



생성형 AI 기술을 적용한 특집 다큐멘터리 ‘영웅의 귀환 레클리스’ 제작기

글. KBS 제주방송총국 기술국



기획 의도와 초기 고민 - 생성형 AI로 재현한 역사

특집 다큐멘터리 <영웅의 귀환 레클리스>는 한국전쟁 당시 실제로 활약한 군마 ‘레클리스’를 조명한 프로그램입니다. 레클리스는 미 해병대 제1사단 소속으로 전장에서 탄약을 나르고 부상병을 이송하는 임무를 수행했으며, 하루 51차례 고지를 오르내리며 5톤에 달하는 탄약을 운반한 공로로 미국 군마 최초로 하사 계급을 부여받았습니다. 그 상징성과 역사적 의미는 충분히 조명할 가치가 있는 이야기였습니다.

그러나 기획 초기부터 명확한 제약이 있었습니다. 영상 자료가 거의 없었다는 점입니다. 일부 남아 있는 사진 역시 대부분 흑백의 정적인 이미지였고, 레클리스가 실제로 전장에서 임무를 수행하는 장면은, 영상은 물론 사진조차 남아 있지 않았습니다.

이런 경우, 통상적으로 재연 방식을 통해 영상화하는 것도 고려할 수 있습니다. 하지만 이번에는 말이 중심이 되는 전투 장면이 필요했고, 동물 안전 문제, 폭음·불꽃 연출에 대한 리스크, 연출 실패 시 재촬영의 어려움 등 여러 현실적 한계가 뒤따랐습니다.

이에 따라, 우리는 생성형 AI 기술을 활용한 영상 재현 방식을 실험적으로 도입하기로 결정했습니다. 당시로서는 방송 제작 현장에서 AI 이미지와 영상 생성 도구를 실제 프로그램에 적용해보는 첫 시도였으며, 기술적인 시행착오와 도전을 각오하고 프로젝트를 시작하게 되었습니다.

전체 제작 과정 - 컷 단위 이미지 기획에서 영상 제작까지

제작 초기, 생성형 AI 툴을 익히는 과정에서 텍스트만으로 영상을 생성하는 방식은 캐릭터나 장면의 일관성을 유지하기 어렵다는 한계를 확인하게 되었습니다. 동일한 프롬프트를 입력해도 결과물이 달라지는 AI의 특성 때문에, 스타일이나 구도를 통일하기 어려웠기 때문입니다.

이에 따라 전체 시나리오를 컷 단위로 나눈 뒤, 각 장면에 필요한 이미지를 먼저 제작하고 이를 기반으로 영상화하는 **Image to Video** 방식으로 접근하였습니다. 이미지를 기준으로 삼으면 장면의 구도, 캐릭터의 외형, 배경의 색감과 분위기 등을 일정하게 유지할 수 있어, 시각적 완성도를 높일 수 있었습니다. 이러한 방식은 영상의 흐름을 안정적으로 구성하는 데 도움이 되었을 뿐 아니라, 전체 제작 과정을 더 체계적이고 효율적으로 진행할 수 있도록 해주었습니다.



그림 1. 레클리스 과거 사진

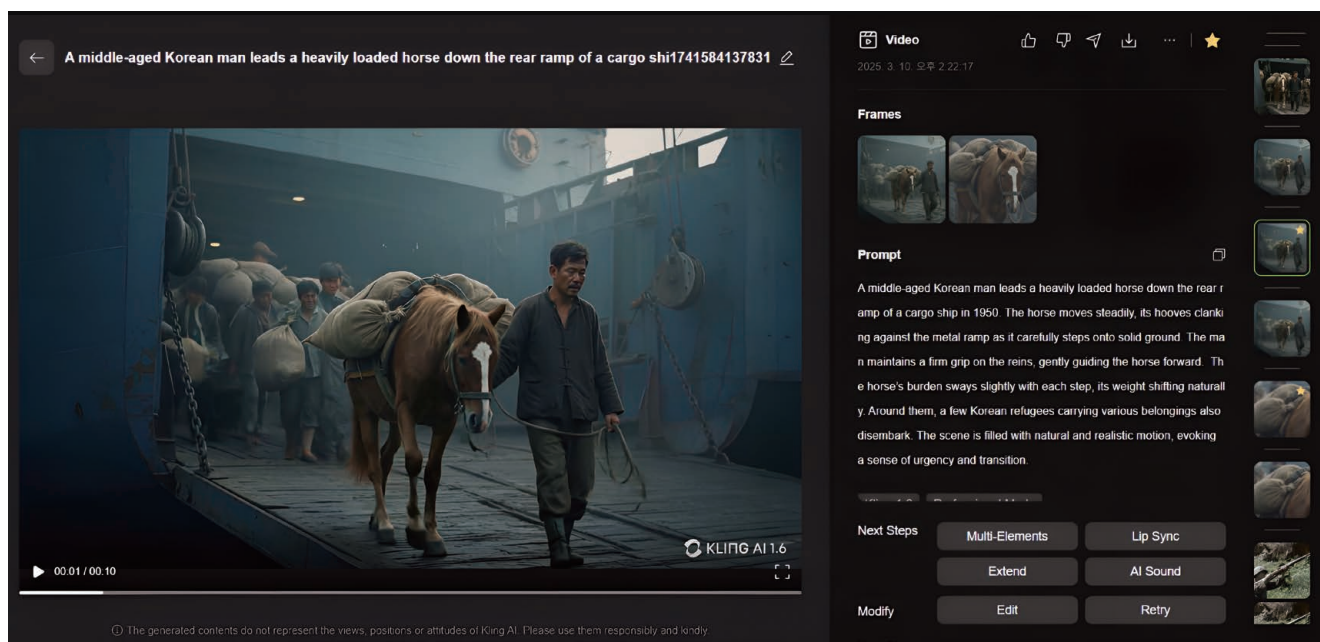


그림 2. Kling의 Image to Video 생성 방식

1단계 : 이미지 컷 기획 및 생성

컷 단위 이미지 제작은 전체 제작 흐름에서 가장 먼저 진행한 핵심 단계였습니다. 전체 시놉시스를 기반으로 필요한 장면들을 컷 단위로 정리하고, 각 장면에 필요한 구도와 상황을 문서로 작성한 후 이미지 생성 작업에 착수하였습니다.

번호	이미지	컷 설명	장면 설명	내용	ft	제작 이미지	피드백
s5c2		C#2 부산항 하역 작업 배 풀샷 설명: 부산항에 대령 화물선이 다가오는 모습 -아이레벨, 풀샷, 화면 아래 부두가의 일부 모습, 배의 이미지 참고		(나레이이션) 여명준 원래 피난민들과 함께 이동했고 부산항에서 짐을 나르는 데 사용되었죠.			
s5c3		C#3 부산항 하역 작업 풀샷 설명: 배에서 내리는 피난민들과 물자 -로우앵글 풀샷, 쏟아지는 사람들로 레클리스가 짐을 싣고 내려오는 모습 레클리스 등에는 가로 세로 1미터, 20명의 사람들이 머리에 붓짐을 지고 있거나, 통짐을 매고 있다.					
s5c4		C#4 부둣가 걷는 레클리스 설명: 피난민들 사이로 짐을 싣고 걸아가는 레클리스 -풀샷, 아이레벨, 화면 좌에서 우로 20명의 피난민들이 이동하는 모습, 한 가운데 레클리스가 가로 세로 1미터 두께 30센티 정도되는 큰 붓짐 두 개를 등 양옆에 매달고 사람들과 발 맞치 지나간다. (		(나레이이션) 시끄러운 소리와 군중들 속에서도 짐착한 모습이 인상적이었습니다.			

그림 3. 컷 단위 이미지 제작 예시

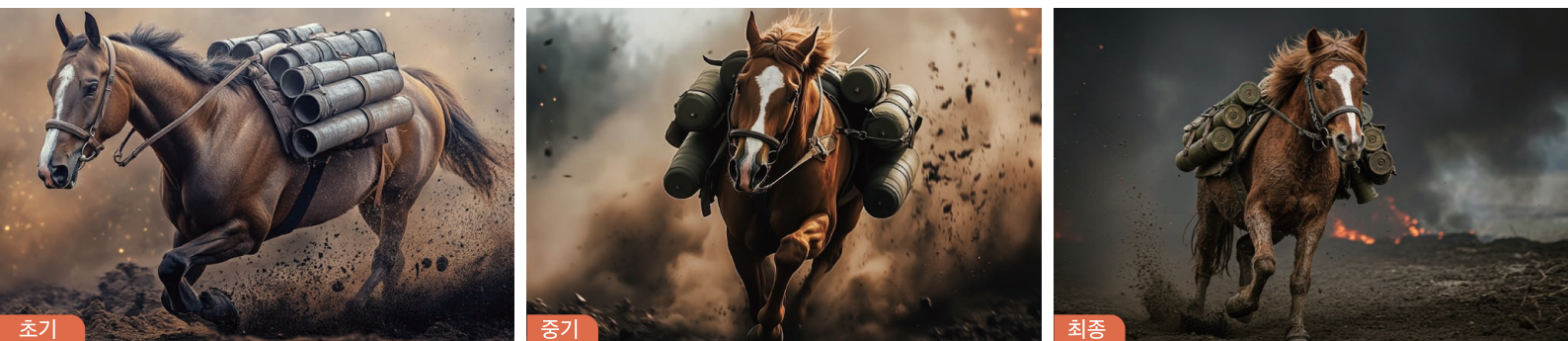


그림 4. 작은 체구의 제주마 특성을 반영한 레클리스 캐릭터 설계 과정(초기-중기-최종)

작업 초기에는 팀원별로 생성한 이미지의 스타일, 구도, 색감 등이 서로 다르게 표현되는 경우가 많았습니다. 동일한 전투 장면임에도 톤이 지나치게 밝거나, 말의 외형이 컷마다 일관되지 않게 표현되는 등의 문제가 발생했기 때문에, 팀원 간 작업 결과를 수시로 공유하고 피드백을 주고받으며 내부 이미지 제작 기준을 점차 정립해 나갔습니다.

특히 반복적으로 등장하는 주요 캐릭터 '레클리스'의 외형을 일관되게 유지하기 위해, 프롬프트에 고정된 키워드 조합을 활용하거나, 특정 이미지를 기반으로 파생 생성하는 방식을 적용하였습니다. 예를 들어 '작은 체구의 갈색 말', '얼굴에 긴 줄무늬'와 같은 키워드를 반복적으로 적용하여 캐릭터 외형의 일관성을 확보하였습니다.

이러한 기준 정립과 프롬프트 관리 과정을 통해 이미지의 시각적 완성도뿐 아니라, 장면 간 연결성과 몰입감까지 함께 높일 수 있었습니다.

2단계 : 이미지 보정 및 해상도 향상

AI를 활용해 원하는 장면을 빠르게 구현할 수는 있었지만, 세부 표현의 정확도나 구도의 완성도 면에서는 뚜렷한 한계가 있었습니다. 예를 들어, 말의 다리 색이나 갈기 색, 얼굴 중앙의 흰색 줄무늬처럼 고정되어야 할 특징들이 이미지마다 다르게 표현되거나 누락되는 경우가 많았습니다. 프롬프트에 해당 요소를 명시하더라도, AI가 이를 일관되게 반영하지 못해 세부 시각 요소에 대한 수동 보정이 필수적이었습니다.

이러한 문제를 해결하기 위해 Photoshop을 활용해 캐릭터 외형을 직접 수정하거나, Midjourney의 Edit 기능으로 이미지의 특정 부분을 지우고 재생성하는 방식으로 정확도를 보완하였습니다.



그림 5. Photoshop을 활용한 말의 다리색과 갈기색 보정

구도 측면에서도 어려움이 있었습니다. 캐릭터의 일관성을 유지하기 위해 프롬프트에 세부 묘사를 모두 포함하면, AI가 캐릭터 중심으로 이미지를 생성해 화면을 과도하게 채우거나, 원근감이 부족해 전체 이미지의 공간감이 제대로 표현되지 않는 경우가 많았습니다. ‘wide shot’, ‘full body’ 등 구체적인 지시어를 사용하더라도 원하는 구도를 얻는 데에는 한계가 있었습니다. 또한 실제 흑백 사진을 활용할 때도 문제가 있었습니다. 서로 다른 비율로 촬영된 탓에 16:9 화면에 맞추려면 여백을 확장하거나 배경을 추가 생성해야 했습니다.

이를 위해 Midjourney의 Edit 기능으로 배경을 확장해 시각적 균형을 맞추고, 장면 전환이 보다 자연스럽도록 조정하였습니다. 아울러 전체 영상의 품질을 높이기 위해 Midjourney와 Topaz의 업스케일링 기능을 활용해 고해상도 이미지를 확보하였고, 이는 이후 안정적인 영상 제작의 기반이 되었습니다.

그림 6-1. Midjourney Edit 기능을 활용한 배경 확장

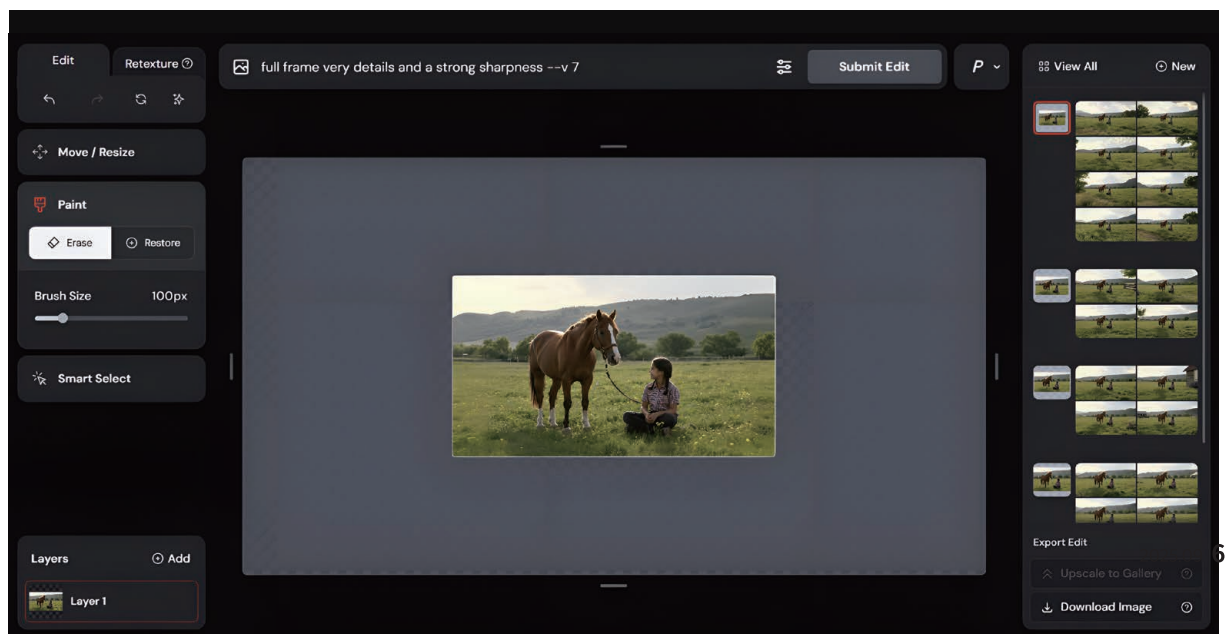




그림 6-2. Midjourney Edit 기능을 활용한 배경 확장

3단계 : 영상 제작 및 색보정 작업

보정이 완료된 이미지를 기반으로 Runway, Kling, Sora, Veo 등 다양한 AI 툴을 활용해 컷 단위 영상 생성을 진행하였습니다. 각 툴마다 결과물의 특성과 장단점이 뚜렷했기 때문에, 본격적인 영상 제작에 앞서 툴별 특성을 파악하고 장면의 성격에 따라 적절히 선택해 활용하였습니다.

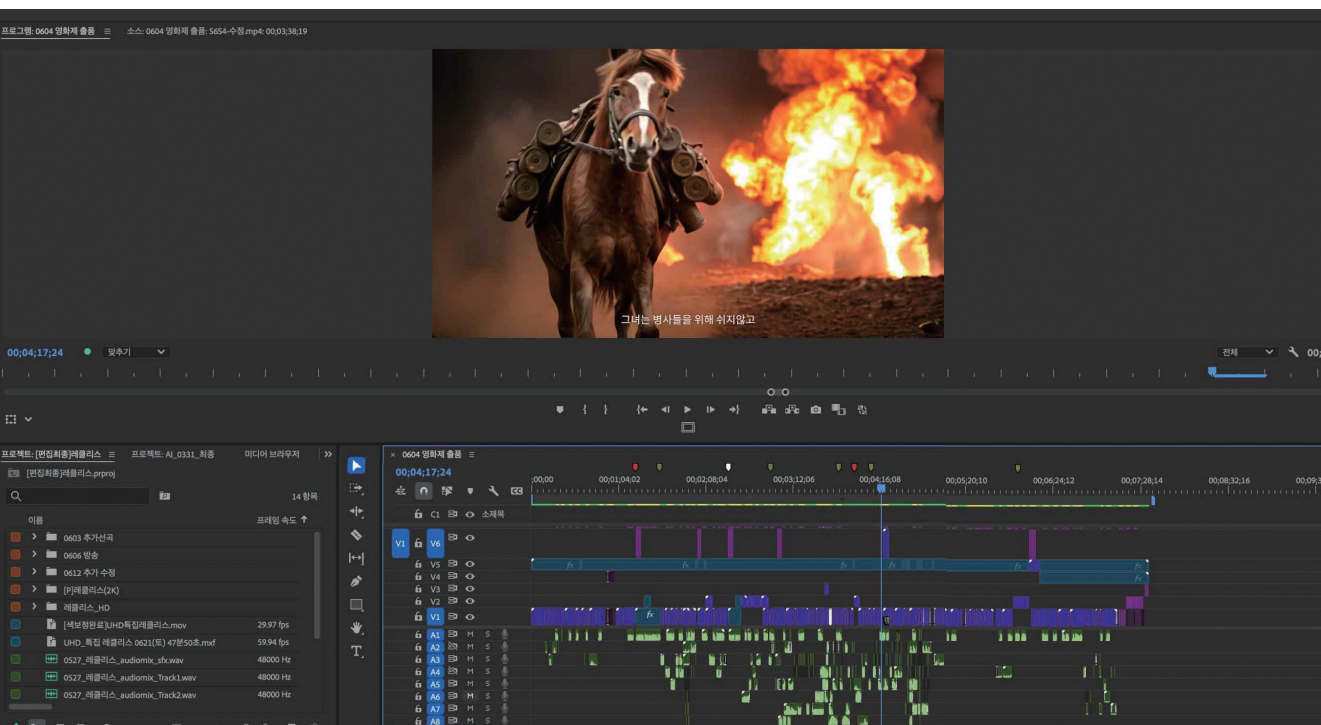


그림 7. Premiere Pro 타임라인

편집 과정에서는 이미지 간 시점, 구도, 캐릭터의 방향 등이 어색하지 않도록 연속성을 확보하는 데 중점을 두었고, 연결이 부자연스러운 장면은 중간 프레임을 추가하거나 배경을 보완해 자연스러운 전환이 이루어지도록 구성하였습니다. 생성형 AI 영상은 현실 영상과 달리, 인물이나 배경이 특정 시점에서 갑작스럽게 왜곡되거나 무너지는 등 프레임의 안정성이 흔들리는 문제가 자주 발생했습니다. 특히 말이 달리는 장면에서는 다리의 움직임이 실제 동작과 맞지 않거나, 관절이 비정상적으로 꺾이는 등 부자연스러운 표현이 나타났습니다. 이러한 문제를 최소화하기 위해 영상 길이와 전환 타이밍을 조정하고, 불안정한 구간은 컷 편집을 통해 제거하여 시청자의 몰입감을 유지하고자 하였습니다.



그림 8. 이미지 영상화 과정 중 발생한 시각적 오류 예시(병사의 다리 부분과 말의 눈 부분)

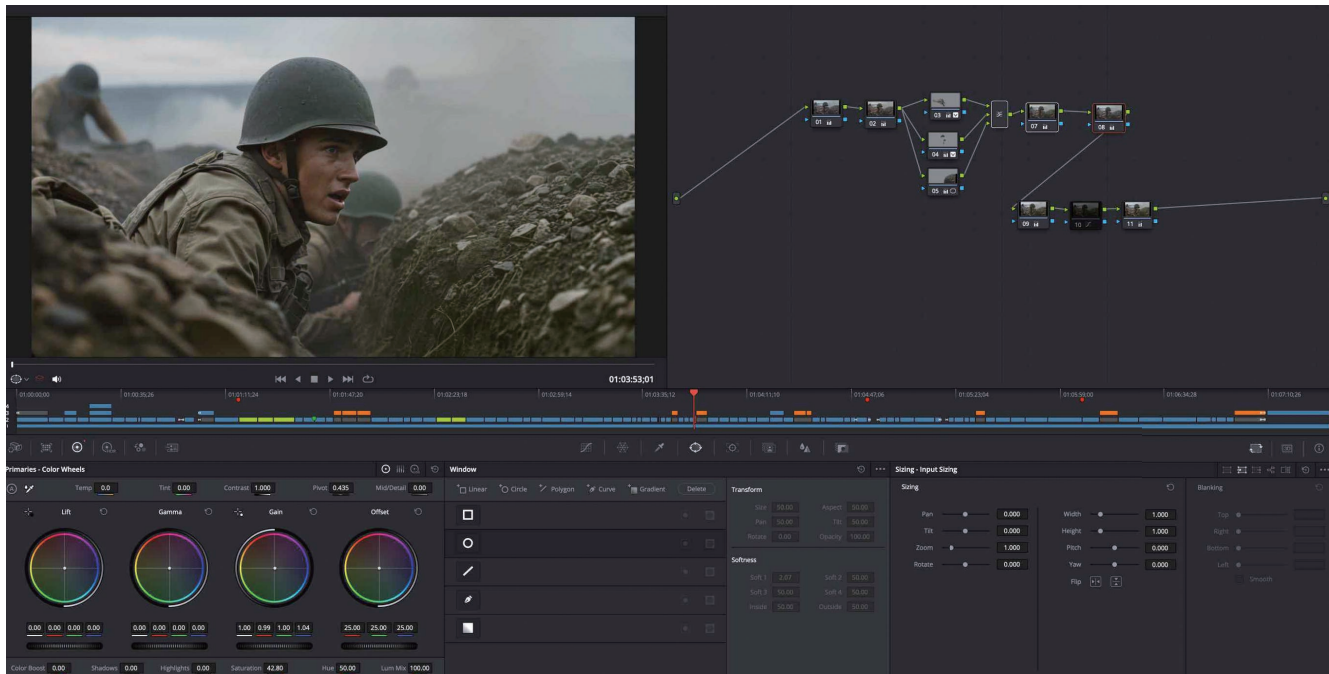


그림 9. DaVinci Resolve 작업 화면

또한 이미지에서 영상으로 전환하는 과정에서 색감이 과도하게 강해지거나 의도와 다른 색조로 표현되는 문제도 발생하였기 때문에, DaVinci Resolve를 활용해 전체 콘텐츠 톤과 각 장면의 분위기에 어울리는 컬러 그레이딩 작업을 병행하였습니다. 이는 단순한 색 보정을 넘어, 장면의 감정과 메시지를 더욱 선명하게 전달하는 연출적 요소로 작용하였습니다.

장면에 따라 달라지는 이미지 톨 선택 전략

이미지 생성 과정에서는 Midjourney와 ImageFX 두 가지 도구를 병행하여 사용하였으며, 각 톨의 특성을 고려해 장면의 목적에 따라 선택적으로 활용하였습니다.

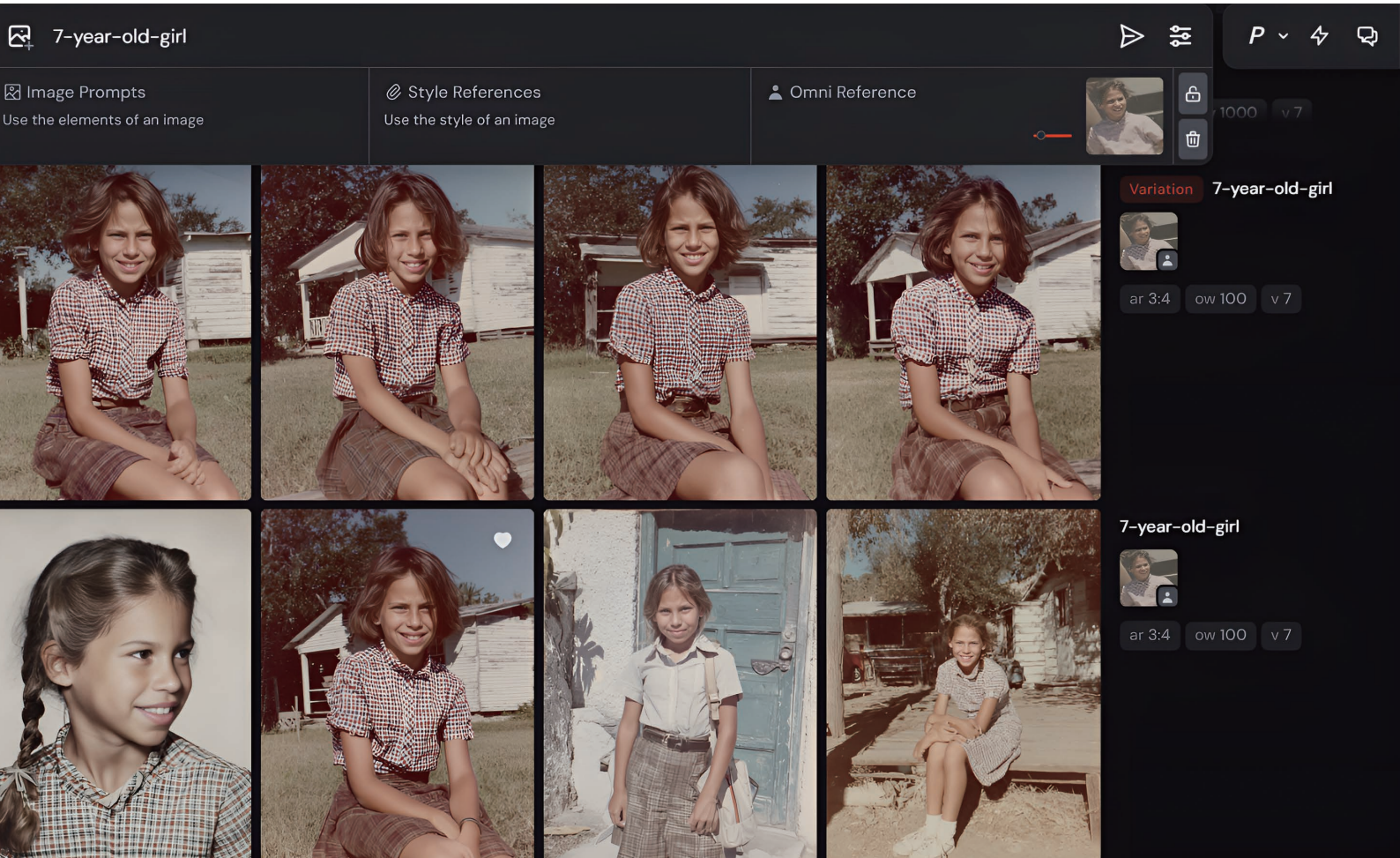


그림 10. 출연자의 과거 사진을 기반으로 한 캐릭터 생성 과정(Midjourney Omni Reference 활용)

Midjourney는 예술적이고 감성적인 이미지 생성에 강점이 있으며, 특히 색감과 분위기 표현에서 뛰어난 결과물을 보여주었습니다. 또한 스타일이나 캐릭터(Omni) 레퍼런스를 설정할 수 있는 기능이 있어, 이미지의 전체적인 톤과 인물의 일관성을 유지하는 데 효과적이었습니다.

ImageFX는 프롬프트 내 키워드를 구조적으로 해석하고 구현하는 데 강점을 보였습니다. 말의 얼굴 생김새, 다리 색, 등에 실린 탄약과 같은 주요 시각 요소들을 정확하게 재현하고, 장면 내 위치 또한 일관되게 유지할 수 있었습니다. 이미지 전체의 질감과 색감도 실제 사진처럼 사실적으로 표현되어, 다큐멘터리 형식의 영상에서 시청자의 몰입도를 높이는 데 효과적으로 작용했습니다.

이러한 특성을 기반으로, 감성적인 장면이나 분위기 중심의 컷에는 Midjourney를, 구조적 정확성과 재현성이 중요한 장면에는 ImageFX를 활용하였습니다. 각 톨의 강점을 적재적소에 적용함으로써 영상 전반의 표현력과 완성도를 효과적으로 끌어올릴 수 있었습니다.

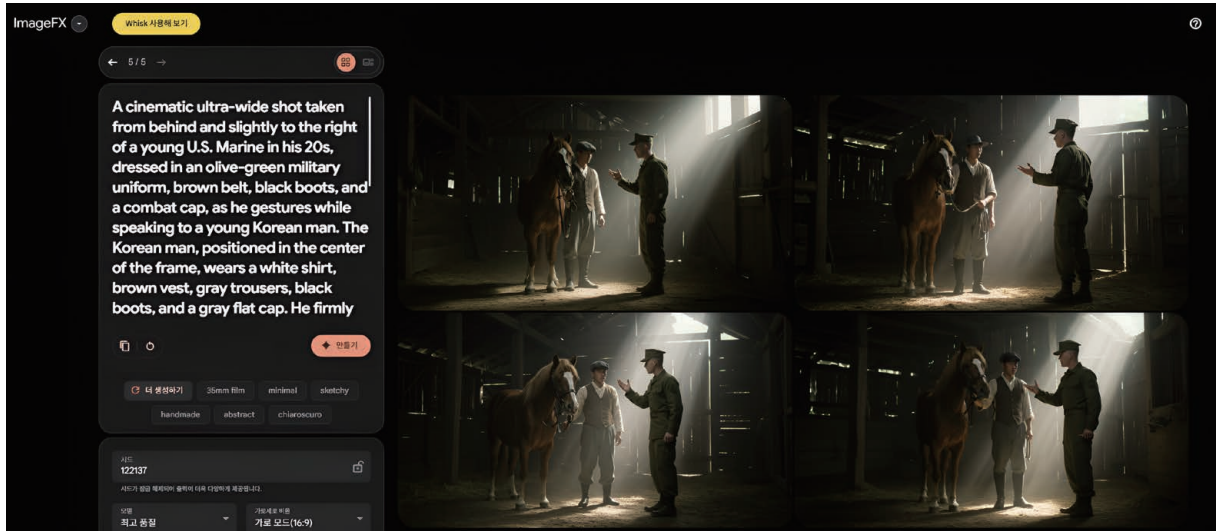


그림 11. 의상과 배경의 세부 요소까지 정밀하게 반영하는 ImageFX

AI 영상 생성 툴 활용 전략 : Kling, Runway, Sora

이미지 기반 컷을 영상으로 전환하는 과정에서는 다양한 AI 영상 생성 도구를 활용하였으며, 이 중에서도 Kling, Runway, Sora 세 가지 툴을 중심으로 특징과 활용 전략을 소개하고자 합니다.

Kling은 프롬프트에 입력한 지시를 정밀하게 수행하는 능력이 뛰어난 영상 생성 도구였습니다. 예를 들어 “말이 전속력으로 달린다”, “군인이 말 등에 짐을 싣는다”와 같은 동작 중심 문장을 입력하면, 해당 장면이 자연스럽게 정확하게 구현되었습니다.

또한 카메라 무빙이 포함된 장면에서도 시점이 이동함에 따라 기존 이미지에 없던 배경을 새롭게 생성해야 하는 경우가 많은데, Kling은 배경 확장 능력이 탁월해 화면 전환이 매끄럽고 왜곡 없이 이어졌습니다. 여러 인물이 동시에 등장하는 장면에서도 특정 인물만 움직이거나 배경과 어긋나는 현상 없이, 등장인물 모두의 움직임이 자연스럽게 표현되어 전체적으로 높은 안정성과 완성도를 보여주었습니다.

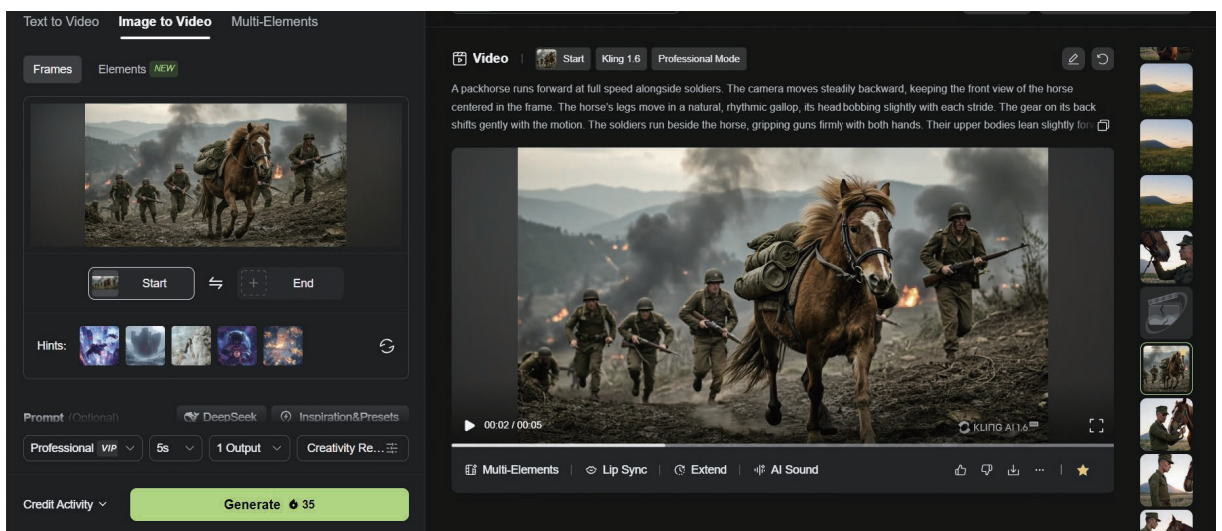


그림 12. Kling을 활용해 말과 군인들이 함께 달리는 장면을 자연스럽게 역동적으로 구현한 예시

Runway는 다른 툴에 비해 영상 생성 속도가 빠르고, 사용자 인터페이스가 직관적이어서 작업 효율이 매우 높은 도구였습니다. 특히 단일 인물을 중심으로 한 짧은 컷에서, 정적인 이미지에 부드러운 무빙을 더해 영상화하는 데 적합했으며, 실제로 남아 있는 흑백 사진을 바탕으로 한 영상 제작에도 효과적으로 활용할 수 있었습니다.

Gen-4로 버전이 업그레이드된 이후에는 Reference 이미지를 기반으로 캐릭터의 외형과 배경 스타일을 일관되게 유지한 채 새로운 이미지를 생성하고, 이를 즉시 영상화할 수 있는 기능이 추가되어 워크플로우의 효율성이 크게 향상되었습니다. 다만, 역동적인 장면을 구현하는 데에는 한계가 있었고, 예를 들어 말의 고삐나 장비 끈과 같은 세밀한 요소들이 영상화 과정에서 흐릿하거나 뭉개지는 경우가 종종 발생하였습니다.

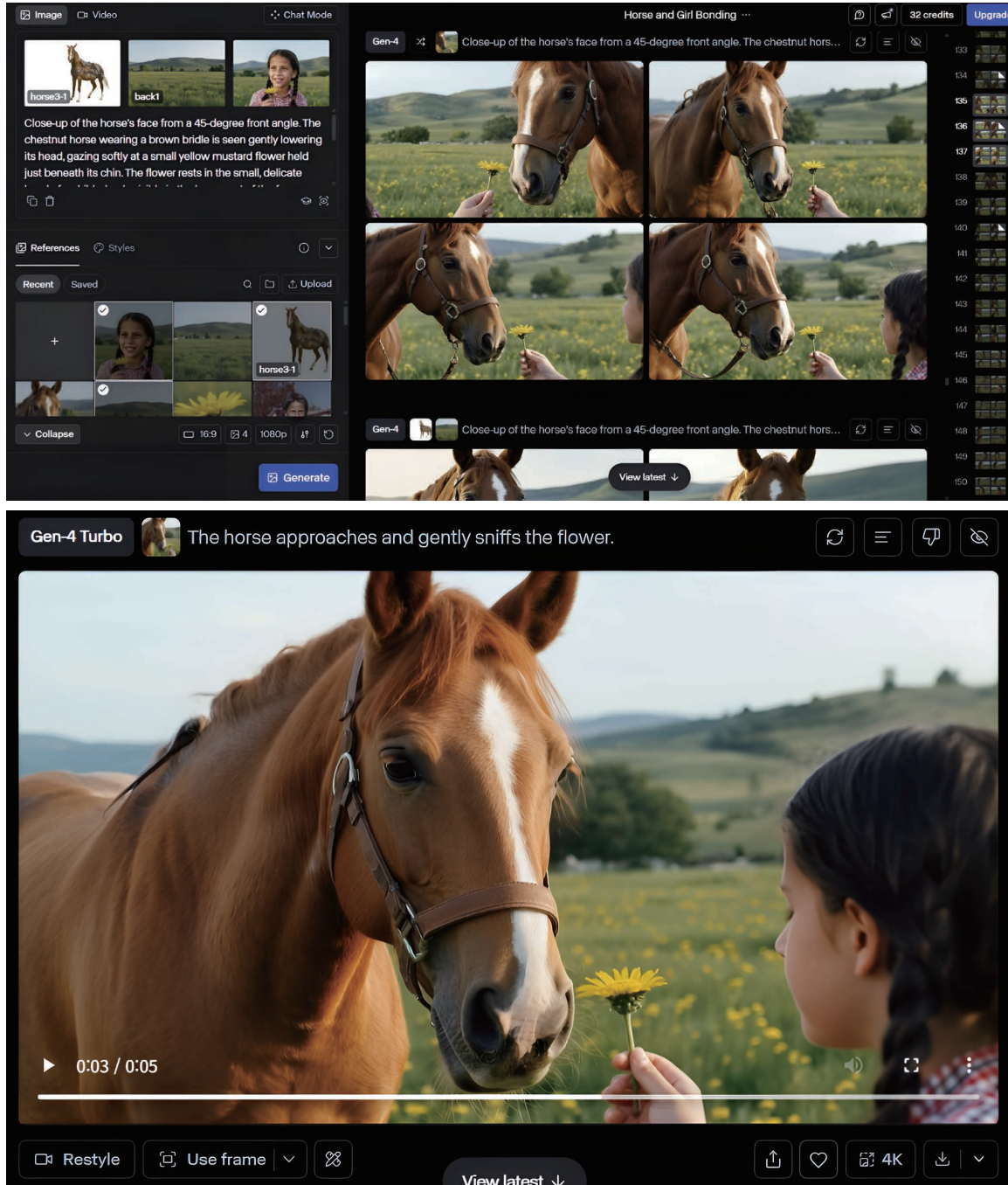


그림 13. Runway의 Reference 기능을 활용하여, 스타일이 반영된 이미지를 생성한 뒤 해당 이미지를 기반으로 즉시 영상화한 예시

Sora는 시네마틱한 영상 스타일을 구현하는 데 강점이 있었으나, Image to Video 기능의 실사용에는 분명한 한계가 있었습니다. 프롬프트를 정확히 반영하지 못하거나, 입력 이미지와 전혀 다른 장면이 생성되는 경우가 잦았고, 인물이나 배경이 어색하게 왜곡되는 현상도 빈번하게 발생했습니다. 이로 인해 정교한 연출이 필요한 작업에서는 안정적인 활용이 어려웠습니다.

반면, 기존에 생성한 짧은 영상을 자연스럽게 확장하는 Video to Video 기능에서는 뛰어난 성능을 보였습니다. 특히 Storyboard 기능을 활용하면 안정적으로 생성된 구간만을 선별해 이어 붙일 수 있었고, 장면의 속도나 움직임에 맞춰 다음 장면이 자연스럽게 이어져 컷 간 연결성과 몰입도를 높이는 데 효과적이었습니다.

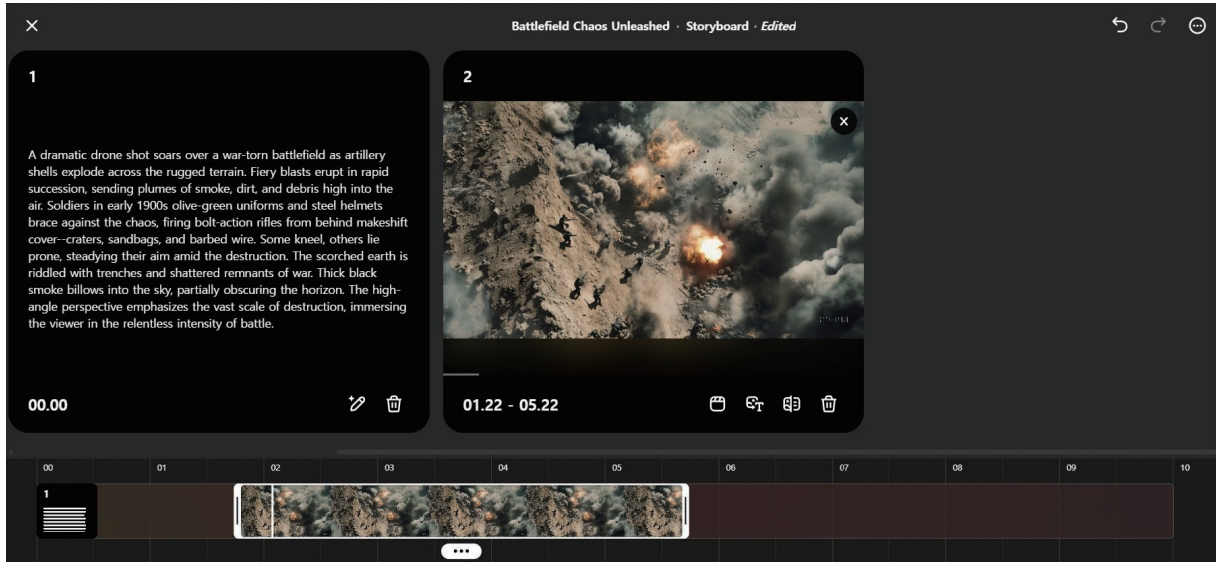


그림 14. Sora의 Storyboard 기능을 활용해 자연스러운 구간을 기반으로 5초 영상을 10초로 확장한 예시

이미지·음성·립싱크까지, 생성형 AI 인터뷰 제작 과정

생성형 AI를 활용해 가상의 퇴역군인 캐릭터를 제작하고, 인터뷰 장면을 연출하는 시도도 함께 진행하였습니다. 인물의 외형부터 음성, 연출까지 전 과정을 AI로 구현해, 실제 인물이 등장하는 듯한 다큐멘터리 인터뷰를 재현하고자 했습니다.

캐릭터 생성에는 ImageFX를 활용했습니다. “고된 전장을 경험한 80대 퇴역군인, 햇볕에 그을린 반점, 진지하고 회상에 잠긴 표정” 등 성격적·시각적 디테일을 반영한 프롬프트를 바탕으로 설계했으며, 정면·측면·45도 각도 등 다양한 시점에서 동일 인물을 생성해 인터뷰 구도에 적합한 컷을 확보하였습니다.

음성 구현에는 ElevenLabs의 AI 음성 생성 기술을 활용하였습니다. 인물의 연령, 성격, 말투, 감정 표현 등을 반영한 텍스트를 기반으로 자연스러운 톤을 구현했고, 발음과 억양을 반복적으로 조정해 실제 인터뷰처럼 들리는 고품질의 음성을 완성할 수 있었습니다.



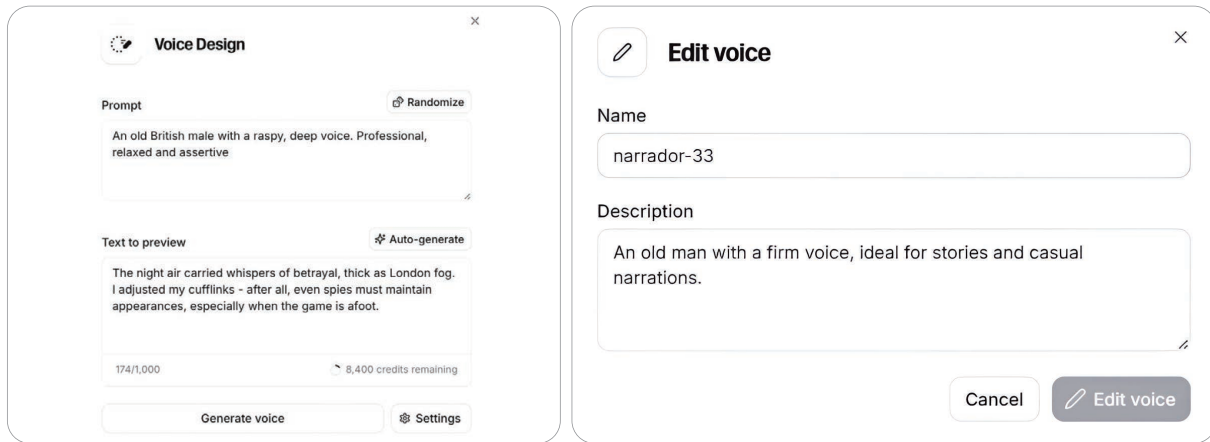


그림 15. ElevenLabs 작업 화면

립싱크 구현은 Hedra를 통해 진행하였습니다. 이미지와 음성을 결합하여 입 모양과 제스처가 일치하는 장면을 연출하였고, 인터뷰 중 손짓이나 표정 변화도 비교적 자연스럽게 표현되어 실제 인물이 인터뷰하는 듯한 몰입감을 전달할 수 있었습니다. 다만 일부 구간에서는 얼굴이나 손동작이 어색하게 무너지는 현상이 발생하여, 편집 과정에서 자연스럽게 표현된 부분만을 선별하여 사용하는 작업이 필요했습니다.

최종 방송에는 포함되지 않았지만, AI가 만들어낸 인물의 표현력과 완성도는 인상적이었으며, 향후 실제 콘텐츠 제작에 응용할 수 있는 충분한 가능성을 보여주었습니다.

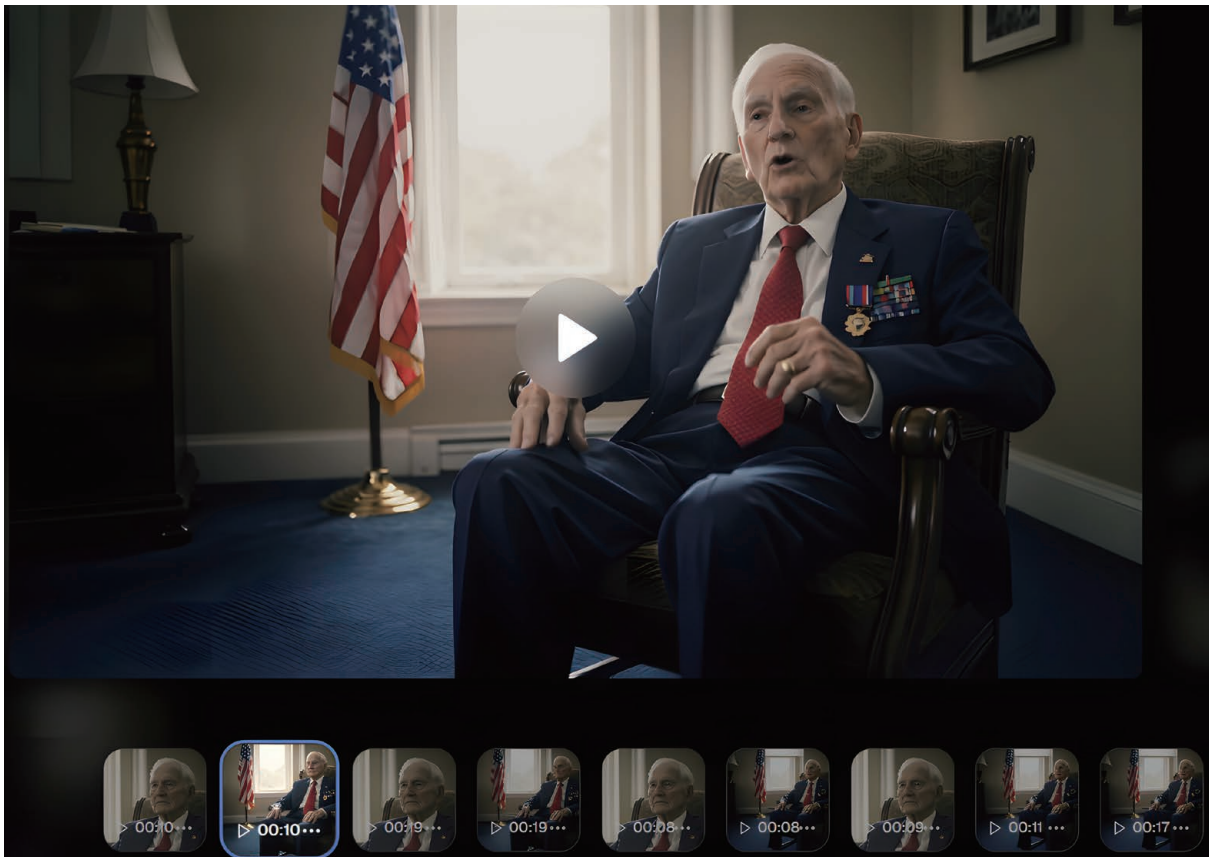


그림 16. Hedra로 생성해 낸 립싱크 장면

좌충우돌, 그럼에도 끝까지

2025년 2월, 생성형 AI를 활용한 다큐멘터리 제작이라는 새로운 도전에 나섰습니다. 처음에는 “이런 것도 가능하네?”라는 놀라움과 기대 속에 출발했지만, 본격적인 제작에 들어서자 쉽지 않은 현실과 마주하게 되었습니다.

긴 호흡의 영상에서는 캐릭터의 얼굴이 중간에 무너지거나 말의 다리가 여섯 개로 늘어나는 등 오브젝트의 일관성을 유지하는 데 어려움이 컸습니다. 프레임 왜곡이나 붕괴 현상도 빈번하게 발생해, ‘과연 이 작업을 끝까지 완성할 수 있을까’하는 회의감이 들기도 했습니다.

하지만 다양한 이미지 생성 및 영상화 툴을 접하면서 점차 실마리를 찾아갈 수 있었습니다. 장면의 특성과 목적에 맞는 툴을 선택하면서 영상의 퀄리티도 조금씩 향상되었고, 반복적인 실험을 통해 “이건 가능하겠다”, “이건 아직 어렵다”처럼 판단 기준이 생기면서 작업의 방향성도 분명해졌습니다. 툴의 한계를 타하기보다는 직접 부딪히며 배우는 과정 자체가 작업 속도를 높이고, 끝까지 프로젝트를 완수할 수 있었던 원동력이 되었습니다.

생성형 AI는 지금, 이 순간에도 빠르게 진화하고 있습니다. Veo 3와 Kling은 효과음까지 포함한 영상 생성을 지원하고, Midjourney 역시 영상 기능을 새롭게 탑재했습니다. 기술이 빠르게 바뀌는 만큼, 생성형 AI를 잘 활용하기 위해서는 직접 실험하고 판단하는 태도, 그리고 새로운 툴을 유연하게 받아들이는 자세가 제작자에게 더욱 필요해질 것 같습니다. 🎬

