구자가 사용해 본 결과 이미지 비율은 한 번만 이야기하면 다음부터는 비율에 대한 명령을 내리지 않더라도 계속 앞에서 지정한 비율로 생성되었다. 또한 챗GPT 4의 한글 인식 능력이 매우 뛰어나므로 달·이 3에서는 다른 이미지 생성 AI와 달리 한글 입력도 가능하며 모바일 앱으로는 음성입력까지 지원한다.

하지만 인터페이스의 변화로 이전보다 생략된 기능도 있다. 한 번에 생성하는 이미지 개수가 4개에서 2개로 줄었으며 특정 부분을 지우고 생성시키는 브러쉬 기능이나 기존 이미지를 바탕으로 추가, 확장하는 Edit 아이콘은 생략되었다. 이는 달·이 3을 창작자의 협업 관계보다는 챗GPT와 함께 이미지 분석 및 생성 기능을 이용하여 사용자의 삶을 도와주는 개인비서 역할로서 방향을 설정하고 있다고 유추할 수 있다.

### 4-3 테스트 프롬프트와 AI 생성 프롬프트 비교

달·이 3은 이전보다 수준이 높고 정확한 이미지를 생성한다고 오픈AI는 홍보하였다. 여기에는 챗GPT를 통해 대화의 주제를 더 깊이 이해하는 것도 포함된다. 무엇보다 놀라운 점은 인간의 프롬프트를 해석하여 AI가 프롬프트를 다시 생성해서 그린다는 점이다. 이에 앞에서 사용한 표 2 Prompt를 적용하여 이미지 수준이 어느 정도 향상되었는지 알아보았다.

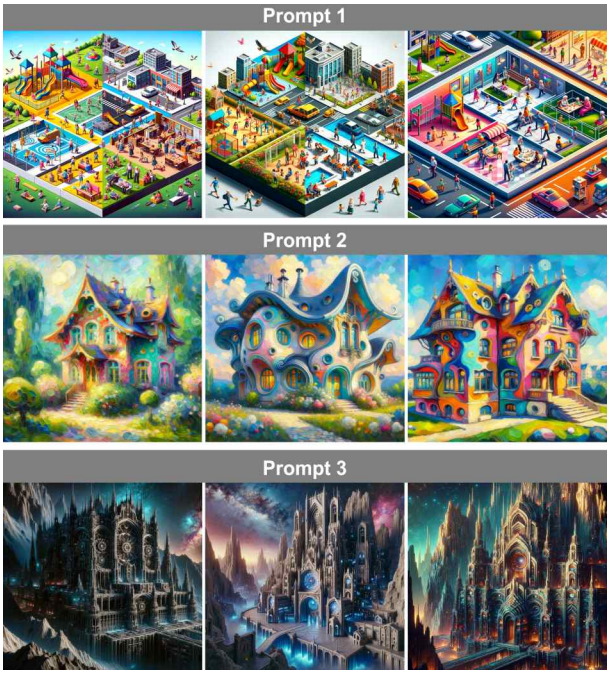


**그림 18.** 달·이 3의 테스트 프롬프트 이미지 생성 결과
**Fig. 18.** DALL·E3 test prompt image creation results

그림 18은 Prompt 3종을 달·이 3에 적용하여 이미지를 생성 결과로써, 그림 16의 생성 이미지들과 비교하면 오픈AI의 홍보 내용처럼 더 다양한 학습데이터와 정교한 학습 알고리즘, 효율적인 데이터 처리 기술 등으로 뚜렷하게 퀄러티가 향상된 것을 확인할 수 있다. 구체적으로 살펴보면 Prompt 1은 인간의 삶을 다양하게 표현하였으며 디테일에서도 사물의 형태나 구조에는 오류가 없었으나 아직 사람 표현에서는 자세히 보면 형태적인 오류가 있었다. 하지만 전체적으로는 매우 뛰어난 해석과 디테일을 볼 수 있었다. Prompt 2는 달·이라는 이름에 걸맞게 두꺼운 유화 붓 터치로 인간이 그린 듯한 페인팅 느낌을 잘 표현하였으며 특이한 집이라는 단어도 잘 표현하였다. Prompt 3에서는 'H. R 기거 이미지 스타일, 고딕 양식'에서는 고딕 양식 쪽으로 좀 더 많이 표현하였지만 그럼에도 이전과 비교할 수 없을 만큼 건축물의 디테일이 좋아졌다.

다음으로 연구자가 작성한 Prompt를 AI가 어떻게 인식하고 있는지를 비교하였다. AI가 생성한 프롬프트는 최초 프롬프트 입력 후 이미지를 생성한 다음 챗GPT 입력창에 이미지 생성에 사용한 프롬프트를 알려달라고 말하거나 생성한 이미지를 클릭한 후 나타나는 오른쪽 느낌표 아이콘을 클릭하면 보여준다.

표 3은 연구자의 Prompt 3종류를 달·이 3이 이해한 후 다시 생성한 프롬프트를 비교한 것이다. 구체적으로 살펴보면 Prompt 1에서는 인간의 다양한 삶을 표현하는 방법으로 4개 구역으로 나누고 어린 시절의 노는 모습, 십 대의 학교생활, 어른들이 일하는 분주한 도시, 노인들이 걷고 쉬는 공원으로 표현하였다. 즉, 인간의 삶을 인생의 4단계로 구분해서 세밀하게 프롬프트를 만들었다. Prompt 2를 살펴보면 다채롭고 특이한 모양의 집, 부드럽고 흐릿한 붓질, 생생한 색상, 빛과 그림자의 유희에 초점을 맞춰 인상주의의 본질을 포착해야 하고, 집은 기발하고 독특해야 하며, 무성한 자연 풍경을 배경으로 평온함과 예술적인 아름다움을 느낄 수 있어야 한다는 것을 설명하고 있다. 최초 내용과 방향성은 같으나 좀 더 섬세하게 명령을 만들었음을 알 수 있다. Prompt 3은 단어들을 연결하여 자연스러운 문장으로 만들었지만, 내용상으로는 특별히 더 추가된 내용은 없는 것을 알 수 있었다. 결과적으로 의미를 함축한 단어를 사용하여 포괄적인 내용으로 작성한 Prompt 1은 AI가 단어의 의미를 해석하고 그에 대한 이미지를 생성하기 위한 구조를 만들고 있음을 알 수 있었으며, 그려야 할 내용은 명확히 제시하면서 그것에 대한 표현을 함축적으로 작성한 Prompt 2는 표현에 관한 단어들을 더 많이 사용하여 섬세한 프롬프트로 만들었음을 확인할 수 있었다. 반면에 연구자가 표 1에서 구분한 프롬프트 구성요소에 맞추어서 작성한 Prompt 3은 AI가 인식하여 생성한 내용에서 큰 차이가 없음을 확인할 수 있었다. 즉, 본 연구에서 정리한 프롬프트 구성요소에 맞게 작성한 Prompt 3은 AI가 인식하고 이미지를 생성하기에 적절한 정보를 제공하고 있음을 알 수 있다.

**표 3.** 테스트 프롬프트, 인간 작성 vs. AI 생성
**Table 3.** Test prompts, human-written vs. AI-generated

| | | |
|---|---|---|
| Prompt 1 | Human | Isometric, 3D illustration, human life |
| | AI | An isometric, 3D illustration showcasing various aspects of human life. The scene is divided into four quadrants, each representing different stages of life. The first quadrant shows a colorful playground with children playing, symbolizing childhood. The second quadrant illustrates a bustling high school scene with teenagers engaging in various activities, representing adolescence. The third quadrant depicts a busy cityscape with adults working and socializing, symbolizing adulthood. The fourth quadrant portrays a serene park with elderly people walking and resting, representing old age. The illustration is vibrant and detailed, capturing the essence of each life stage. |
| Prompt 2 | Human | Impressionist Painter's painting, colorful and unusually shaped house |
| | AI | An impressionist style painting featuring a colorful and unusually shaped house. The artwork should capture the essence of Impressionism with soft, blurred brush strokes, vibrant colors, and a focus on the play of light and shadow. The house should be whimsical and distinctive, set against a lush, natural landscape, evoking a sense of tranquility and artistic beauty. |
| Prompt 3 | Human | outdoors, night, structure in a huge rock mountain, gothic structure, H.R. Giger style, extraterrestrial civilization, complex structures, colorful nebula sky, pale colored buildings, wide angle, low angle, blue volume light glowing from the floor, high resolution, high details, realistic photography, realistic lighting |
| | AI | A realistic, high-resolution photograph depicting an outdoor night scene with a gothic structure built into a huge rock mountain, inspired by H.R. Giger's style. This scene represents an extraterrestrial civilization with complex structures. The sky is filled with a colorful nebula, and the buildings are pale-colored. The perspective is wide angle and low angle, with blue volume light glowing from the floor, emphasizing realistic lighting and high detail. |

## V. 결 론

4개의 이미지 생성 AI 중에서 미드저니는 연구자가 정의한 프롬프트 구성요소에 속한 단어들을 사용하는 유무에 상관없이 안정되고 섬세하며 퀄러티가 높은 이미지를 생성한다. 그러나 이미지 구성이나 아이디어 면에서는 프롬프트 단어가 포괄적일수록 생성된 이미지의 편차도 커지므로 계속 원하는 형식과 내용의 이미지가 생성될 때까지 반복해야 한다. 이는 4종류의 이미지 생성 AI 모두에 해당하는 내용으로, 생성 AI를 활용하여 작업할 때는 프롬프트 구성요소에 맞추어 세밀히 단어를 나열한다면 원하는 이미지를 빨리 얻을 수 있다.

오픈AI의 달·이는 최초의 이미지 생성 AI라는 중요한 역할에 비하여 점차 관심이 줄어들었으나 달·이 3의 등장으로 새로운 방향으로 발전할 것으로 예상된다. 달·이 3은 일반인들이 이미지 생성 AI를 사용할 때, 시행 결과를 보고 다시 시행

하는 반복행위가 아니라 AI와 대화하면서 자신이 어떤 그림을 그리고 싶은지를 알게 되는 과정 가운데 이미지를 생성할 수 있는 단계로 발전하고 있다. 이같이 챗GPT의 도움으로 가능해진 달·이 3의 프롬프트 생성 기능을 활용하여 인간의 프롬프트를 AI가 어떻게 이해했는지를 살펴본 결과 연구자가 정의한 프롬프트 구성요소를 고려하여 작성한 Prompt 3의 경우 AI가 생성한 프롬프트와 큰 차이가 없는 것을 볼 때 본 연구의 목적인 프롬프트 구성요소가 생성 AI를 제어하기 위한 적절한 프롬프트 언어의 틀이 될 수 있음을 확인할 수 있었다.

프롬프트 작성 교육은 글쓰기를 통해 이루어지며, 이미지뿐만 아니라 미래의 모든 분야 생성형 AI와 상호작용하는 데 필수적이다. 따라서 이는 필수 교양과목으로 지정될 필요가 있다. 생성형 AI는 세상이 주목한 지 1년이 되지 않았지만, 그 이후로 새롭게 출시된 서비스의 수와 발전 속도는 캄브리아기 동물들이 폭발적으로 등장한 것에 비견될 만큼 엄청나다. 할리우드 작가 조합과 배우 조합은 AI에 위협을 느끼고 파업을 벌이고 있으며, 국내 웹툰 업계는 생성 AI로 그린 이미지를 인정해야 할지를 논란 중이다. 게임과 일러스트 등 2D이미지를 다루는 분야에서도 인간과 생성형 AI가 조화롭게 협력하는 방법이 어떠한 것인지를 아직 찾지 못하고 있다. 영상 업계에서는 배경 제작이나 합성을 위해 일부 이미지 제거 등에 이미 사용되고 있지만 앞으로는 촬영하지 않은 컷을 생성할 만큼 기술력이 높아진다면 영상 제작 전반에 걸쳐서 활용될 것이다. 근 미래에는 창작 작업 전반에 걸쳐 단순 반복 작업은 점차 생성 AI에 맡기고 인간은 창의적인 작업에 집중하는 시대가 될 것으로 예상되지만 반면에 자신만의 세계관과 창의적인 시각을 가진 창작자가 아닌 경우, 할 수 있는 일이 점차 없어지는 상황이 될 것이다. 그럼에도 많은 노동력을 요구하는 작업에서 생성 AI의 협업이 일반화된다면 개인 창작자들에게는 작업 수준을 높일 수 있는 좋은 도구이자 동반자가 될 수 있으므로 생성 AI의 발전을 긍정적으로 만들 수 있는 것은 결국 사용자에 달려 있다고 할 수 있다.

본 연구에서 비교한 달·이로부터 달·이 3이 나올 때까지의 기간은 1년 6개월 있었다. 하지만 점차 AI가 인간에게 배우는 지도학습을 넘어 스스로 학습하는 자율학습 단계로 나아가고 있으므로 더욱 빠르게 발전할 것으로 예상된다. 그러므로 본 연구에서 언급한 파이어플라이, 미드저니, 달·이, 스테이블 디퓨전 외에도 앞으로 출시될 영상 제작, 3D 애니메이션, 사운드 편집 등 다양한 분야의 생성 AI를 잘 다루기 위해서는 사용자가 직접 프롬프트 명령을 작성하고 그 결과를 체험해 보는 것이 중요하며, 교육 현장에서는 프롬프트 명령을 실제로 사용하면서 시행착오를 줄일 기회를 제공해야 할 것이다. 본 프롬프트 구성요소 연구가 이미지뿐만 아니라 앞으로 등장할 다양한 분야의 생성 AI를 이해하고 활용하는 데 도움이 되길 기대한다.

## 참고문헌


[1] Electronic Times Newspaper. ‘ChatGPT’ Craze Shakes Up AI Market [Internet]. Available: https://www.etnews.com/20230130000092.

[2] Sequoia. Generative AI: A Creative New World [Internet]. Available: https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/.

[3] T. W. Park, *Taewoong Park’s AI Lecture*, Seoul: Hanbit Biz, 2023.

[4] C. H. Moon and D. H. Kwon, “A Study of User-Level Understanding of Image-Generative AI and How to Apply It to Animation Production Education,” *Cartoon & Animation Studies*, No. 72, pp. 213-262, September 2023. http://dx.doi.org/10.7230/KOSCAS.2023.72.213

[5] M.-J. Lee and E. Choi, “A Study on Creative Nail Art Design Generation Based on Text Prompt: Focused on Image-Generating Artificial Intelligence Models, DALL-E 2 and Bing Image Creator,” *Journal of the Korean Society of Cosmetology*, Vol. 29, No. 4, pp. 1058-1065, August 2023. https://doi.org/10.52660/JKSC.2023.29.4.1058

[6] S. Park and S. Kim, “A Comparative Study of Image Generation Artificial Intelligence Service Used in Idea Development Process - Focusing on Text Prompt Based Image Generation Design Service -,” *Journal of Korea Design Forum*, Vol. 28, No. 2, pp. 7-16, May 2023. http://dx.doi.org/10.21326/ksdt.2023.28.2.001

[7] S.-H. Lee and K.-S. Song, “Exploring the Possibility of Using ChatGPT and Stable Diffusion as a Tool to Recommend Picture Materials for Teaching and Learning,” *Journal of the Korea Society of Computer and Information*, Vol. 28, No. 4, pp. 209-216, April 2023. http://dx.doi.org/10.9708/jksci.2023.28.04.209

[8] S.-H. Lee and K.-S. Song, “Prompt Engineering to Improve the Performance of Teaching and Learning Materials Recommendation of Generative Artificial Intelligence,” *Journal of the Korea Society of Computer and Information*, Vol. 28, No. 8, pp. 195-204, August 2023. http://dx.doi.org/10.9708/jksci.2023.28.08.195

[9] H. Han, The Study of User Customization of AI images Training Model through Instance Prompt settings, Master’s Thesis, Pusan National University, Busan, August 2023.

[10] Artificial Intelligence Times. Open AI Evolves from GPT-3 for Writing... Unveils ‘DALL·E’ and ‘CLIP’, AI Models that Read Text and Draw Pictures [Internet]. Available: https://www.aitimes.kr/news/articleView.html?idxno=18892.

[11] Y. Hwang, “The Emergence of Generative AI and PROMPT Literacy: Focusing on the Use of ChatGPT and

DALL-E for English Education," *Journal of the Korea English Education Society*, Vol. 22, No. 2, pp. 263-288, May 2023.

[12] Y. M. Yi, "New Paradigm and Literacy in the Digital Era : Focusing on Digital Literacy and AI Literacy," *The Journal of General Education*, No. 20, pp. 35-60, July 2022. http://dx.doi.org/10.24173/jge.2022.07.20.2

[13] Notion. KEEPKWAN [Internet]. Available: https://vine-hell-55c.notion.site/KEEPKWAN-7cdb13264ec14bd191e979e7fa3bbe0a.

[14] Civitai. Fine Tuning Parameter Model Sharing Site [Internet]. Available: https://civitai.com/.

[15] CNN Business. 'Verified' Twitter Accounts Share Fake Image of 'Explosion' Near Pentagon, Causing Confusion [Internet]. Available: https://www.donga.com/news/Inter/article/all/20230524/119446695/1

[16] Download Location. [Internet]. Available: https://edition.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion/index.html

[17] Adobe. Firefly for Adobe Stock Contributors [Internet]. Available: https://helpx.adobe.com/kr/stock/contributor/help/firefly-faq-for-adobe-stock-contributors.html

[18] Fast Company. Adobe Is So Confident Its Firefly Generative AI Won't Breach Copyright that It'll Cover Your Legal Bills [Internet]. Available: https://www.fastcompany.com/90906560/adobe-feels-so-confident-its-firefly-generative-ai-wont-breach-copyright-itll-cover-your-legal-bills.

[19] Adobe. Frequently Asked Questions About Generated Credits [Internet]. Available: https://helpx.adobe.com/kr/firefly/using/generative-credits-faq.html

[20] DALL·Ery GALL·Ery. DALL·E 2 Prompt Book [Internet]. Available: https://dallery.gallery/wp-content/uploads/2022/07/The-DALL%C2%B7E-2-prompt-book-v1.02.pdf.

**권동현(Dong-Hyun Kwon)**

1994년 : 서울대학교 미술대학 조소과 (미술학사)

2007년 : 연세대학교 영상대학원 (現커뮤 니케이션대학원) 영상디자인전공 (MFA, 예술학석사)

1996년~2000년: ㈜ON MEDIA (투니버스) Visual Arts Team Art Director

2000년~2002년: Video venture company ㈜Motion Factory Team Manager

2002년~2003년: ㈜TOONIPOP 콘텐츠 기획 제작 Team Manager

2006년~2009년: ㈜M-AGE Pictures 제작 이사

2010년~2019년: 경기대학교 애니메이션학과 교수

2020년~현 재: 경기대학교 애니메이션학과 초빙교수

※관심 분야 : 영상 편집, VR 특수영상, 생성 AI