생성형 AI 사극 영화 장면의 인물 복식 고증을 위한 평가 요소 도출 및 벤치마크 데이터셋 개발 연구

- 조선시대 후기를 중심으로

Keywords
: 생성형 AI 영화, 사극, 복식 고증 평가, 벤치마크, 도메인특화자동평가

선행연구
생성형 AI 영화의 배우 구현 유형 및 특성 연구
실사-AI 혼합 영화의 AI 영상 생성 시점에 따른 유형 및 특성 연구

연구자 : 박선주 (홍익대학교 AI·실감미디어콘텐츠학 박사과정)
지도교수 : 한정엽 (홍익대학교 AI·실감미디어 콘텐츠학)

목 차

제1장 서론

1.1. 연구 배경 및 목적
1.1.1. 연구 배경
1.1.2. 연구 문제
1.1.2. 연구 목적

1.2. 연구 범위 및 방법
1.2.1 연구의 범위
1.2.2 연구의 방법

제2장 이론적 배경

2.1. 생성형 AI 사극 영화와 인물 복식 고증
2.1.1. 생성형 AI 영화의 개념 및 특성
2.1.2. 사극 영화의 인물 복식 기능과 역사적 맥락
2.1.3. 사극 영화의 고증 개념과 인물 복식 부합도

2.2. 조선 후기 복식 고증의 기준과 평가 요소
2.2.1 복식 고증 평가를 위한 기준과 사료의 위계
2.2.2 복식사 분류와 신분별 착장 규범
2.2.3. AI 사극 영화 장면의 인물 복식 고증 평가 요소

2.3. AI 영상 평가 벤치마크와 복식 고증 평가 지표
2.3.1. 벤치마크 데이터셋 설계원리와 평가 기준
2.3.2. 전문가 판단 기반 복식 고증 평가
2.3.3. 도메인 특화 자동 평가 지표

제3장 연구 설계

3.1. 연구 방법 및 절차
3.1.1. 연구의 개요 및 흐름
3.1.2. 단계별 연구 수행 절차

3.2. 평가요소 도출 및 델파이 검증 설계
3.2.1. 인물 복식 고증 평가 요소 도출
3.2.2. 전문가 델파이 검증

3.3. KCCB 데이터셋 구축 설계
3.3.1. 프롬프트와 고증 기준 구성
3.3.2. 생성 장면 및 데이터셋 구성

3.4 전문가 평가 및 검증 설계
3.4.1. 전문가 평가 절차
3.4.2. 고증 평가 점수와 오류 유형 분석

3.5 사극 AI 영화 장면 고증 평가 설계
3.5.1. 사극 영화 선정
3.5.2. 작품별 복식 고증 오류 진단

제4장 연구 결과 (예상)

4.1. 평가 요소 도출·검증 결과
4.2. 데이터셋 구축 및 지표 개발 결과
4.3 전문가 평가 기반 부합성 검증 결과
4.4. 사극 AI 영화 장면 부합성 평가 결과
4.5. 종합 결과

제5장 결론

5.1. 연구의 요약 및 결과 논의
5.2. 연구의 시사점
5.3. 연구의 한계점 및 향후 연구

1.1.1. 연구 배경

산업적 배경

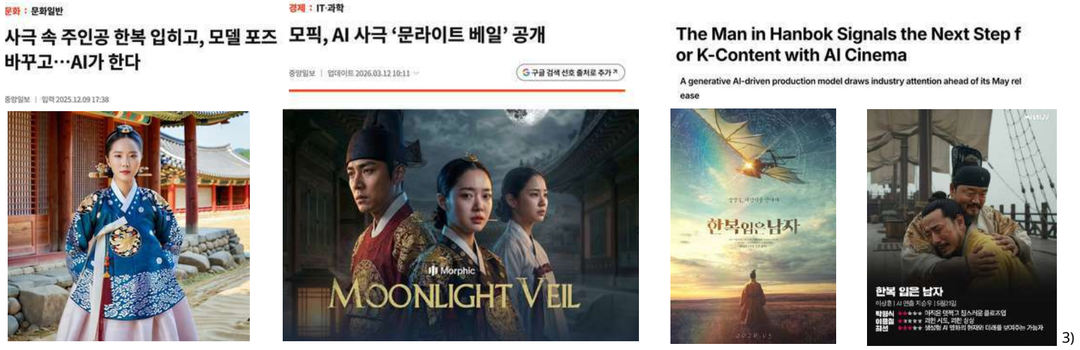
T2V 모델의 발전으로 AI 영화 부상 → 한국 사극 장르로 진입

- 확산모델 기반 T2I·T2V 생성 기술이 시각 콘텐츠의 생산·소비 구조를 근본적으로 재편 [1)]
- 'AI 영화'가 신생 매체로 부상 - 촬영 이전 단계에서 장면을 직접 합성 [2)]
- Kling·Veo 등 상용 AI 영상 모델을 활용하여 고증이 요구되는 사극 영화 제작 확산 [3)]

사회적 배경

생성형 AI의 한국 문화유산 왜곡과 복식 생성 모델의 빈번한 양식 혼동 오류

- 사극에서 복식은 시대·신분·성별·의례를 드러내는 핵심 시각 기호로, 재현 정확성이 역사적 신뢰성을 좌우 [4)]
- 확산 모델은 과거의 시대 양식을 평균화·정형화하여, 시대착오 등 역사적으로 부정확한 시각 표상을 산출 [5)]
- 한복은 생성 과정에서 중국·일본 양식과 혼효되기 쉽고, VLM 역시 한복을 오인·혼동하는 해석의 오류를 보임 [6)]



3)

7)

**확산모델 기반 T2I·T2V 생성 모델이 사극 장면 제작에 진입하면서 한국 복식이 인접 문화와 혼동·시대착오되는 오류가 빈번해지고 있으나,
이를 직접 진단하는 복식 고증 특화 평가 도구의 부재**

1) Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents.
Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models.
2) Zhang, R., Yu, B., Min, J., Xin, Y., Wei, Z., Shi, J. N., Huang, M., Kong, X., Xin, N. L., Jiang, S., Bahuguna, P., Chan, M., Hora, K., Yang, L., Liang, Y., Bian, R., Liu, Y., Campillo Valencia, I., Morales Tredinick, P., … Rao, A. (2025).
Generative AI for film creation: A survey of recent advances.
3) 최민지. (2025, 12월 9일). 사극 속 주인공 한복 입히고, 모델 포즈 바꾸고…AI가 한다. 중앙일보. https://www.joongang.co.kr/article/25388703
한아름. (2026, 3월 11일). 모픽, AI 사극 '문라이트 베일' 공개. 중앙일보. https://news.nate.com/view/20260311n29741
[Shin Jeong-sun]. (2026, 5월 27일). [AI Films Face 15-Second Wall, Require Thousands of Prompts]. The Chosun Daily.
https://www.chosun.com/english/travel-food-en/2026/05/27/YUUPRYQZ3VD7DHWTEKNKRSZNRM/
최선. 씨네21. (2026, 5월 20일). [리뷰] 생성형 AI영화의 현재와 미래를 보여주는 가늠자, '한복 입은 남자'. https://www.cine21.com/news/view/?mag_id=109975
4) 김유선. (2013). 영화의상. 커뮤니케이션북스 / 장민정. (2008). 한국 사극영화 복식의 표현성에 관한 연구 [박사학위논문, 성신여자대학교]
5) De Rosa Palmini, M.-T., & Cetinić, E. (2025). Synthetic history: Evaluating visual representations of the past in diffusion models.
6) Lee, H., Song, J., Jang, G., & Yoo, Y. (2025). Lost in translation: A position paper on probing cultural bias in vision-language models via Hanbok VQA.
7) 양형모. (2025, 2월 13일). 생성형 AI의 한국 문화 유산 왜곡 심각, 국가적 대응과 AI 대한민국 홍보대사 육성해야. 스포츠동아. https://sports.donga.com/world/article/all/20250213/131022581/1

1.1.1. 연구 배경

학술적 배경

AI 생성 평가 벤치마크의 확장과 한계 → 한국 복식 고증 도메인 특화 평가의 공백

- **국외연구 - AI 생성 평가 벤치마크 연구는 범용·문화·역사·진위 탐지에 국한되어 한국 복식·고증 도메인 특화 측정 부재**
  CUBE·CuRe (문화 역량) · Synthetic History (시대 표상) · T2VWorldBench (세계지식·문화 포함). [1)2)3)4)]

- **국내연구 - 사극 복식 고증 연구는 작품 단위의 질적 분석에 머물러 정량화된 평가요소의 부재, 척도화·벤치마크 데이터셋 연구로 확장되지 못함**
  <장희빈> · <스캔들> · <미인도> 등 작품 단위 고증 실태를 유물·문헌 대조로 분석, Lost in translation: Hanbok VQA로 VLM의 한복 인식 한계를 문제로 제기. [5)6)7)8)]

- **국내 도메인 특화 AI 평가 연구 - 광고 분야에서 생성형 AI 이미지 평가를 위한 도메인 특화 벤치마크와 자동평가 지표 개발 가능성을 제시**
  <광고사진 표현 형식 기반 생성형 AI 이미지 벤치마크 설계 및 상용 모델 성능 비교> [9)]

구분	연구 (연도)	대상	벤치마크 선행연구 내용 요약	본 연구의 시사점
범용 영상 평가	VBench (2024) / VBench-2.0 (2025)	T2V	화질, 시간 일관성, 프롬프트 충실도, 물리·상식·인체 충실도 평가	특정 문화권 복식 규범은 미측정
문화 역량	CUBE (2024) / CuRe (2025)	T2I	문화 인식, 문화 다양성, 문화 대표성 평가	한복 세부 착장과 시대 고증 기준 부족
역사 표상	Synthetic History (2026)	T2I	시대 양식, 시대착오, 역사 표상 왜곡 평가	조선 후기 복식의 법전·유물·도상 기반 판정 부재
한국 복식	Hanbok VQA (Lee et al., 2025)	VLM	한복 인식 한계를 문제로 제기	생성물의 복식 고증 평가 체계는 미제시
국내 사극 복식 연구	김주영(2003), 남수경(2004), 박가영(2010)	사극 영화 복식	작품 단위 복식 분석, 시대별 의상 변천, 영화의상 표현성 검토	AI 생성 장면 평가용 척도·데이터셋·정량 지표로는 미확장
국내 도메인 특화 평가	심현준 (2026)	T2I 광고 이미지	광고사진 표현형식 기반 AD-ImageBench, 전문가 2AFC, AD-LPIPS, AD-Score 제안	도메인 특화 벤치마크 설계 절차 참조 , 복식 고증 기준은 별도 재규정 필요

**국외 벤치마크는 범용 영상 품질·문화 역량·역사 표상 평가로 확장되었고, 국내에서도 광고 이미지의 도메인 특화 자동평가 시도가 있으나,
조선 후기 사극 영화의 인물 복식 고증을 위한 평가 요소와 벤치마크 연구는 부재**

1) Kannen, N., Ahmad, A., Andreetto, M., Prabhakaran, V., Prabhu, U., Dieng, A. B., Bhattacharyya, P., & Dave, S. (2024). Beyond aesthetics: Cultural competence in text-to-image models
Rege, A., Nie, Z., Ramesh, M., Raskar, U., Yu, Z., Kusupati, A., Lee, Y. J., & Vinayak, R. K. (2025). CuRe: Cultural gaps in the long tail of text-to-image systems
2) De Rosa Palmini, M.-T., & Cetinić, E. (2025). Synthetic history: Evaluating visual representations of the past in diffusion models (arXiv:2505.17064)
3) Chen, Y., Guo, X., Shi, Z., Song, Z., & Zhang, J. (2025). T2VWorldBench: A benchmark for evaluating world knowledge in text-to-video generation (arXiv:2507.18107).
4) Ghosh, D., Hajishirzi, H., & Schmidt, L. (2023). GenEval: An object-focused framework for evaluating text-to-image alignment
5) 김주영. (2003). 영상물 「장희빈」의 복식 분석 [석사학위논문, 이화여자대학교]
6) 남수경. (2004). 영화 〈스캔들〉 의상에 관한 연구 [석사학위논문, 중남대학교]
7) 박가영. (2010). 신윤복 풍속화와 영화 〈미인도〉 복식 비교 연구: 형태, 색채를 중심으로 [석사학위논문, 숙명여자대학교].
8) Lee, H., Song, J., Jang, G., & Yoo, Y. (2025). Lost in translation: A position paper on probing cultural bias in vision-language models via Hanbok VQA Workshops (pp. 4479–4486)
9) 심현준. (2026). 광고사진 표현 형식 기반 생성형 AI 이미지 벤치마크 설계 및 상용 모델 성능 비교 [Designing a representation form-based benchmark for generative AI images in advertising photography and comparing the performance of commercial models

1.1.2. 연구 문제

[연구문제 1]

생성형 AI 사극 영화 장면의 조선 후기 인물 복식 고증 평가 요소는 무엇이며, 어떻게 구성·검증 되는가?

1-1. 조선 후기 인물 복식 고증의 개념과 기준은 무엇인가?
1-2. 조선 후기 인물 복식 고증의 평가 요소는 어떻게 구성되는가?
1-3. 도출된 평가 요소는 전문가 내용 타당도 검증을 통해 신뢰도를 확보할 수 있는가?

[연구문제 2]

평가 요소 기반 벤치마크 데이터셋은 타당하게 구축·검증 될 수 있는가?

2-1. 평가 요소 기반 벤치마크 데이터 셋은 어떻게 설계·구축 되는가?
2-2. 구축된 데이터셋의 평가 항목과 코딩 기준은 전문가 다차원 평정을 통해 내용 타당도와 평가자 간 신뢰도를 확보할 수 있는가?

[연구문제 3]

개발된 데이터셋을 적용할 때, 사극 영화 장면의 인물 복식 고증 평가는 어떠한가?

3-1. 사극 영화 장면의 인물 복식 고증은 평가 요소별로 어떤 고증 부합도 수준을 보이는가?
3-2. 반복적으로 나타나는 복식 고증 오류는 무엇이며, 이는 향후 사극 AI 영화 평가 기준에 어떤 시사점을 제공하는가?

1.1.3. 연구 목적

**생성형 AI 사극 영화 장면의 조선 후기 인물 복식 고증 부합도 평가 요소를 도출·검증하고,
이를 기반으로 벤치마크 데이터셋과 자동 평가 지표를 개발하여
→ 적용 가능성과 한계를 실증적으로 검토하고자 함**

사료에 기반한 고증 평가 체계를 구성하고, 전문가 평정·자동 평가 지표를 설계하여
도메인 특화 벤치마크 데이터셋을 AI 사극 영화의 복식 고증 품질 진단에 적용한다.

사료 기반 복식 고증 평가 요소·기준 부재	생성형 AI 사극영화 인물 복식 고증 평가 요소 도출 + 전문가 델파이 검증
한국 복식 고증 도메인 특화 벤치마크 데이터셋 부재	한국 사극영화 조선 후기 인물 복식 데이터셋 KCCB(Korean Cinematic Costume Benchmark) 구축
한국 복식 고증 전문가 판단 기반 자동 지표 부족	전문가 2AFC 쌍대비교 판단을 학습신호로 활용한 KCCB-LPIPS 다차원 품질 평정값 예측 보조 지표 KCCB-Score 개발
생성형 AI 사극 영화 고증을 진단하는 평가 도구 부재	독립 데이터셋 적용·복식 고증 및 오류 진단

연구의 범위

구분	범위 내용
대상·시대적 범위	한국 AI 사극 영화 장면 인물 복식, 조선 후기
형식적 범위	T2V 5초 내외 장면 / SME(Start-Middle-End) 3프레임 캡처 기준 전문가 평가 프롬프트 500개 × 2개 모델 = 1000개 생성 장면 생성형 AI 모델 - 상용 2종 - Kling 3.0 · Veo.3 (동·서양 대표, 연구 시점 최신 모델 사용) 생성형 AI 사극영화 · 일반 사극 영화 · 사극 드라마
내용적 범위	복식 고증 평가요소 검증 - 요소 도출· 전문가 델파이 타당도·합의 검증 / KCCB 검증 : 전문가 고증 부합도 평정 · 자동평가 지표 학습 · 독립 테스트셋 적용
방법론적 범위	선행연구 분석·사료 연구(1차 평가요소 도출) - 수정 델파이 3라운드 (타당성 ·합의 검증) 데이터셋 구축·전문가 평가(2AFC 쌍대비교 + 다차원 평정 + 오류태그) - 자동지표 개발 - 소규모 독립 테스트셋 진단

연구의 방법

제1장 서론

연구의 배경 및 필요성
연구문제와
연구 목적
연구 범위 한정 및
연구 방법 수립

▶

제2장 이론적 배경

생성형 AI 사극 영화와
인물 복식 고증
조선 후기 복식 고증의
기준과 평가 요소
AI 영상 평가 지표와
복식 고증 부합도 개념 정의

▶

제3장 연구 설계

인물 복식 고증 평가 요소 도출
(델파이조사 1–3라운드)

연구 1 설계
한국 사극 인물 복식 고증 벤치마크(KCCB)구축
지표(전문가 평가·고증 부합도) 설계

연구 2 설계
독립 테스트셋 적용

▶

제4장 연구 결과(예상)

델파이 조사 결과

연구 1 결과
KCCB 데이터셋 구성 **신뢰도 검증**
KCCB·LPIPS, KCCB·Score 성능 검증

연구 2 결과
독립 테스트셋
고증 부합도 평가

종합 결과

▶

제5장 결론

연구의 요약 및
결과 논의
학술적·방법론적·
실무적 함의
후속 연구 제언

연구는 '평가요소 도출(델파이)'과 'KCCB 구축·지표 개발'을 연구 1로, '독립 테스트셋 진단'을 연구 2로 분리하여 순차 수행

2.1.1. 생성형 AI 영화의 개념 및 특성

생성형 AI 영화의 개념

- 생성형 AI 영화는 T2I·I2V·T2V 생성 모델을 활용하여 인물, 배경, 움직임, 스타일을 합성·변형·연결하는 영상 제작 방식[1)2)3)]
- 기존 영화 제작은 촬영 이후 편집·합성을 중심으로 제작되지만, 생성형 AI 영화는 촬영 이전 단계에서 프롬프트와 참조 이미지를 통해 장면을 직접 산출 [3)]
- AI를 부분 보조하는 ‘AI 활용 제작’부터 배우 촬영 없이 전 과정을 합성하는 ‘완전 생성(100% AI) 제작’이 있음 - 본연구는 완전 생성 AI 영화를 기준으로 함 [3)]

생성형 AI 영화의 특성

- **조건 기반 장면 합성**: 프롬프트·참조 이미지·시드·모델 버전이 장면 산출을 좌우하며, 동일 조건에서도 결과가 달라지는 비결정성을 보임 [1)2)]
- **반복 생성과 선별**: 생성형 AI 영화 제작은 단일 촬영이 아니라 다수 후보 장면의 생성·비교·선택·후반 조정으로 제작됨 [3)]
- **시간적 일관성의 한계**: 프레임·샷 사이에서 인물 정체성·복식 형태·색상이 변형되는 오류가 빈번하다 [4)]

5) 7)

**생성형 AI 영화는 T2I·I2V·T2V 생성 모델을 활용하여 프롬프트, 참조이미지 등으로 장면을 반복 생성·선별하는 제작 방식으로,
비결정성·드리프트로 인해 복식을 포함한 시각 요소의 일관성 오류가 빈번하여, 이를 점검할 체계적 접근이 요구된다.**

1) Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models.
2) Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S. W., Fidler, S., & Kreis, K. (2023). Align your latents: High-resolution video synthesis with latent diffusion models.
3) Zhang et al. (2025). Generative AI for film creation: A survey of recent advances.
4) Huang et al. (2024). VBench: Comprehensive benchmark suite for video generative models.
5) [이승미]. (2026, 4월 28일). ['100% AI 영화' 안방 넘어 극장으로...예술의 확장인가 진정성의 훼손인가]. 스포츠동아. https://sports.donga.com/ent/article/all/20260428/133827084/1
6) 박수빈. (2025, 12월 24일). 100% ‘AI 합성’ 영화 나온다...감독만 6명, ‘상업성’ 시험대. 한국경제. https://www.hankyung.com/article/202512243089g

2.1.2. 사극 영화의 인물 복식 기능과 역사적 맥락

사극 영화와 인물 복식

- 사극 영화는 특정 역사적 시대·사건·인물을 영화적으로 재구성하는 장르이며, 복식은 시대성과 인물 정보를 가장 직접적으로 전달하는 시각 요소 [1)2)]
- 관객은 인물의 복식을 통해 장면의 시대, 신분, 상황, 서사적 분위기를 직관적으로 해석 [2)3)]

인물 복식의 주요 기능

- 관객은 한 벌의 복식으로 '언제·누구·어떤 상황·무엇'인가를 즉각 독해 [4)]

시대 지시 기능 “언제인가”
복식만으로 배경 시대가 분별되는 시간 좌표 기능
(박가영, 2010)

신분·역할 표지 기능 “누구인가”
신분·지위·계급·성격의 상징적 암시
(김유선, 2013; 권혁산, 2012; 김미경·이은주, 2020)

상황·의례 표지 기능 “어떤 장면인가”
의례·공무·일상 등 상황에 적합한 착장 요구
(조미숙, 2005)

심미·서사 기능 “무엇을 말하는가”
인물 심리·감정·극 전개를 보조하는 표현적 변형
(장민정, 2007)

역사적 맥락

- 조선 후기 복식은 신분·성별·의례·상황에 따른 착장 규범을 지니며, 사극 영화는 이를 바탕으로 고증과 영화적 변형 사이의 적정 범위를 구성 [5)6)]
- 따라서 복식 고증은 단순한 형태 비교가 아니라, 장면 속 인물·시대·상황 맥락에 부합하는지를 판단하는 평가 차원 [6)7)]

사극 영화에서 인물 복식은 시대, 신분, 상황, 서사를 드러내는 핵심 시각 요소이며, 복식 기능은 작품의 역사적 신뢰성을 구성한다.
→ 인간 제작 사극뿐 아니라 생성형 AI가 합성한 장면에서도 검토되어야 할 평가 차원

1) 김유선. (2013). 영화의상. 커뮤니케이션북스
2) 조미숙. (2005). 영상물 〈장희빈〉 복식의 고증실태와 분석 [석사학위논문, 전남대학교]
3) 박가영. (2010). 신윤복 풍속화와 영화 〈미인도〉 복식 비교 연구: 형태, 색채를 중심으로 [석사학위논문, 숙명여자대학교]
4) 권혁산. (2012). 조선시대 무관초상화와 흉배에 관한 연구
5) 김미경, 이은주. (2020). 조선시대 공신초상(功臣肖像)의 관복(제1기) 고찰
6) 장민정. (2008). 한국 사극영화 복식의 표현성에 관한 연구: 프로덕션 디자이너 정구호의 작품을 중심으로
7) De Rosa Palmini, M.-T., & Cetinić, E. (2025). Synthetic history: Evaluating visual representations of the past in diffusion models (arXiv:2505.17064)

2.1.3. 사극 영화의 고증 개념과 인물 복식 부합도

사극 영화의 고증 개념

- **사극 영화의 고증은 역사적 사실의 단순 복원이 아니라, 사료에 근거한 영화적 재현 과정** [1)2)]
  사극은 특정 시대·인물·사건을 영화적으로 재구성하는 장르이며, 고증은 장면의 시대성·사회적 맥락·시각적 설득력을 확보하기 위한 기준으로 작동
- **인물 복식은 시대·신분·상황을 식별하게 하는 핵심 시각 요소** [3)4)]
  복식은 인물의 신분, 성별·연령, 역할, 의례적 상황을 드러내는 미장센 요소이며, 사극 장면의 역사적 신뢰성을 구성하는 주요 고증 대상
- **영상 매체의 고증은 '완전 복원'이 아니라 '허용 가능한 재현 범위' 안에서 판단** [5)]
  사극 의상은 사료에 근거해야 하지만, 영화적 연출·장면 맥락·가시성의 요구를 함께 고려 → 고증 평가는 단일 정답 이미지와의 일치보다, 사료 기반 기준에 대한 '부합 정도'를 판단해야 함

인물 복식 고증 부합도 [조작적 정의]

인물 복식 고증 부합도 = 생성형 AI 사극 영화 장면 속 인물 복식이
사료 기반 조선 후기 복식 고증 기준(Ground Truth: GT)과 **프롬프트에 명시된 장면 조건**에 부합하는 정도

*GT는 단일 정답 이미지가 아니라 사료 기반 전문가 판단 기준점

사극 영화의 고증은 사료 기반 역사 재현과 영화적 허용 범위가 결합된 개념
→ 본 연구에서 '인물 복식 고증 부합도'는 생성형 AI 사극 장면의 복식이 조선 후기의 시대·신분·상황·착장 기준에 얼마나 부합하는지를 판단하는 도메인 특화 평가 개념으로 정의

1) 박가영. (2010). 신윤복 풍속화와 영화 〈미인도〉 복식 비교 연구: 형태, 색채를 중심으로 [석사학위논문, 숙명여자대학교]
2) 신다혜, 나현신. (2016). TV 사극드라마 의상에 관한 기초연구
3) Roach-Higgins, M. E., & Eicher, J. B. (1992). Dress and identity. Clothing and Textiles Research Journal
4) 남수경. (2004). 영화 〈스캔들〉 의상에 관한 연구 [석사학위논문, 충남대학교].
5) 임정미. (2017). 한국 사극 의상 디자인에서 보여지는 재매개 연구: 2000년대 이후 TV 사극을 중심으로 [박사학위논문, 중앙대학교]

2.2.1. 복식 고증 평가를 위한 기준과 사료의 위계

조선 후기 복식 고증의 기준

- **복식 고증은 단일 이미지와의 일치가 아니라, 사료에 근거한 허용 가능한 고증 범위를 설정하는 과정**[1)2)] → 장면 안에서 시대성·신분성·상황성이 설득력 있게 유지되는지를 판단하는 기준
- **조선 후기 복식 고증은 제도·규범 자료, 실물·유물 자료, 도상·시각 자료의 교차검증에 근거** [3)] → 사료는 "누가 무엇을 입었는가", "실제 의복은 어떤 형태였는가", "장면 속에서 어떻게 착용되었는가"를 서로 보완

고증 사료	검증 내용	고증 기준으로서의 역할
제도·규범 (법전·의궤·실록)	**누가·언제 무엇을 입는가**	신분, 품계, 관복, 의례 복식의 규범 기준 제공
실물·유물 (출토복식·박물관 소장품현전 유물)	**실제 의복은 어떤 형태로 존재했는가**	형태, 색채, 문양, 재질의 물질적 기준 제공
도상·시각 자료 (초상화·풍속화·기록화)	**실제 장면에서 어떻게 착용되었는가**	착장 방식, 생활 맥락, 신분별 시각 표지의 기준 제공

[고증 사료의 층위]

2.2.2 복식사 분류·용어와 신분별 착장 규범

- **선행연구에서 반복되는 복식 분석 차원 : 형식 차원 - 형태 · 색채 · 문양 · 소재/재질 · 착장/구성 · 두식/관모 | 맥락 차원 - 신분/품계 · 상황/용도 · 인물 역할** [4)]
  → 이 범주들은 1960년대 이후 영상 사극 복식 연구에서 작품·인물군이 달라도 반복 관찰되는 분석 차원

[5)]

기준(GT: Ground Truth)의 개념
하나의 정답 이미지가 아닌 **법전·의궤·실록, 출토복식·박물관 유물, 초상화·풍속화·기록화를 교차 검토해 설정한 사료 기반 고증 기준점**
→AI 생성 장면은 이 기준점에 따라 **< 부합 / 부분 부합 / 불일치>**의 수준에서 평가

조선 후기 복식 고증의 기준점은 단일 정답 이미지가 아니라, 제도·규범 / 실물·유물 / 도상·시각의 세 사료를 교차 검토하여 고증 기준(GT)을 구성
→ 본 연구는 이 기준을 바탕으로 다음 절에서 AI 사극 영화 장면의 인물 복식 고증 평가요소를 도출

1) De Rosa Palmini, M.-T., & Cetinić, E. (2026). Synthetic history: Evaluating visual representations of the past in diffusion models
2) Qadri, R., Díaz, M., Wang, D., & Madaio, M. (2025). The case for "thick evaluations" of cultural representation in AI. 김월계. (2009). 사극 드라마에 나타난 조선시대 궁녀복식에 관한 연구강지현 (2009)
3) 김월계. (2009). 사극 드라마에 나타난 조선시대 궁녀복식에 관한 연구,장민정, & 심화진. (2007). 조선시대를 배경으로 한 영화에서 보여지는 귀부녀의 복식에 관한 연구, 한경희. (2004). 영화 〈청풍명월〉의 갑주(甲冑) 연구, 전여선. (2007). 한국 영화와 TV드라마 전통복식에 나타난 색의 변화와 이미지의 상징성
4) 김선. (2016). 1960년대 이후 사극영화에 나타난 왕실 여성의상 연구, 남수경. (2004). 영화 〈스캔들〉 의상에 관한 연구, 조미숙. (2005). 영상물 〈장희빈〉 복식의 고증실태와 분석, 김주영. (2003). 영상물 〈장희빈〉의 복식 분석
5) De Rosa Palmini & Cetinić (2025); Qadri et al. (2025)

2.2.3. AI 사극 영화 장면의 인물 복식 고증 평가 요소

선행 연구(복식사·영상 사극·고증 평가)의 교차·수렴 - 고증층 → 평가 요소 → 조작적 정의 ▶ 관찰지표 / **핵심 고증 평가축 5개 + 영상 보조 검증축 1개 도출**

생성형 AI 사극 영화 장면의 인물 복식 고증 평가 요소

인물 복식 고증 부합도

조선 후기 인물 복식을 법전·유물·도상 사료의 위계로 '입어야 할 바'로 규정하고, AI 생성 장면의 고증 부합도를 평가하는 요소

분류	고증층	내용	평가요소	내용 (조작적 정의)	관찰지표	참고문헌
인물 복식 고증 부합도 평가축	제도·규범 (법전·의궤)	'누가·언제 무엇을 입는가' 법전·의궤가 규정한 시대·신분 질서로 색인되는 규범 (경국대전·가례도감의궤)	시대	인물 복식이 조선 후기 시간 범위에 부합하고, 타시대·근현대·타문화 요소가 부적절하게 혼재하지 않는 정도	시대착오 요소, 근현대 양장화, 중국·일본풍 혼종, 판타지화	김선(2017); 조미숙(2005)
			신분·역할	인물의 신분, 품계, 성별·연령, 직역에 맞는 복식 등급과 표지가 구현된 정도	관복/편복 구분, 관모, 흉배, 품계 표지, 궁녀·내관·기녀·무관·양반·상민 표지	김주영(2003); 김미경·이은주(2020)
	형식·물성 (현전 유물)	'무엇을 어떻게 갖춰 입었는가' 현전 유물·도판과 직접 대조하는 물질적 재현	형식·구성	의복의 기본 형태, 실루엣, 품목 조합, 착장 방식이 조선 후기 복식 자료와 부합하는 정도	일습 완비(저고리·치마·포·관모·두식) 형태·품목 조합·착장 방식 (결손·오대체 검출)	김선(2017); 한경희(2004)
			색채·문양·재질	색채·문양·직물이 신분·상황 규범에 부합하는 정도 (사료 신뢰도 낮음→명백 위반만 감점)	신분색·금지색 위반, 문양 과잉/오용, 흉배 오류, 금박·장식 과잉, 재질감 불일치	전여선(2007); 박가영(2010)
	행위·맥락 (도상·의궤)	'언제·어떤 상황에 입었는가' 도상·의궤 기반 맥락 (초상화·풍속화 자료)	상황·용도	장면의 상황(일상·공무·의례·외출·상장)에 적합한 복식이 선택·착용된 정도	일상복/의례복/공무복/외출복/노동복 구분, 혼례·상례·궁중·장터·노동 장면 적합성	김월계(2009)
영상 보조 검증축	(영상 지속성)	고증 총점에서 제외하고 별도 보고하는 영상 차원의 시간적 지속성	복식 일관성 (총점 제외)	동일 인물의 복식 구조와 주요 표지가 5초 클립 안에서 의도 없이 변형·소실·전이되지 않고 프레임 간 시간적으로 일관되게 유지되는 정도	동일 인물 복식 지속 착장 구조 유지(드리프트·모핑 부재)	VBench(2024); VBench-2.0(2025)

김선. (2017). 1960년대 이후 사극영화에 나타난 왕실 여성의상 연구 [석사학위논문, 성균관대학교]
조미숙. (2005). 영상물 〈장희빈〉 복식의 고증실태와 분석 [석사학위논문, 전남대학교].
김주영. (2003). 영상물 〈장희빈〉 의 복식 분석 [석사학위논문, 이화여자대학교]
김미경, 이은주. (2020). 조선시대 공신초상(功臣肖像)의 관복(제1기) 고찰
한국 영화와 TV드라마 전통복식에 나타난 색의 변화와 이미지의 상징성 [석사학위논문, 연세대학교].

박가영. (2010). 신윤복 풍속화와 영화 〈미인도〉 복식 비교 연구: 형태, 색채를 중심으로 [석사학위논문, 숙명여자대학교]
김월계. (2009). 사극 드라마에 나타난 조선시대 궁녀복식에 관한 연구
Huang, Z., He, Y., Yu, J., Zhang, F., Si, C., Jiang, Y., ... Liu, Z. (2024). VBench: Comprehensive benchmark suite for video generative models.
Zheng, D., Huang, Z., Liu, H., Zou, K., He, Y., Zhang, F., Gu, L., Zhang, Y., He, J., Zheng, W.-S., Qiao, Y., & Liu, Z. (2025). VBench-2.0: Advancing video generation benchmark suite for intrinsic faithfulness

2.3.1. 벤치마크 데이터셋의 설계원리와 평가 기준

벤치마크 설계원리

- **공통 설계 원리 = 다차원·위계적 분해** → 모델 성능을 단일 미학 점수로 통합하지 않고, 화질·일관성·정합성처럼 **세부 차원으로 쪼개 진단해야 모델의 강점·약점·오류 패턴이 드러남** [1)]
- **평가체계의 요건** - 표성·포괄성·균형성·재현가능성·접근성을 충족해야 '비교 가능한' 평가가 성립 [2)]
- **객체·요소 단위 검증** → 장면을 통째로 보지 않고 '객체의 존재·개수·색·위치'처럼 요소 단위로 코드화해 점검 [3)]

이미지 벤치마크 (지표가 세대적으로 진화) **: 픽셀 → 구조 → 지각 → 분포 → 의미로 발전 / 영상 벤치마크 (이미지 + 시간축)**

프롬프트 집합 설계 → 모델별 영상 생성 → 프레임/클립 단위 특징 추출 → 차원별 지표 자동 채점(필요 시 인간 평가나 VLM, 시각-언어 모델) 판정을 결합) → 차원 점수를 가중·집계 → 모델 간 순위와 오류 패턴 보고

1세대 - 표면적 충실도
분포 거리·표면 정합: FID(Heusel et al., 2017), CLIPScore(Hessel et al., 2021) · 화질·시간 일관성·프롬프트 정합(Huang et al., 2024)
→ **"영상이 그럴듯한가"**

▶

2세대 - 내재적 충실도
물리 법칙·상식·인체 구조·구성 일관성의 5차원 18능력 (Zheng et al., 2025)
→ **"세계 원리에 부합하는가"**

▶

본 연구 - 도메인 특화 고증 부합도
특정 문화권의 역사 규범을, 사료 기반 GT·참조 이미지로 시대·신분·품목 등의 평가요소 단위 고증 부합으로 측정
→ **"조선 후기 복식 고증 기준에 맞는가"**

벤치마크(연도)	대상	평가 차원	본 연구의 시사점
GenEval (2023)	T2I	객체의 존재·개수·색·위치 - 요소·속성 단위 검증(Ghosh et al., 2023)	요소 단위 코드화 형식 차용 → 복식 품목·속성으로 전이
VBench (2024)	T2V	화질·시간 일관성·프롬프트 정합의 계층적 16+차원(Huang et al., 2024)	분해 원리 차용 - 복식 고증 항목은 부재
VBench-2.0 (2025)	T2V	내재 충실도: 인체·물리·상식·제어·창의 5차원 18능력(Zheng et al., 2025)	'표면적 그럴듯함' 극복 - 한국 복식 규범은 미측정
T2VWorldBench (2025)	T2V	세계 지식 6범주(문화 포함)·1,200 프롬프트(Chen et al., 2025)	문화 범주 포함하나 복식 고증의 세분 기준 없음
CUBE (2024) · CuRe (2025)	T2I	문화 역량(인식·다양성)·장기꼬리 문화 격차(Kannen et al., 2024; Rege et al., 2025)	문화 일반 - 조선 복식의 신분·시대는 미세분
Synthetic History (2026)	T2I	HistVis 30,000장 - 암묵적 양식 주입·시대착오·인구 표상(De Rosa Palmini & Cetinić, 2026)	역사 정확성·시대착오 진단 → 평가요소(시대 부합성)의 직접 근거
VISTAR (2025) · Hanbok VQA (2025)	T2I·VLM	델파이(120명)·역할 기반 하이브리드 채점 / 한복 인식·thick evaluation(Jiang et al., 2025; Lee et al., 2025)	KCCB의 방법 모델·직접 선례 - '생성물 채점'은 미수행
AD-ImageBench (2026)	T2I	광고사진 표현형식 기반 AD-ImageBench, 전문가 2AFC, AD-LPIPS, AD-Score 제안 (심현준, 2026)	도메인 특화 벤치마크 설계, 자동지표 학습 절차 참조
KCCB (본 연구)	**AI 사극 영상**	**사극 영화 장면의 인물 복식 고증 평가 (사료 기반)**	**한국 조선 후기 복식 고증 도메인 특화 · GT 기반 진단형**

AI 이미지·영상 평가 벤치마크는 '표면적 그럴듯함 → 세계 원리 부합 → 도메인 특화·역사 규범 부합으로 진화
도메인 특화 평가는 기준 이미지(GT) 구조를 전제하며, 본 연구는 조선 후기 인물 복식의 고증 부합을 진단형 평가로 구체화

1) Huang, Z., He, Y., Yu, J., Zhang, F., Si, C., Jiang, Y., Zhang, Y., Wu, T., Jin, Q., Chanpaisit, N., Wang, Y., Chen, X., Wang, L., Lin, D., Qiao, Y., & Liu, Z. (2024). VBench: Comprehensive benchmark suite for video generative models.
2) Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021). AI and the everything in the whole wide world benchmark.
3) Ghosh, D., Hajishirzi, H., & Schmidt, L. (2023). GenEval: An object-focused framework for evaluating text-to-image alignment

2.3.2 전문가 판단 기반 복식 고증 평가

복식 고증 평가와 전문가 판단

- **학습 신호로서의 인간 판단** - 학습 기반 지각 지표·생성형 AI 평가에서 인간 판단은 사후 검증을 넘어 모델 학습 신호로 직접 활용 [1)]
- **전문가 영역의 판단 -** 일반 사용자 선호로는 복식 고증 품질 측정이 부족하며, 복식사·영화의상·영상미술· AI 영상관련 전문가의 판단 영역 [2)]
- **쌍대 비교 (2AFC)의 안정성 -** 절대 점수 부여보다 두 자극의 상대적 차이를 판단하는 방식이 더 안정적 [3)]
- **내용타당도 기반 평가 요소 선정** - 수정 델파이로 도출한 복식 고증 평가 요소를 내용타당도 비율(CVR)로 선별하되, 패널 수별 임계값으로 채택 기준을 적용 [4)]
- **합의 신뢰도의 정량 검증** - 전문가 간 합의의 신뢰도를 정량 지표로 검증 [5)6)]

전문가 판단 자료의 유형과 학습 신호로서의 기능

쌍대 비교 자료 (2AFC)

후보 A **vs** 후보 B → **상대적 고증 부합 판단**

→ 상대적 고증 부합 판단
→ **KCCB-LPIPS 학습 신호** (Thurstone, 1927; Bradley & Terry, 1952)

다차원 평정 자료 (요소별)

시대 | **신분·역할** | **형식·구성** | **색채·문양·재질** | **상황·용도**

→ 시대 · 신분·역할 · 형식·구성 · 색채·문양·재질 · 상황·용도 (평가요소)
→ **KCCB-Score 학습 신호** 요소별 절대 점수·프로파일 (Krippendorff, 2018)

복식 일관성 (총점 제외)

복식 고증 평가는 전문가의 쌍대 비교(2AFC)와 요소별 다차원 평정을 학습 신호로 삼는다.
KCCB는 수정 델파이로 타당화한 복식 고증 평가요소를 축으로 전문가 판단을 자동 지표의 학습·검증 기준으로 구성한다.

1) Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric.
2) 김미경, 이은주. (2020). 공신초상의 관복 고증
3) Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons
4) Lawshe, C. H. (1975). A quantitative approach to content validity
5) Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters
6) Krippendorff, K. (2018). Content analysis: An introduction to its methodology
7) Thurstone, L. L. (1927). A law of comparative judgment

2.3.3. 도메인 특화 자동 평가 지표

기술적 구조

- **자동 평가 지표의 한계** → 자동 지표는 지각·분포·비전-언어 의미 기반 세 계열로 발전해 왔지만, 복식 고증 규범을 직접 반영하도록 학습된 지표는 부재 [1)2)3)]
- **도메인 재보정** →범용 지표를 식별된 불일치 조건에 맞춰 재보정하는 것은 확립된 방법론 → 전문가 2AFC 기반 도메인 재보정의 근거 [4)5)6)]

자동 평가 지표

지각 기반 지표 · LPIPS
심층 특징 거리로 인간 지각 유사도를 근사 (Zhang et al., 2018)
→ 사전학습 신경망의 심층 특징 공간에서의 거리로 인간 지각 유사도를 예측

의미 기반 지표 · CLIPScore
이미지–텍스트 의미 정합을 참조 없이 평가 (Hessel et al., 2021)
→ CLIP 임베딩의 이미지–텍스트 코사인 유사도로 의미 정합을 참조 없이 측정

본 연구(KCCB)에 사용하는 지표

KCCB-LPIPS · 지각 차이 기반 재보정
참조 스틸(GT)·생성 프레임 → CNN 특징 추출 + 전문가 2AFC 재보정
→ 채널 가중치 재학습 → 변별 점수 (Ghazanfari et al., 2023; Ghildyal & Liu, 2022)

KCCB-Score · 다차원 품질 예측
CNN 시각 특징 + 비전-언어 의미 임베딩 + 복식 도메인 특징 → 회귀 모델 (심현준, 2026)
→ 복식 고증 요소별 평정값 예측

범용 지표는 도메인 특화 복식 고증 규범을 직접 반영하지 못함
→ KCCB는 전문가 2AFC로 재보정한 KCCB-LPIPS와 평가 요소를 예측하는 KCCB-Score로 참조 이미지(GT) 대비 도메인 특화 자동 평가를 구성

1) Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric.
2) Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium
3) Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., & Choi, Y. (2021). CLIPScore: A reference-free evaluation metric for image captioning
4) Ghazanfari, S., Garg, S., Krishnamurthy, P., Khorrami, F., & Araujo, A. (2023). R-LPIPS: An adversarially robust perceptual similarity metric
5) Ghildyal, A., & Liu, F. (2022). Shift-tolerant perceptual similarity metric
6) 심현준 (2026) 광고사진 표현 형식 기반 생성형 AI 이미지 벤치마크 설계 및 상용 모델 성능 비교

3.1.1. 연구의 개요 및 흐름

연구 1

KCCB **지표 개발 · 내부 성능 검증**

자료	KCCB 데이터셋 - 층화 프롬프트 500 × 2모델 = 1,000 **장면** + GT 카드, 참조이미지
전문가	2AFC **쌍대 비교** + **평가요소**(E1-E6) 7**인 다차원 평정**(0·1·2) + 오류 태깅
산출물	KCCB-LPIPS (2AFC 학습) + KCCB-Score (평가요소 평정 결과 학습)

연구 2

독립 테스트셋 적용 · 외부 검증

자료	**실제 사극 영상** 6**편** (**사극** AI**영화** 2 **· 일반 사극 영화** 2 **· 사극 드라마** 2 (학습 미참여)
적용지표	KCCB-LPIPS · KCCB-Score + 범용 LPIPS · CLIPScore
검증	**복식 전문가** 3**인의** 2AFC **다수결** (연구 1 미참여)

RQ1

1. **평가 요소 도출·델파이**

문헌연구 · 수정 델파이 3R · 평가요소(E1-E6)

RQ2

연구1

2. KCCB **구축·전문가 평가**

프롬프트 500×2모델 · 2AFC · E1-E6 평정· 오류 태깅

연구1

3. **지표 학습·내부 검증**

KCCB-LPIPS(2AFC) · KCCB-Score(E1–E6) · 내부 성능 검증

RQ3

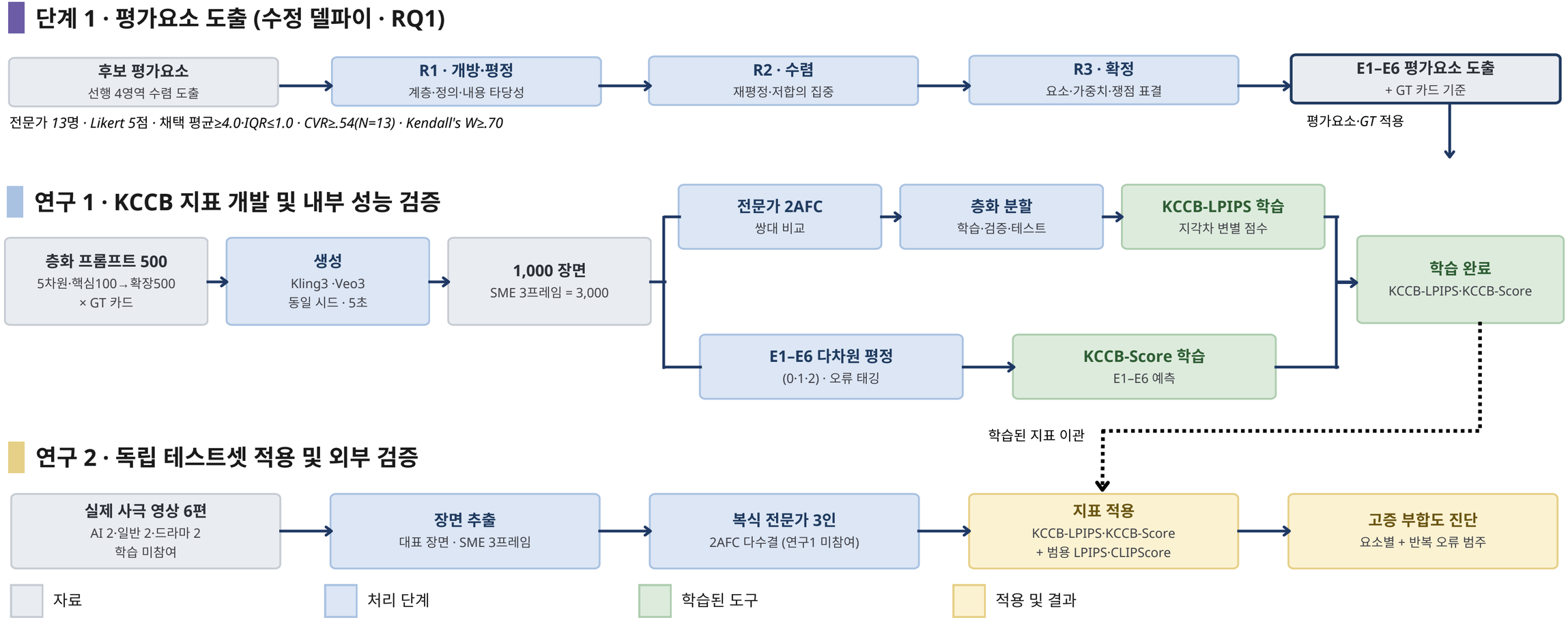
연구2

4. **독립 테스트셋 적용**

사극 영상 6편 · 범용 지표 비교 · 전문가 3인 2AFC

본 연구는 생성형 AI 사극 영화 장면의 인물 복식 고증 부합도를 평가하기 위해, 복식사·영상 사극 연구·고증 평가·AI 벤치마크 선행연구에 기반하여
문헌연구 - 평가 요소 도출 - 전문가 델파이 검증 – 고증 벤치마크(KCCB) 구축 – 전문가 평가 – 독립 테스트셋 진단의 절차를 설계

3.1.2. 단계별 연구 수행 절차



[KCCB 연구의 흐름] - 평가요소 도출 · 연구 1 · 연구 2

* 연구는 순차 수행되며 각 단계는 선행 단계의 산출물을 입력으로 받는다.

1. 조사 개요

조사 목적

선행 영역 수렴으로 도출한 후보 평가 요소의 채택·세부 항목·요소별 가중치를
전문가 합의로 검증·확정

조사 인원 (Ayre & Scally, 2014)

복식사·고증·역사·AI영화·사극의상·벤치마크 전문가
총 13명 구성된 패널 n ≥ 13.
- 복식사·고증 전문가 3인 (김OO 외)
- 사극 영화/드라마 의상 전문가 3인 (양OO 외)
- 생성형 AI·영화 전문가 4인 (백OO 외)
- 벤치마크·평가방법론 전문가 3인 (심OO 외)

전문가 선정 기준

1. 조선 복식사·고증 또는 생성형 AI 영상·평가 분야에서 전문 경력 보유
2. 사극 영화 의상 감독 및 실무 전문 경력을 보유
3. 해당 분야 논문·작품·프로젝트 등 전문 성과 보유

수정 델파이 채택 근거 (Hsu & Sandford, 2007)

본 연구는 선행 연구를 바탕으로 도출된 평가 요소를 1차안으로 제시하여 라운드 수와 응답 부담을 통제하고 합의 수렴을 가속한다.

2. 수정 델파이 라운드별 운영

라운드	조사 목적	주요 평가 항목
R1 개방·평정	고증층–평가 요소–세부 항목 구조의 적절성·내용 타당성 검증, 삭제·수정·추가 제안 의견 수렴	복식 고증 평가 요소에 대한 타당성 항목 구성의 계층·정의·내용 적절성 (5점 평정) 삭제·수정·추가 자유 제안, 개방응답 → 수정안 반영
R2 수렴	1R 통계 피드백 후 재평정 · 저합의·신규 항목 집중 검증 및 심층 의견 수렴	1차운드 통계 피드백 제공 합의 수준 낮은 항목 집중 평가 피드백 기반 신규 항목 검증
R3 확정	최종 평가 요소·가중치 확정 및 쟁점 표결 → 평가요소 기반 GT시안 확정	최종 평가 평가항목 합의 도출 평가자별 이견 항목 종합

3. 분석 기준

항목	기준	선행연구
Likert 5점 척도	5(매우 적합) ~ 1(전혀 부적합)	Likert (1932)
채택 · 합의도	**평균 ≥ 4.0 · IQR ≤ 1.0**	Hsu & Sandford (2007)
내용타당도 (CVR)	N=13 → CVR ≥ .54 (정확이항 임계값; 항목당 ≥10/13 '필수')	Ayre & Scally (2014)
순위 합의 (Kendall's W)	**W ≥ .70** (강한 합의; 중등도 .50 대비 상향)	Schmidt (1997)

4. 수정 델파이 조사 설계

생성형 AI 사극 영화 장면의 인물 복식 고증 평가 요소 도출				
구분	**평가요소**	**내용 (조작적 정의)**	**관찰지표**	**내용 적절성**
E1	**시대**	인물 복식이 조선 후기 시간 범위에 부합하고, 타시대·근현대·타문화 요소가 부적절하게 혼재하지 않는 정도	시대착오 요소, 근현대 양장화, 중국·일본풍 혼종, 판타지화	1~5점
E2	**신분·역할**	인물의 신분, 품계, 성별·연령, 직역에 맞는 복식 등급과 표지가 구현된 정도	관복/편복 구분, 관모, 흉배, 품계 표지, 궁녀·내관·기녀·무관·양반·상민 표지	
E3	**형식·구성**	의복의 기본 형태, 실루엣, 품목 조합, 착장 방식이 조선 후기 복식 자료와 부합하는 정도	일습 완비 (저고리·치마·포·관모·두식) 형태·품목 조합·착장 방식 (결손·오대체 검출)	
E4	**색채·문양·재질**	색채·문양·직물이 신분·상황 규범에 부합하는 정도 (사료 신뢰도 낮음→명백 위반만 감점)	신분색·금지색 위반, 문양 과잉/오용, 흉배 오류, 금박·장식 과잉, 재질감 불일치	
E5	**상황·용도**	장면의 상황(일상·공무·의례·외출·상장)에 적합한 복식이 선택·착용된 정도	일상복/의례복/공무복/외출복/노동복 구분, 혼례·상례·궁중·장터·노동 장면 적합성	
E6	**복식 일관성 (총점 제외)**	동일 인물의 복식 구조와 주요 표지가 5초 클립 안에서 의도 없이 변형·소실·전이되지 않고 프레임 간 시간적으로 일관되게 유지되는 정도	동일 인물 복식 지속 착장 구조 유지 (드리프트·모핑 부재)	

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation
Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. Practical Assessment, Research, and Evaluation
Lawshe, C. H. (1975). A quantitative approach to content validity
Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology
Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques

3.3.1. 프롬프트와 고증 기준 구성

프롬프트는 평가 요소 기반으로 균형있게 표본을 구성하고,
GT 카드를 이원화 하여 '텍스트 규범(허용 대역·인간 채점용)'과 '참조 스틸 1장(대표값·자동지표 참조용)'으로 작성

참조 스틸은 허용 대역 내 대표값을 구현한 표준 포맷 1장으로, 모델 산출물을 보기 전 복식사 검수로 확정하여 독립성과 누출 차단을 보장하며,
근거는 1차 사료·전문가 인증에 둔다.

1) 프롬프트 - 500개

(평가 요소 조합을 균형 있게 포함하도록 프롬프트를 표본 구성)

시대(조선 후기)	**신분·역할**
성별·연령	**상황·의례**
복식 요소 초점	

2) 고증 정답지(GT 카드) - 프롬프트와 1:1 대응 구조

텍스트 규범 - 허용 대역 · 인간 평정 기준

- 신분·상황 규정 [법전·의궤] · 필수 품목 일습 [유물·도판]
- 색·문양·재질 허용 대역 [도상·문헌] · 맥락·예법 [의궤·예서]
- 착장·실루엣 기준 · 전형적 오류 메모

참조 스틸 1장 - 대표값 · KCCB-LPIPS의 참조(ref)

- **역할:** KCCB-LPIPS의 참조 이미지(ref)
- **위상:** 인간 채점을 근사하는 상대 변별 보조기
- 표준 포맷 : 정면·전신·중성 회색 배경·균일 조명·단일 인물
- 모델 생성 전 복식사 검수·동결 - 동일 플랫폼 생성 금지 (누출 차단)

3.3.2. 생성 장면 및 데이터셋 구성

동일 프롬프트로 통제한 상용 모델 2종에 500개 프롬프트로 1,000장면을 생성하고,
장면당 SME 3프레임(총 3,000장)을 평가 입력으로 삼아 프롬프트→생성→전문가 평정→진단의 파이프라인으로 KCCB 데이터셋을 구성

생성 모델·조건

동서양 모델 2종 고정 - Kling 3.0 · Veo3 (연구 시점 최신 모델)
동일 프롬프트·동일 시드 통제

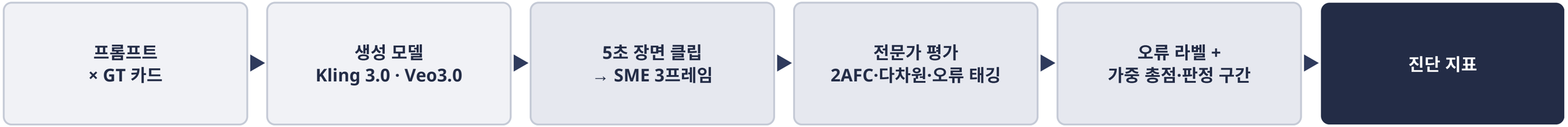
분석 단위

5초 내외 장면 클립
- 시작·중간·끝(SME) 3프레임 캡처 이미지를 평가 입력으로 사용
  (전수 시청 부담 통제 + 드리프트 포착)

데이터셋 규모

500 × 2모델 = 1,000 장면 (SME 프레임 → 3,000 장)

KCCB 데이터 구조 및 처리 흐름



***** 상용 모델의 비결정성은 동일 시드 고정, 다중 생성·표본 확장으로 관리 · 데이터셋 구성·한계·용도를 명시적으로 문서화(데이터시트 투명성; Gebru et al., 2021)

Gebru, T., et al. (2021). Datasheets for datasets. Communications of the ACM

3.4.1. 전문가 평가 절차

평정단 13인, 평가요소(GT 카드) 숙지 → ① 2AFC 강제선택 · ② 요소별 다차원 평정(0·1·2) · ③ 오류 유형 태깅을 병행하여, 자동 지표·진단을 위한 신호를 분리 생성

양식	절차	산출 · 통계	학습 · 검증 신호
2AFC 강제선택	동일 프롬프트·GT에서 더 고증 정합적인 장면을 선택(좌우 무작위·모델명 비공개)	선택 비율 → 소프트 라벨 → 쌍대비교 척도화	→ **KCCB-LPIPS**
요소별 다차원 평정	각 장면을 평가요소 기반 관찰지표로 0·1·2 평정	요소 점수 · 가중 총점 · 판정 구간	→ **KCCB-Score**
오류 유형 태깅	고증 오류를 유형 라벨로 다중 부착 (시대·신분·형식·색채·상황)	오류 빈도 · 유형 분포	→ **오류 진단 해석**
복식 일관성 점검	영상 클립 내 동일 인물 착장 지속 점검 (총점 제외)	드리프트 · 모핑 빈도	→ **영상 보조 진단**

신뢰도 확보 (IRR) 및 일치도 해석

평정 신뢰도

순위형 평정(요소별 0·1·2 다차원 평정) : Krippendorff α로 평가자 간 신뢰도 점검

명목형 판단(2AFC 선택·오류 유형 태깅) : Fleiss κ로 평가자 간 일치도 점검

전문가 판단은 2AFC(쌍대 선호) · 요소별 평정(요소 품질)· 오류 태깅(진단)으로 수집되어
각각 KCCB-LPIPS·KCCB-Score·오류 진단의 학습·검증 신호로 분리 사용

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs
Thurstone, L. L. (1927). A law of comparative judgment
Krippendorff, K. (2018). Content analysis
Fleiss, J. L. (1971). Psychological Bulletin,
심현준. (2026). 광고사진 표현 형식 기반 생성형 AI 이미지 벤치마크 설계 및 상용 모델 성능 비교

3.4.2. 전문가 평가와 자동 지표(KCCB-LPIPS · KCCB-Score) 학습 설계

전문가 2AFC 선호를 학습신호로 LPIPS(VGG16)의 채널 가중치를 재보정하여 KCCB-LPIPS를 만들고, VGG16·CLIP·요소 특징을 결합한 회귀로 KCCB-Score를 학습

KCCB-LPIPS (2AFC 학습)

- **원리: 전문가가 더 고증에 맞다고 고른 쪽에 지표의 거리 판단을 일치시킴 (참조에 가까울수록 = 더 부합)**
- **학습 신호:** 전문가 2AFC 선호 비율(y) → p_A = σ(d_B − d_A), BCE 손실
- **학습 방식** 범용 LPIPS(VGG16)는 고정 → 가중치 층(lin 층)만 재보정 학습 → 적은 데이터로 도메인 적응
- **입력 통제**: 복식 외 요소(배경·피부·얼굴) 제거 → 거리값이 복식 차이만 반영
- 전처리: 참조 장면이미지(ref)·모델 A·B 프레임에 동일 전처리 - 배경 제거→의복 영역 마스킹→정렬→ 256px

KKCCB-Score (회귀)

시각 VGG16 | **의미 CLIP** | **복식 평가 요소**

- **원리: 세 특징을 결합한 회귀로 고증 총점·요소별 점수 동시 예측**
- **검증:** 교차검증으로 전문가 평정과의 상관 확인
- 가중 총점·요소 프로파일 동시 예측

전처리 → 강도 결정 → 재보정 학습

① 정규화 전처리
배경 제거 → 의복 영역 마스킹(얼굴·맨살 제외) → 정렬 → 256px

▶

② 전처리 강도 결정
전처리 단계별 비교 실험을 통해
검증셋 2AFC 정확도가 가장 높은 조건 채택

▶

③ KCCB-LPIPS 재보정
VGG16 동결 + lin 층만 전문가 2AFC 선호로 학습

**LPIPS 학습을 위한 참조 스틸과 생성 프레임의 도메인 격차는 정규화 전처리와 2AFC 상대 비교로 통제하고, 전처리 강도는 단계별 비교 실험을 통해 정하며,
VGG16을 동결한 채 lin 층만 전문가 2AFC 선호로 재보정**

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric
Li, P., Xu, Y., Wei, Y., & Yang, Y. (2022). Self-correction for human parsing
심현준. (2026). 광고사진 표현 형식 기반 생성형 AI 이미지 벤치마크 설계 및 상용 모델 성능 비교
Zhang et al. (2018). LPIPS. CVPR. · Radford et al. (2021). CLIP. ICML. · Hessel et al. (2021). CLIPScore. EMNLP
Li et al. (2022) SCHP; Kirillov et al. (2023) SAM

**학습에 쓰지 않은 독립 자료(신규 프롬프트 + 실제 사극 6편)에 학습된 KCCB 지표를 적용하고,
미참여 전문가 3인 2AFC 다수결을 정답으로 삼아 지표의 외부 적용 가능성과 한계를 진단**

3.5.1 독립 테스트셋 구성

- **동일 조건 원칙**

  연구 1과 동일한 프롬프트 ·GT 카드·5초 단위·SME 3프레임·평가 입력 포맷으로 구성

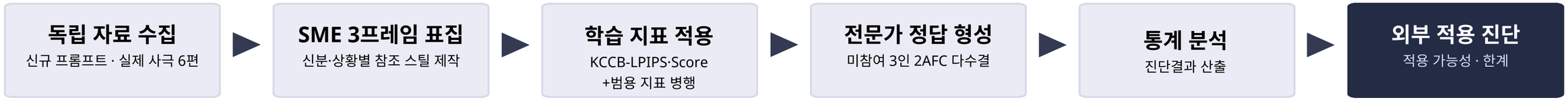
  → 측정 일관성·재현성 확보
- **독립 생성셋:** 학습 미사용 신규 프롬프트를 동일 조건으로 자동 생성, 약 100 장면(장면당 SME 3프레임)
- 실제 영상물(6편) : AI 사극 영화 · 일반 사극 영화 ·사극 드라마

  → 복식 고증이 가능한 장면을 SME 3프레임으로 표집
- 참조 스틸 각 작품 신분·상황별 참조 스틸을 동일 규약으로 제작해 KCCB-LPIPS 산출

3.5.2 학습된 지표 기반 진단 [1)]

- 진단 지표 KCCB-LPIPS · KCCB-Score(요소 종합) - 기준선 범용 LPIPS · CLIPScore 병행
- 주 분석 모델·범주 요인 혼합설계 분산분석 + 비교 세트별 우세 모델 판정
- 전문가 대조 미참여 전문가 3인 2AFC 다수결을 정답으로, 지표 판정과의 일치도를 확인 [2)]
- 분석 모델·범주 요인 혼합설계 + 진단 결과 산출

독립 테스트셋 진단 흐름



설계 원칙 - 학습 미참여 자료로 외부 타당도 확인 · 쏠림·우연을 보정해 단순 일치율의 착시 방지 · 데이터셋 구성·한계·용도의 명시적 문서화 [3)]

1) Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric
Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., & Choi, Y. (2021). CLIPScore: A reference-free evaluation metric for image captioning. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing
2) Thurstone, L. L. (1927). A law of comparative judgment / Cohen, J. (1960). A coefficient of agreement for nominal scales / Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme
3) Gebru, T., et al. (2021). Datasheets for datasets

5. 참고문헌

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. Measurement and Evaluation in Counseling and Development, 47(1), 79–86.
Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S. W., Fidler, S., & Kreis, K. (2023). Align your latents: High-resolution video synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika, 39(3/4), 324–345.
Brooks, T., Peebles, B., Holmes, C., DePue, W., Guo, Y., Jing, L., Schnurr, D., Taylor, J., Luhman, T., Luhman, E., Ng, C., Wang, R., & Ramesh, A. (2024). Video generation models as world simulators. OpenAI.
Candy, L., & Edmonds, E. (2018). Practice-based research in the creative arts: Foundations and futures from the front line. Leonardo, 51(1), 63–69.
Chen, Y., Guo, X., Shi, Z., Song, Z., & Zhang, J. (2025). T2VWorldBench: A benchmark for evaluating world knowledge in text-to-video generation (arXiv:2507.18107) [Preprint]. arXiv. https://arxiv.org/abs/2507.18107
Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.
De Rosa Palmini, M.-T., & Cetinić, E. (2026). Synthetic history: Evaluating visual representations of the past in diffusion models. In Proceedings of the International Conference on Learning Representations (ICLR 2026).
Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378–382.
Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92.
Ghazanfari, S., Garg, S., Krishnamurthy, P., Khorrami, F., & Araujo, A. (2023). R-LPIPS: An adversarially robust perceptual similarity metric (arXiv:2307.15157) [Preprint]. arXiv.
Ghildyal, A., & Liu, F. (2022). Shift-tolerant perceptual similarity metric. In Proceedings of the European Conference on Computer Vision (ECCV 2022). Springer.
Ghosh, D., Hajishirzi, H., & Schmidt, L. (2023). GenEval: An object-focused framework for evaluating text-to-image alignment. Advances in Neural Information Processing Systems (NeurIPS 2023).
Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., & Choi, Y. (2021). CLIPScore: A reference-free evaluation metric for image captioning. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (pp. 7514–7528). Association for Computational Linguistics.
Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In Advances in Neural Information Processing Systems (NeurIPS 2017).
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.
Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65–70.
Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. Practical Assessment, Research & Evaluation, 12(10), 1–8.
Huang, K., Huang, Y., Wang, X., Lin, Z., Ning, X., Wan, P., Zhang, D., Wang, Y., & Liu, X. (2025). FilMaster: Bridging cinematic principles and generative AI for automated film generation (arXiv:2506.18899) [Preprint]. arXiv.
Huang, Z., He, Y., Yu, J., Zhang, F., Si, C., Jiang, Y., Zhang, Y., Wu, T., Jin, Q., Chanpaisit, N., Wang, Y., Chen, X., Wang, L., Lin, D., Qiao, Y., & Liu, Z. (2024). VBench: Comprehensive benchmark suite for video generative models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 21807–21818).
Jiang, K., Sun, R., Cao, Y., Xu, Y., Zhang, X., Guo, J., & Deng, C. (2025). VISTAR: A user-centric and role-driven benchmark for text-to-image evaluation (arXiv:2508.06152) [Preprint]. arXiv.
Kannen, N., Ahmad, A., Andreetto, M., Prabhakaran, V., Prabhu, U., Dieng, A. B., Bhattacharyya, P., & Dave, S. (2024). Beyond aesthetics: Cultural competence in text-to-image models. In Advances in Neural Information Processing Systems (Vol. 37, pp. 13716–13747).
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4015–4026).
Krippendorff, K. (2018). Content analysis: An introduction to its methodology (4th ed.). SAGE Publications.
Lawshe, C. H. (1975). A quantitative approach to content validity. Personnel Psychology, 28(4), 563–575.
Lee, H., Song, J., Jang, G., & Yoo, Y. (2025). Lost in translation: A position paper on probing cultural bias in vision-language models via Hanbok VQA. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 4479–4486).
Li, P., Xu, Y., Wei, Y., & Yang, Y. (2022). Self-correction for human parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6), 3260–3271.
Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 22(140), 5–55.
Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA) – Protein Structure, 405(2), 442–451.
Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77.
Qadri, R., Díaz, M., Wang, D., & Madaio, M. (2025). The case for "thick evaluations" of cultural representation in AI. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 8(3), 2067–2080.
Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (PMLR 139, pp. 8748–8763).
Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021). AI and the everything in the whole wide world benchmark. Advances in Neural Information Processing Systems Datasets and Benchmarks Track (NeurIPS 2021).
Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents (arXiv:2204.06125) [Preprint]. arXiv.
Rege, A., Nie, Z., Ramesh, M., Raskar, U., Yu, Z., Kusupati, A., Lee, Y. J., & Vinayak, R. K. (2025). CuRe: Cultural gaps in the long tail of text-to-image systems. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 15680–15691).
Roach-Higgins, M. E., & Eicher, J. B. (1992). Dress and identity. Clothing and Textiles Research Journal, 10(4), 1–8.
Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684–10695).
Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. Decision Sciences, 28(3), 763–774.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420–428.
Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review, 34(4), 273–286.
Yadav, S., Tilton, L., Antoniak, M., Arnold, T., Li, J., Pawar, S. M., Karamolegkou, A., Frank, S., An, Z., Rostamzadeh, N., Hershcovich, D., Belongie, S., & Shutova, E. (2025). Cultural evaluations of vision-language models have a lot to learn from cultural theory (arXiv:2505.22793) [Preprint]. arXiv.
Zeng, Y., Dong, F., Zhao, J., Zheng, P., Li, J., & Zhou, H. (2025). Towards culturally fair multimodal generation: Quantifying and mitigating Orientalist biases in text-to-visual models. In Proceedings of the 33rd ACM International Conference on Multimedia.
Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 586–595).
Zhang, R., Yu, B., Min, J., Xin, Y., Wei, Z., Shi, J. N., Huang, M., Kong, X., Xin, N. L., Jiang, S., … Rao, A. (2025). Generative AI for film creation: A survey of recent advances. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVEU).
Zheng, D., Huang, Z., Liu, H., Zou, K., He, Y., Zhang, F., Gu, L., Zhang, Y., He, J., Zheng, W.-S., Qiao, Y., & Liu, Z. (2025). VBench-2.0: Advancing video generation benchmark suite for intrinsic faithfulness (arXiv:2503.21755) [Preprint]. arXiv.

강성식. (2012). 조선시대를 배경으로 제작된 사극의 뷰티디자인 요소 분석. 석사학위논문, 서경대학교
강지현. (2009). 조선시대 기녀복색(妓女服色)의 특징: 16~18세기를 중심으로. 민속학연구, 25. 권혁산. (2012). 조선시대 무관초상화와 흉배에 관한 연구. 미술사연구, 26, 165–192.
김미경, 이은주. (2020). 조선시대 공신초상(功臣肖像)의 관복(제1기) 고찰. 문화재, 53(2), 180–203. 김선. (2016). 1960년대 이후 사극영화에 나타난 왕실 여성의상 연구. 석사학위논문, 성균관대학교
김월계. (2009). 사극 드라마에 나타난 조선시대 궁녀복식에 관한 연구. 한복문화, 12(1).
김유선. (2013). 영화의상. 커뮤니케이션북스.
김주영. (2003). 영상물 「장희빈」의 복식 분석. 석사학위논문, 이화여자대학교
남수경. (2004). 영화 〈스캔들〉 의상에 관한 연구. 석사학위논문, 충남대학교
박가영. (2010). 신윤복 풍속화와 영화 〈미인도〉 복식 비교 연구: 형태, 색채를 중심으로. 석사학위논문, 숙명여자대학교
신다혜, 나현신. (2016). TV 사극드라마 의상에 관한 기초연구. 한복문화, 19(2), 41–54. 심정택, 정재훈. (2026). 한국 전통문양의 생성형 AI 학습 데이터 구축과 활용 사례 및 전략 연구. 경영학연구
심현준. (2026). 광고사진 표현 형식 기반 생성형 AI 이미지 벤치마크 설계 및 상용 모델 성능 비교. 한국공간디자인학회논문집, 21(2), 83–96.
이성웅. (1987). 델파이 기술예측 방법론에 관한 연구. 박사학위논문, 전북대학교
임정미. (2017). 한국 사극 의상 디자인에서 보여지는 재매개 연구: 2000년대 이후 TV 사극을 중심으로 박사학위논문, 중앙대학교
장민정. (2008). 한국 사극영화 복식의 표현성에 관한 연구: 프로덕션 디자이너 정구호의 작품을 중심으로 박사학위논문, 성신여자대학교
전여선. (2007). 한국 영화와 TV드라마 전통복식에 나타난 색의 변화와 이미지의 상징성. 석사학위논문, 연세대학교
정두희. (2025). 회화 문화유산 복원을 위한 AI 활용 방안: 고지도 및 어진을 중심으로. 미술이론과 현장, 39, 63–75.
조미숙. (2005). 영상물 「장희빈」 복식의 고증실태와 분석. 석사학위논문, 전남대학교
한경희. (2004). 영화 〈창풍명월〉의 갑주(甲冑) 연구. 석사학위논문, 중앙대학교

양형모. (2025, 2월 13일). 생성형 AI의 한국 문화유산 왜곡 심각, 국가적 대응과 AI 대한민국 홍보대사 육성해야. 스포츠동아. https://sports.donga.com/world/article/all/20250213/131022581/1
이승미. (2026, 4월 28일). '100% AI 영화' 안방 넘어 극장으로: 예술의 확장인가 진정성의 훼손인가. 스포츠동아. https://sports.donga.com/ent/article/all/20260428/133827084/1
최민지. (2025, 12월 9일). 사극 속 주인공 한복 입히고, 모델 포즈 바꾸고...AI가 한다. 중앙일보. https://www.joongang.co.kr/article/25388703
최선. (2026, 5월 20일). [리뷰] 생성형 AI영화의 현재와 미래를 보여주는 가늠자, '한복 입은 남자'. 씨네21. https://www.cine21.com/news/view/?mag_id=109975
한아름. (2026, 3월 11일). 모픽, AI 사극 '문라이트 베일' 공개. 중앙일보. https://www.joongang.co.kr/article/25410978
Shin, J.-s. (2026, May 27). AI films face 15-second wall, require thousands of prompts. The Chosun Daily. https://www.chosun.com/english/travel-food-en/2026/05/27/YUUPRYQZ3VD7DHWTEKNKRSZNRM/
박수빈. (2025, 12월 24일). 100% 'AI 합성' 영화 나온다: 감독만 6명, '상업성' 시험대. 한국경제. https://www.hankyung.com/article/202512243089g

경청해 주셔서 감사합니다.

생성형 AI 사극 영화 장면의 인물 복식 고증을 위한 평가 요소 도출 및 벤치마크 데이터셋 개발 연구
- 조선시대 후기를 중심으로

연구자 : 박선주 (홍익대학교 AI·실감미디어콘텐츠학 박사과정)