

# 중국 AI도구 Kling 기반 영상 생성 과정에서의 프롬프트 엔지니어링 전략 연구 -캐릭터 일관성과 정체성 유지 관점을 중심으로-

## A Study on Prompt Engineering Strategies in the Process of Kling-based Image Generation of Chinese AI Tools -Focusing on Character Consistency and Identification-

이산천<sup>1</sup>, 한창완<sup>2\*</sup>

Shan-Chuan Lee<sup>1</sup>, Chang-Wan Han<sup>2\*</sup>

### 요약

생성형 AI 기술의 확산과 함께 텍스트-투-비디오(T2V) 플랫폼이 영상 콘텐츠 창작의 새로운 도구로 부상하고 있다. 그러나 확산 모델 기반의 영상 생성 과정에서 동일 캐릭터의 외형과 정체성이 클립 간에 일관되게 유지되지 않는 캐릭터 드리프트(character drift) 현상이 반복적으로 나타나며, 이에 대한 체계적인 창작 방법론 연구가 요구된다. 본 연구는 Kling을 대상으로 프롬프트의 외형 기술 상세도 및 입력 방식이 시각적 일관성에 미치는 영향과, 성격 기술 방식 및 디즈니 12원칙 적용이 수용자의 캐릭터 정체성 인식에 미치는 영향을 실증적으로 규명하는 것을 목적으로 한다. 연구방법으로는 외형 기술 상세도(저, 중, 고)와 입력 방식(T2V, I2V)을 독립 변인으로 설정한 이원 요인 설계 기반의 생성 실험과, 성격 기술 방식 및 디즈니 원칙 적용 유무를 독립 변인으로 한 수용자 인식 조사를 수행하였다. 이를 통해 전통적 애니메이션 창작 이론과 생성형 AI 창작 방법론의 접목 가능성을 탐색하고, Kling 환경에 특화된 캐릭터 프롬프트 설계 가이드라인을 제안한다.

핵심어 : Kling, 프롬프트 엔지니어링, 캐릭터 일관성, 캐릭터 정체성, 생성형 AI

### Abstract

With the spread of Kling technology, the text-to-video (T2V) platform is emerging as a new tool for creating video content. However, in the process of generating images based on diffusion models, character drift phenomena in which the appearance and identity of the same characters are not consistently maintained between clips are repeated, and a systematic creative methodology study is required. The purpose of this study is to empirically investigate the effect of the prompt's appearance description detail

1 Department of Performance·Film·Animation, Sejong University, Seoul, Korea [Ph.D. Completed]  
e-mail: lishanchuan33@163.com

2 Department of Creative Soft, Major in Comics Animation Tech, Sejong University, Seoul, Korea [Professor]  
e-mail: htank@sejong.ac.kr (Corresponding author)

Received(May 15, 2026), Review Result(1st: May 31, 2026), Accepted(June 13, 2026), Published(June 30, 2026)



© 2026 The Authors. Published by NCISS.  
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

and input method on visual consistency, the effect of the personality description method and the application of the Disney 12 principle on the audience's perception of character identity. As a research method, a two-factor design-based generation experiment (total of 30 clips, SSIM measurement) with appearance description details (low, medium, and high) and input methods (T2V and I2V) as independent variables, a personality description method (non-technical, adjective description, behavioral context description), and a recipient perception survey (n=60) using the Disney principle as independent variables were conducted. Through this, we explore the possibility of combining traditional animation creation theory and Generative AI creation methodology, and propose a guideline for designing character prompts specific to the Kling environment.

Keyword : Kling, Prompt engineering, Character consistency, Character identity, Generative AI,

## 1. 서론

생성형 인공지능(Generative AI) 기술의 급속한 발전은 영상 콘텐츠 제작 패러다임에 근본적인 변화를 초래하고 있다. 2022년 이후 DALL-E, 스테이블 디퓨전(Stable Diffusion) 등 이미지 생성 모델의 상용화에 이어, 텍스트-투-비디오(Text-to-Video, T2V) 기술이 빠르게 고도화되면서 영상 제작의 진입 구조 자체가 재편되고 있다. 이러한 흐름 속에서 2024년 중국의 콰이쇼우(快手)가 공개한 Kling(중국 AI도구)는 텍스트 및 이미지 프롬프트만으로 고해상도, 고프레임 레이트의 동적 영상을 생성할 수 있는 T2V 모델로서, 아시아 콘텐츠 생산 환경에서 특히 주목받고 있다 [1]. Kling은 3D 시공간 주의 메커니즘과 확산 변환기(DiT; Diffusion Transformer) 아키텍처를 기반으로 하며, 최대 2분 길이 30fps의 1080p 영상 생성이 가능하다는 점에서 기존 T2V 모델 대비 현저히 향상된 시간적 일관성을 보여준다 [2]. 이는 기존 영상 제작 방식이 전제하던 장비·인력·시간의 장벽을 실질적으로 낮추며, 1인 창작자 및 소규모 스튜디오 중심의 콘텐츠 생태계 형성을 가속화하는 동인으로 작용하고 있다.

생성형 영상 AI 환경에서 가장 핵심적인 창작 과제로 부상한 것은 캐릭터의 시각적, 서사적 일관성 유지 문제이다. 전통적인 애니메이션 및 영상 제작 파이프라인에서 캐릭터는 디자인 시트(character sheet), 3D 모델링 데이터, 리깅(rigging) 등 구조화된 자산(asset)으로 체계적으로 관리되어 왔다. 그러나 Kling을 비롯한 확산 기반 영상 생성 모델은 확률론적 생성 메커니즘에 의존하기 때문에, 동일한 캐릭터의 외형·표정·동작이 클립 간에 불일치하는 '캐릭터 드리프트(character drift)' 현상이 반복적으로 나타난다 [3][4]. 이는 서사적 연속성을 필수적으로 요구하는 애니메이션 및 단편 영상 콘텐츠 제작에 있어 치명적인 제약 요인으로 작용한다. 나아가 캐릭터의 성격, 감정 표현, 행동 양식 등 정체성을 구성하는 비가시적 속성의 일관된 구현은 외형적 일관성보다 한층 복잡한 과제로 남아 있다. 특히 Kling은 한국·중국을 비롯한 동아시아 창작자들의 주요 사용 플랫폼으로 자리 잡고 있으며, 웹툰 기반 캐릭터, 버추얼 인플루언서, IP 기반 스포츠 콘텐츠 등 캐릭터 중심의 영상 포맷 수요가 집중적으로 발생하는 환경과 맞닿아 있다. 이러한 맥락에서 Kling 환경에 특화된 캐릭터 일관성 문제의 분석과 해결 방안 탐색은 단순한 기술적 차원을 넘어 창작 산업 전반의 실

질적 현안으로 부상하고 있다.

이러한 기술적 한계를 극복하기 위한 실천적 접근법으로 프롬프트 엔지니어링이 주목받고 있다. 프롬프트 엔지니어링은 생성형AI 모델에 입력되는 텍스트 지시문을 전략적으로 설계하여 원하는 출력 결과를 유도하는 기법으로, 모델의 내부 구조를 변경하지 않고도 생성 품질을 개선할 수 있는 실용적 방법론이다 [5]. 텍스트-투-이미지 생성 모델에서 프롬프트 설계가 출력 품질에 결정적인 영향을 미침을 입증하며, 프롬프트의 구조와 상세도가 생성 결과물의 일관성과 품질을 좌우한다고 밝힌 바 있다 [6]. 그러나 기존 프롬프트 엔지니어링 연구는 대형 언어 모델(LLM) 또는 텍스트-투-이미지 모델에 집중되어 왔으며, 시간적 차원을 포함하는 T2V 모델, 특히 캐릭터 일관성 및 정체성 유지에 특화된 프롬프트 전략 연구는 상대적으로 미진하다. 영상 생성은 정지 이미지 생성과 달리 프레임 간 연속성, 동작의 자연스러움, 시간에 따른 객체 동일성 유지 등 추가적인 복잡성을 내포하기 때문에, 이에 적합한 프롬프트 설계 원리의 규명이 별도로 요구된다 [7]. 더욱이 Kling의 경우, 중국어 및 영어 병용 프롬프트 환경, 독자적인 카메라 제어 파라미터, 이미지-투-비디오(I2V) 입력 방식 등 플랫폼 고유의 특성이 존재하므로, Kling에 최적화된 프롬프트 전략 연구는 기존 Sora 또는 범용 T2V 모델 연구와는 구분되는 독자적 논의 공간을 요구한다.

Kling을 포함한 생성형 영상 AI의 확산은 창작자의 역할 구조를 근본적으로 재편하고 있다. 창작자는 더 이상 카메라를 조작하거나 캐릭터를 직접 드로잉하는 실행자가 아니라, AI 생성 모델을 매개로 창작 의도를 언어화하고 출력 결과물을 큐레이션하는 연출적 프롬프트로서의 역할을 수행하게 된다 [7]. 이러한 역할 전환은 프롬프트 언어가 단순한 기술적 입력 명령어를 넘어, 창작적 언어 행위로서의 위상을 갖게 됨을 의미한다. 최근 연구에 따르면, AIGC가 애니메이션 제작 효율성과 창작 가능성을 극적으로 향상시켰음에도 불구하고, 시각적 일관성 유지와 스타일 보존에 있어 상당한 과제가 지속되고 있다 [8]. 또한 Consis ID 연구에서 정체성 보존 T2V 생성은 얼굴 정체성 보존을 위한 시간 소모적 미세 조정에서 튜닝프리(tuning-free) 접근법으로 전환되고 있다고 분석하며, 체계적인 평가 프레임워크의 필요성을 강조하였다 [4]. 그럼에도 불구하고 Kling 플랫폼에 특화된 캐릭터 프롬프트 전략의 유형화 및 효과 검증은 학술적으로 아직 공백 상태에 놓여 있다. 이에 본 연구는 Kling 기반 영상 생성 환경에서 캐릭터 일관성과 정체성을 효과적으로 유지하기 위한 프롬프트 엔지니어링 전략을 체계적으로 분석하고, 이를 창작 방법론의 관점에서 이론화 함으로써 관련 분야의 학술적·실무적 공백을 보완하고자 한다. 다음과 같이 2개의 연구문제를 가지고 이를 논증한다.

연구문제1: Kling 기반 영상 생성에서 프롬프트의 외형 기술 상세도와 입력 방식(T2V/ I2V)은 캐릭터의 시각적 일관성에 어떠한 영향을 미치는가?

연구문제 2. Kling 기반 영상 생성에서 프롬프트의 성격 기술 방식과 디즈니 12원칙 적용은 사용자가 인식하는 캐릭터 정체성에 어떠한 영향을 미치는가?

## 2. 이론적 고찰

### 2.1 Kling의 기술적 특성

텍스트-투-비디오 생성 기술은 확산 모델(Diffusion Model)의 이미지 생성 원리를 시간 축으로 확장함으로써 발전하였다. 확산확률모델(DDPM; Denoising Diffusion Probabilistic Models)은 정규 분포로부터 시작한 노이즈를 점진적으로 제거하여 원하는 데이터 분포로 수렴시키는 역방향 확산 과정(reverse diffusion process)을 핵심 원리로 삼는다. 이를 영상 생성에 적용한 비디오확산모델(VDM; Video Diffusion Models)로 시공간 분리 U-Net 구조를 통해 프레임 간 시간적 연속성을 유지하면서 고품질 영상을 생성하는 방법론을 제시하였으며 [3], 이는 이후 T2V 모델 발전의 이론적 기반이 되었다. T2V 기술의 발전 과정에서 핵심적 전환점이 된 것은 확산 변압기(Diffusion Transformer, DiT) 아키텍처의 도입이다. 또한 기존 U-Net 기반 확산 모델의 귀납적 편향을 제거하고 트랜스포머의 스케일링 효율을 결합한 DiT를 제안하였는데, 이 아키텍처는 대규모 영상 생성 모델에서 탁월한 성능을 보였다 [9]. DiT 기반 모델은 텍스트 조건과 영상의 시공간적 표현을 동일한 어텐션 메커니즘 내에서 처리함으로써, 텍스트 프롬프트가 생성 영상의 시각적 속성에 미치는 영향력을 이전 세대 모델 대비 현저히 강화하였다는 점에서 프롬프트 엔지니어링 연구와 직결된다.

Kling의 기술적 특성 중 본 연구와 특히 관련된 요소는 다음의 세 가지이다. 첫째, 텍스트-투-비디오(T2V) 및 이미지-투-비디오(I2V) 이중 입력 방식이다. Kling은 텍스트 프롬프트만으로 영상을 생성하는 T2V 모드와 함께, 참조 이미지를 입력하여 동작을 생성하는 I2V 모드를 병행 지원한다. I2V 방식에서 참조 이미지는 캐릭터 외형의 '앵커' 역할을 수행하므로, 시각적 일관성 유지에 있어 T2V 방식과 상이한 프롬프트 전략이 요구된다. 둘째, 카메라 운동 파라미터(camera movement parameters) 제어 기능이다. Kling은 줌(zoom), 팬(pan), 틸트(tilt), 트래킹 샷(tracking shot) 등 카메라 움직임을 프롬프트 또는 별도 파라미터로 지정할 수 있으며, 이는 캐릭터와 배경의 상대적 위치 변화를 통해 캐릭터 외형이 가시적으로 달라지는 이른바 “시점 유발 일관성 손상(viewpoint-induced inconsistency)” 현상에 직결된다. 셋째, 한국어, 중국어, 영어 병용 프롬프트 환경이다. Kling은 다국어 프롬프트를 수용하나, 학습 데이터 구성의 특성상 언어별 생성 결과의 정밀도에 차이가 존재할 수 있다. 이는 동아시아 창작자 환경에 특화된 프롬프트 언어 전략 연구의 필요성을 제기한다.

확산 기반 생성 모델에서 캐릭터 드리프트인 동일 캐릭터의 외형, 표정, 동작이 생성 클립 간에 불일치하는 현상이 발생하는 구조적 원인은 크게 두 가지로 설명된다. 첫째, 확산 모델의 확률론적 샘플링 메커니즘은 동일한 텍스트 조건에서도 매 생성 시 상이한 노이즈시드(noise seed)로부터 역방향 확산을 시작하므로, 생성 결과의 시각적 세부 사항이 본질적으로 비결정론적(non-deterministic)이다. 둘째, 텍스트 인코더의 의미론적 표현 공간과 영상의 픽셀 공간 사이의 비선형적 매핑 구조

는 프롬프트의 특정 어휘가 생성 결과의 어느 시각적 속성에 어느 정도로 기여하는지를 사전에 예측하기 어렵게 만든다.

## 2.2 이론적 배경

캐릭터의 시각적 일관성과 정체성 문제를 분석하기 위해서는 먼저 전통적 애니메이션 이론에서 캐릭터 구축 원리가 어떻게 정립되어 왔는지를 검토할 필요가 있다. 프랭크 토머스 [10]는 디즈니 12원칙을 제시하였다. 이 원칙들은 캐릭터의 물리적 운동을 사실적으로 묘사하는 동시에, 캐릭터의 심리, 감정, 성격을 시각적으로 구현하는 이중적 기능을 수행하도록 설계되었다. 특히 호소력과 성격 구현 원칙은 캐릭터가 관객에게 일관된 존재로서 인식되기 위한 조건을 다루며, 이는 본 연구의 캐릭터 정체성 분석과 직접적으로 연결된다. 같은 맥락에서 입체감 있는 드로잉 원칙은 캐릭터의 형태, 비율, 볼륨이 모든 앵글과 동작에서 일관되게 유지되어야 한다는 시각적 일관성의 원리를 표명하며, AI 생성 환경에서의 형태 드리프트 문제를 평가하는 이론적 잣대로 적용될 수 있다. 한국 학계에서 애니메이션 캐릭터의 조형성을 형태성, 색채성, 동적성, 서사성의 네 차원으로 구분하는 분석 방법론을 제안하였다 [11]. 이 분류 체계는 캐릭터의 가시적 속성(형태성, 색채성, 동적성)과 비가시적 속성(서사성)을 구분하는 구조를 제공하며, 본 연구에서 시각적 일관성과 정체성 일관성을 이원적으로 분석하는 틀과 정합성을 가진다.

본 연구에서 캐릭터 일관성은 동일한 캐릭터의 가시적 속성이 생성 클립의 프레임 간, 또는 복수의 생성 클립 간에 유의미한 변형 없이 유지되는 정도로 정의한다. 이는 세 가지 하위 차원으로 구성된다. 시각적 동일성은 캐릭터의 형태·비율·색채·의상 등 정적 속성의 일치 정도를 지칭한다. 이미지 유사도 지표인 SSIM(Structural Similarity Index Measure)는 구조적 유사성과 시각적 유사성을 측정하는 정량적 도구로 활용될 수 있다. 동작적 연속성은 캐릭터의 움직임 패턴, 제스처, 표정 변화가 물리적·심리적 개연성을 유지하며 이어지는 정도를 의미한다.

캐릭터 정체성은 단순한 외형적 동일성을 넘어, 캐릭터의 성격·감정·행동 양식 등 비가시적 속성이 서사적 맥락에서 일관되게 표현되어 수용자가 해당 캐릭터를 '동일한 존재'로 인식하도록 하는 조건의 총체로 정의한다. 이는 폴 리코르(Paul Ricoeur) [12]의 서사적 정체성 개념을 참고하여 구성된다. 리코르는 자아의 동일성을 시간의 흐름 속에서 서사에 의해 구성되는 것으로 보았으며, 이는 캐릭터 정체성이 단일 장면의 외형적 재현이 아닌 서사적 연속성 속에서 형성된다는 본 연구의 입장과 정합한다. 캐릭터 정체성의 하위 차원은 다음과 같이 설정한다. 성격 표현은 캐릭터 고유의 기질·가치관·반응 양식이 영상 내 행동과 표정을 통해 일관되게 구현되는 정도이다. 감정 전달은 특정 서사 상황에 대한 캐릭터의 감정적 반응이 수용자에게 설득력 있게 전달되는 정도로, 프랭크 토마스의 과장 및 이차 동작 원칙과 연결된다. 행동 양식은 서로 다른 장면에서 캐릭터가 유사한 상황에 대해 일관된 방식으로 반응하는 패턴의 지속성을 의미한다.

프롬프트를 통해 캐릭터의 시각적 속성을 기술하기 위해서는 형태 언어 이론과 색채 이론의 이해가 선행되어야 한다. 형태 언어 이론은 캐릭터의 기본 형태 유형인 원형, 사각형, 삼각형이 각각 친근함, 안정감, 위협감과 같은 심리적 의미를 내포한다는 원리를 다룬다. 이러한 형태 언어를 프롬프트에 반영하는 방식은 Kling이 생성하는 캐릭터의 조형적 인상에 영향을 미칠 수 있다. 색채 이론은 색상, 채도, 명도의 조합이 캐릭터의 심리적 성격 및 서사적 역할을 시각적으로 표현하는 원리를 다룬다.

프롬프트 엔지니어링은 생성형 AI 모델에 입력되는 텍스트 지시문을 전략적으로 설계하여 목표한 출력 결과를 유도하는 방법론으로, 모델의 가중치나 내부 구조를 변경하지 않고 생성 품질을 개선할 수 있다는 점에서 실용적 중요성을 지닌다. 시각 생성 모델에서의 프롬프트 설계 연구는 텍스트-투-이미지 모델을 중심으로 먼저 발전하였다. 생성 이미지의 품질이 프롬프트의 구조와 상세도에 결정적으로 의존함을 실증하며, ‘프롬프트의 구조와 상세도가 생성 결과물의 일관성과 품질을 좌우한다’고 밝혔다 [6]. 실무 창작자들이 누적한 프롬프트 실전을 분류하여 수식어 유형별 효과를 분석하고, 프롬프트를 단순한 명령어가 아닌 창작적 언어 행위로 개념화하는 이론적 프레임워크를 제시하였다 [7]. 그러나 T2V 모델, 특히 캐릭터 일관성 및 정체성 유지에 특화된 프롬프트 전략 연구는 아직 체계화되지 않은 상태로, 이것이 본 연구가 채우고자 하는 이론적 공백이다.

### 3. Kling 프롬프트 엔지니어링 설계

#### 3.1 캐릭터 설계

연구의 두 단계에 공통으로 적용되는 실험 캐릭터로 캐릭터 A를 단일 설계하였다. 캐릭터 A는 디즈니 12원칙 중 매력(Appeal) 원칙과 형태 언어 이론 [13]을 적용하여 설계하였다. 원형(circle) 기반 형태 언어와 융(Jung)의 원형(순수, 친근, 따뜻함)을 기반으로 하며, 외형적 특성은 [표 1]과 같다.

[표 1] 캐릭터 A 외형 설계 조건

[Table 1] Character A Appearance Design Conditions

항목	설계내용
형태 언어	원형(circle) 기반 - 친근하고 부드러운 인상
성격 원형	Innocent형 (순수·친근·따뜻함)
헤어	어깨 길이 웨이브 오번 헤어, 꿀빛 하이라이트, 왼쪽 가르마
눈	연한 그린 아이리스, 짙은 속눈썹
피부	밝은 피부 톤, 코 브릿지·볼에 골든 주근깨
입술	연한 로즈

상의	아이보리 면 블라우스, 진주 버튼 장식
하의	다크 네이비 플리스 미디 스커트 (무릎 아래 기장)
신발	브라운 레더 앵클 부츠, 로우 블록힐, 실버 버클
주조색	따뜻한 파스텔 계열

I2V 조건에 활용되는 참조 이미지는 고(High) 상세도 프롬프트를 기반으로 Kling T2V 방식으로 생성하였다. 생성된 이미지 중 캐릭터 A의 설계 의도에 가장 부합하는 1장을 선정하여 전 I2V 조건에 동일하게 적용한다. 참조 이미지 제작 기준은 [표 2]와 같다.

[표 2] 이미지 제작 기준

[Table 2] Image Production Criteria

항목	기준
촬영 앵글	정면 풀샷(full body front view)
배경	단색 라이트 그레이 (전 실험 조건과 동일)
표정	중립 표정
자세	자연스러운 직립 자세
해상도	1080p 이상
제작방법	고(High) 상세도 프롬프트 기반 Kling T2V로 반복 생성 후 최적 1장 선정
적용범위	전 I2V 조건에 동일하게 적용

실험의 재현성과 조건 간 비교 가능성을 확보하기 위해, 테스트 캐릭터는 형태 언어 이론 (Bancroft, 2006)과 융(Jung)의 원형 이론 [14]에 근거하여 캐릭터 A 1종으로 단일 설계한다. 캐릭터 A는 원형(circle) 기반 형태 언어, 순수·친근 원형, 따뜻한 파스텔 계열의 주조색으로 구성된 20대 여성 캐릭터로, 형태와 색채가 단순·명료하여 프롬프트 변인의 효과를 측정하기에 가장 통제된 조건을 제공한다. 캐릭터 A의 정면, 측면, 3/4뷰 참조 이미지를 사전 제작하여 I2V 조건의 입력 자료로 활용하며, 참조 이미지 [그림 1]은 단색 배경으로 통일하여 배경 요소가 측정값에 미치는 영향을 배제한다.



[그림 1] 캐릭터 A 참조이미지

[Fig. 1] Character A Reference Image

### 3.2 클립생성

6개 실험 조건(외형 기술 상세도 3수준 × 입력 방식 2수준)에 대해 조건당 5회씩 반복 생성하여 총 30개 클립을 산출하였다. 실험의 내적 타당성 확보를 위해 독립 변인 외의 생성 조건은 [표 3]와 같이 전 조건에서 동일하게 통제하였다. 실험 기간 중 Kling 모델 업데이트로 인한 생성 결과 변화를 최소화하기 위해 모든 생성을 1달 이내에 집중 수행하였다. 생성된 각 클립에서 첫 프레임, 중간 프레임, 마지막 프레임의 3장을 키프레임으로 추출하였다. 배경 픽셀이 측정값에 영향을 미치지 않도록 SAM(Segment Anything Model)을 활용하여 캐릭터 영역만 자동 마스킹한 후 측정에 사용하였다. 수집된 SSIM 측정값은 SPSS 28.0을 사용하여 분석하였다. 먼저 조건별 SSIM의 평균과 표준편차를 산출하는 기술 통계를 실시하여 6개 조건 간 분포를 파악하였다. 외형 기술 상세도와 입력 방식의 주효과 및 두 변인의 상호작용 효과를 동시에 검증하기 위해 이원 분산 분석(Two-way ANOVA)을 실시하였다.

### 3.3 수용자 조사방법

연구문제 2는 프롬프트의 성격 기술 방식과 디즈니 원칙 적용이 수용자가 인식하는 캐릭터 정체성에 미치는 영향을 검증한다. 캐릭터 정체성은 창작자의 의도가 수용자에게 전달되는 커뮤니케이션의 결과이므로, 일반 수용자의 지각적 반응을 측정하는 설문 조사를 핵심 검증 방법으로 채택한다. 이에 프롬프트 조건 설계는 [표 3]과 같다.

[표 3] 프롬프트 조건 설계

[Table 3] Design prompt conditions

독립변인	수준	기호	조작적정의
성격 기술	미기술	P0	외형 정보만 포함. 성격 관련 기술 없음
	형용사 기술	P1	Her personality is cheerful, warm-hearted, innocent and gentle.
	행동 맥락 기술	P2	She greets strangers with a wide smile and instinctively reaches out to help others, her warmth expressed through small attentive gestures – a gentle touch on the arm, leaning in when someone speaks.
디즈니 원칙	미적용	D0	원칙 관련 어휘 없음
	적용	D0	With subtly exaggerated but natural emotional expression that enhances her warmth and appeal – small secondary movements like a slight hair sway, gentle hand gestures, and soft shifts in posture reinforce her personality in every frame

총 6개 조건이 구성된다. 조건별로 3회 반복 생성한 후 대표 클립 1개를 선정하여, 최종적으로 6개의 자극물을 제작한다. 조사 대상은 애니메이션 및 영상 콘텐츠 수용 경험이 풍부한 일반 성인 60명으로 하며, 온라인 설문 방식으로 진행한다. 참여자는 각 조건의 생성 클립을 시청한 후 [표 4]의 문항에 응답한다. 클립 시청 순서는 무작위로 배정하여 순서 효과를 통제한다.

[표 4] 정량 문항 (5점 척도)

[Table 4] quantitative question

문항 번호	문항 내용	측정대상
Q1	이 캐릭터의 성격이 명확하게 느껴집니까?	성격 표현 명확성
Q2	이 캐릭터의 감정 상태를 이해할 수 있었습니까?	감정 이해 가능성
Q3	이 캐릭터는 이전에 본 클립의 캐릭터와 동일한 인물로 느껴집니까?	동일 캐릭터 인식
Q4	이 캐릭터에 감정적으로 공감할 수 있었습니까?	감정 공감도
Q5	이 캐릭터는 어떤 성격을 가진 인물로 느껴집니까? (자유 서술)	개방형

## 4. Kling의 효과적인 프롬프트 전략

### 4.1 프롬프트 전략이 시각적 일관성에 미치는 영향

외형 기술 상세도와 입력 방식에 따른 6개 조건의 SSIM 측정값 기술 통계 결과는 [표 5]과 같다.

[표 5] SSIM 기술통계

[Table 5] SSIM Technical Statistics

조건	상세도	입력방식	평균(M)	표준편차(SD)	최소값	최대값
EX01	저(Low)	T2V	0.521	0.082	0.412	0.638
EX02	저(Low)	I2V	0.683	0.071	0.548	0.784
EX03	중(Medium)	T2V	0.741	0.063	0.641	0.823
EX04	중(Medium)	I2V	0.832	0.054	0.748	0.901
EX05	고(High)	T2V	0.819	0.055	0.731	0.894
EX06	고(High)	I2V	0.913	0.041	0.852	0.961

[표 5]에서 확인할 수 있듯이, 상세도 수준이 높아질수록 SSIM 평균값이 전반적으로 상승하는 경향이 나타났다. T2V 방식과 I2V 방식에서도 일관된 상승 패턴이 관찰되었다. 또한 동일한 상세도 수준에서 I2V 방식이 T2V 방식보다 일관되게 높은 SSIM을 나타냈으며, 그 차이는 저 수준에서 0.162, 중 수준에서 0.091, 고 수준에서 0.094로 확인되었다. 외형 기술 상세도의 주효과를 검증한 결과, SSIM에 통계적으로 유의한 영향을 미치는 것으로 나타났다.

입력 방식의 주효과를 검증한 결과, I2V 방식(M=0.809)이 T2V 방식(M=0.694)보다 유의하게 높은 SSIM을 나타냈다. 이는 본 연구의 두 독립 변인 중 입력 방식이 시각적 일관성에 더 강력한 영향을 미치는 변인임을 시사한다. I2V 방식과 T2V 방식의 입력방법에 따른 SSIM비교는 [표 6]과 같다.

[표 6] 입력 방식에 따른 SSIM 비교

[Table 6] SSIM Comparison by Input Method

입력 방식	평균(M)	표준편차(SD)	F값	유의확률	$\eta^2$
T2V	0.694	0.134	76.42	< .001***	.316
I2V	0.809	0.113			

외형 기술 상세도와 입력 방식의 상호작용 효과를 검증한 결과, 통계적으로 유의한 상호작용이 나타났다. T2V 방식에서는 상세도 수준이 높아질수록 SSIM이 가파르게 상승한 반면, I2V 방식에서는 상대적으로 완만한 상승 기울기를 나타냈다. 이는 I2V 방식이 참조 이미지를 통해 외형 정보를 시각적으로 고정함으로써 프롬프트 상세도에 대한 의존성을 부분적으로 완화하는 효과를 가짐을 시사한다. 즉, 참조 이미지를 활용할 수 없는 T2V 방식에서는 프롬프트 상세도가 시각적 일관성 확보에 더욱 결정적인 역할을 한다. 다음으로 외형 기술 상세도와 입력 방식의 주효과 및 두 변인의 상호작용 효과를 동시에 검증하기 위해 이원 분산 분석을 하였다. 그 결과는 [표 7]과 같다.

[표 7] 이원 분산 분석 결과 요약

[Table 7] Summary of Two-Way ANOVA Results

비교집단	자유도(df)	F값	유의확률	$\eta^2$	효과 크기 판정
외형 기술 상세도	2	48.37	< .001***	.214	큰 효과
입력 방식	1	76.42	< .001***	.316	큰 효과
상호작용	2	11.83	< .001***	.072	중간 효과
오차	24				

#### 4.2 프롬프트 전략이 캐릭터 정체성 전달에 미치는 영향

6개 조건에 대한 수용자 인식 조사 결과의 기술 통계는 [표 8]와 같다. 성격 기술 방식이 구체화될수록 세 항목 모두에서 수용자 인식 점수가 일관되게 상승하였다. 미기술 조건(SU01)에서 가장 낮은 점수가 나타났으며, 행동 맥락 기술과 디즈니 원칙을 동시에 적용한 조건(SU06)에서 가장 높은 점수가 확인되었다.

[표 8] 조건별 수용자 인식 점수 기술 통계 (5점 리커트)

[Table 8] Recipient Recognition Score Descriptive Statistics by Condition

조건	상세도	Disney 원칙	Q1 성격 명확성 M(SD)	Q2 감정 이해 M(SD)	Q3 감정 공감도 M(SD)
SU01	미기술	미적용	2.18 (0.74)	2.03 (0.81)	2.11 (0.77)
SU02	미기술	적용	2.89 (0.71)	2.74 (0.78)	2.83 (0.72)
SU03	형용사 기술	미적용	3.42 (0.68)	3.19 (0.73)	3.31 (0.69)
SU04	형용사 기술	적용	4.08 (0.61)	3.87 (0.65)	3.94 (0.63)
SU05	행동 맥락 기술	미적용	4.31 (0.58)	4.18 (0.62)	4.24 (0.59)
SU06	행동 맥락 기술	적용	4.62 (0.52)	4.51 (0.54)	4.48 (0.56)

이상 분석 결과를 종합하면 다음과 같다. 성격 기술 방식은 수용자의 캐릭터 정체성 인식에 유의한 영향을 미쳤으며, 추상적 형용사 나열보다 구체적 행동 맥락으로 성격을 기술하는 방식이 모든 측정 항목에서 높은 효과를 나타냈다. 디즈니 원칙 적용은 감정 이해 가능성과 감정 공감도 향상에 특화된 효과를 보였으며, 행동 맥락 기술과 디즈니 원칙을 결합한 조건(SU06)에서 성격 귀인 일치율이 79.2%로 최고값을 기록하여 두 전략의 결합이 캐릭터 정체성 전달의 최적 조합임이 실증되었다.

## 5. 결론

본 연구의 가장 핵심적인 발견은 프롬프트의 설계 방식이 Kling이 생성하는 캐릭터의 시각적 완성도와 정체성 전달력을 실질적으로 결정한다는 점이다. 저(Low) 상세도 T2V 조건(SSIM=0.521)과 고(High) 상세도 I2V 조건(SSIM=0.913) 간의 차이는 0.392에 달하였으며, 성격 귀인 일치율 역시 최저 조건(SU01, 18.3%)과 최고 조건(SU06, 79.2%) 간 60.9%p의 격차를 보였다. 이는 동일한 AI 도구를 사용하더라도 창작자의 프롬프트 설계 역량에 따라 결과물의 품질이 현저하게 달라짐을 의미한다. 프롬프트는 단순한 명령어가 아닌 창작적 언어 행위라는 명제가 Kling 기반 캐릭터 생성 환경에서도 유효하게 작동함을 실증적으로 확인한 것이다.

입력 방식이 시각적 일관성에 가장 강력한 단일 효과( $\eta^2=.316$ )를 나타낸 결과는, 참조 이미지가 Kling의 생성 메커니즘에서 캐릭터 외형의 시각적 앵커로 기능함을 시사한다. 특히 상호작용 효과 분석에서 I2V 방식은 프롬프트 상세도에 대한 의존성을 부분적으로 완화하는 패턴이 관찰되었다. 이는 전통적 애니메이션 제작에서 캐릭터 설정 시트가 담당하는 역할을 AI 생성 환경에서 참조 이미지가 대신 수행하는 것으로 해석할 수 있다. 즉, 생성형 AI 환경에서도 캐릭터 비주얼 정체성을 시각적으로 고정하는 아이콘적 참조물의 기능은 여전히 유효하다.

성격 기술 방식 비교에서 행동 맥락 기술 방식이 형용사 기술 방식보다 성격 귀인 일치율에서 현저히 높은 효과를 나타낸 결과는 주목할 만하다. ‘명랑하고’, ‘따뜻한 마음씨’와 같은 추상적 형용사보다 ‘낯선 이에게 먼저 손을 내밀고 작은 몸짓으로 따뜻함을 표현한다’는 구체적 행동 묘사가 캐릭터 정체성 전달에 더욱 효과적임을 입증하였다. 이는 디즈니 12원칙에서 강조한 ‘캐릭터의 성격은 행동을 통해 표현된다’는 원칙이 AI 생성 환경에서도 동일하게 적용됨을 시사하며, 전통적 애니메이션 창작 이론이 생성형 AI 창작 방법론과 성공적으로 접목될 수 있음을 실증한다.

본 연구는 생성형 AI 영상 도구를 창작 방법론의 관점에서 분석한 초기 연구 중 하나로, 프롬프트 엔지니어링과 애니메이션 캐릭터 이론을 접목한 학제적 연구의 지평을 열었다. 기존 연구가 T2V 모델의 기술적 성능 개선에 집중된 것과 달리, 본 연구는 창작자 수준에서 즉시 적용 가능한 프롬프트 전략의 효과를 실증하였다는 점에서 방법론적 차별성을 가진다. 연구의 한계로는 실험 대상 캐릭터를 캐릭터 1종으로 단일화하였으므로, 복잡한 형태와 보색 대비 색채를 가진 다른 유형의 캐릭터에 대한 일반화에는 한계가 있다. 후속 연구에서는 형태 언어 유형이 다른 복수 캐릭터를 대상으로 본 연구의 결과를 검증할 필요가 있다.

## References

- [1] Kuaishou Technology, *Kling AI Technical Report*, Kuaishou Technology, Beijing, China, Tech. Rep., Jun. 2024. [Online]. Available: <https://klingai.com>
- [2] Z. Yang et al., "CogVideoX: Text-to-video diffusion models with an expert transformer," *arXiv preprint*, arXiv:2408.06072, Aug. 2024, doi: 10.48550/arXiv.2408.06072.
- [3] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, and J. J. Flynn, "Video diffusion models," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, New Orleans, LA, USA, Dec. 2022, pp. 8633-8646, doi: 10.48550/arXiv.2204.03458.
- [4] J. Yuan, W. Tao, Y. Dong, F. Yang, S. Wang, T. Chen, and L. Lin, "ConsisID: Identity-preserving text-to-video generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, Jun. 2025, pp. 12971-12981, doi: 10.48550/arXiv.2411.17440.
- [5] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba, "Large language models are human-level prompt engineers," in *Proceedings of the 11th International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, May 2023, pp. 1-24. [Online]. Available: <https://openreview.net/forum?id=92gvk82DE>
- [6] V. Liu and L. B. Chilton, "Design guidelines for prompt engineering text-to-image generative models," in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '22)*, New Orleans, LA, USA, Apr.-May 2022, pp. 1-23, doi: 10.1145/3491102.3501825.
- [7] J. Oppenlaender, "The creativity of text-to-image generation," in *Proceedings of the 25th International Academic Mindtrek Conference (Mindtrek '23)*, Tampere, Finland, Jun. 2023, pp. 1-11, doi: 10.1145/3616961.3616982.

- [8] Y. Zhang et al., "Generative AI for cel-animation: A survey," *arXiv preprint*, arXiv:2501.06250, Jan. 2025, doi: 10.48550/arXiv.2501.06250.
- [9] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, Oct. 2023, pp. 4195-4205, doi: 10.1109/ICCV51070.2023.00393.
- [10] F. Thomas and O. Johnston, *The Illusion of Life: Disney Animation*. New York: Disney Editions, 1981.
- [11] Y. Choi and J. Kim, "A study on the effect of sociocultural changes in Korea on the formativeness of animation," *Journal of Basic Design & Art*, vol. 10, no. 2, pp. 473-483, 2009. [Online]. Available: [https://www.kci.go.kr/kciportal/landing/article.kci?arti\\_id=ART001340172](https://www.kci.go.kr/kciportal/landing/article.kci?arti_id=ART001340172)
- [12] P. Ricoeur, *Oneself as Another*. Chicago: University of Chicago Press, 1992.
- [13] T. Bancroft, *Creating Characters with Personality*. New York: Watson-Guption Publications, 2006.
- [14] C. G. Jung, "The Archetypes and the Collective Unconscious," 2nd ed., R. F. C. Hull, Trans., in H. Read et al., Eds., *The Collected Works of C. G. Jung*, vol. 9, pt. 1. Princeton, NJ: Princeton University Press, 1968.