

소라[Sora] 이후 동영상 생성형 AI 서비스 출시 동향

[목차]

1. 들어가며
2. 소라 터보[Sora Turbo] 상용화, 그리고 베오2[Veo2] 공개
3. 동영상 생성형 AI의 성능 비교
4. 동영상 생성형 AI의 적용 사례
5. 마치며

요약문

“

2024년 2월 챗GPT를 만든 오픈AI가 동영상 생성형 AI ‘소라(Sora)’를 공개한 이후 영상 제작 업계뿐만 아니라 일반인들에게도 텍스트 투 비디오(Text to Video, T2V)에 대한 관심이 커졌다. 메타, 오픈AI, 구글 등 빅테크 기업뿐만 아니라 루마 AI (Luma AI, Dream Machine), 런웨이(Runway, Gen-3 Alpha), 피카랩스(Pika Labs) 등 중소기업체와 콰이쇼우(Kling AI), 바이트댄스, 알리바바 등 중국기업까지 다양한 기업들이 연이어 새로운 동영상 생성형 AI 서비스를 발표하며 경쟁하고 있다. 본 글에서는 작년 2월 ‘소라(Sora)’가 공개된 이후, 다양한 사업자들이 출시 또는 공개한 동영상 생성형 AI 서비스 동향에 대해 알아보고자 한다.

”



들어가며

2024년 2월 오픈AI가 소라(Sora)를 발표하고 난 후 동영상 생성형 AI에 대한 관심이 크게 늘었다. 기존 서비스들로부터 제작된 어설피고 오류가 많은 영상이 아니라 실제 인간이 촬영한 듯한 영상이 공개되면서 영상 제작 업계뿐만 아니라 일반인들의 관심도 커졌다. 이전에는 주로 스타트업이나 중소기업체 위주로 서비스 개발과 발표가 이루어졌다면 이후부터는 메타, 구글 등 글로벌 IT기업들도 본격적으로 연이어 서비스를 발표하면서 이 분야에서 활발한 경쟁을 벌이기 시작했다.

2024년 5월에는 구글이 베오(Veo)를 발표했고, 6월에는 런웨이가 기존 서비스(Gen-2)를 개선한 젠-3

알파(Gen-3 Alpha)를 발표했으며 루마 AI는 드림 머신(Dream Machine)을 상용화했다. 7월에는 중국의 콰이쇼우¹⁾가 클링 AI(Kling AI)를 글로벌로 출시했고, 10월에는 메타가 무비젠(Movie Gen)을 공개했다. 소셜미디어 페이스북 운영사인 메타가 10월에 공개한 무비젠은 텍스트를 입력하면 최대 16초 길이의 동영상을 생성할 수 있다. 이 서비스는 인물 사진을 넣어 그 사람이 등장하는 영상을 생성할 수도 있고 영상에 배경음악이나 효과음을 넣는 것도 가능하다. 그러나 메타는 무비젠을 일반인들이 이용할 수 있도록 상용화한 것이 아니라 영화 제작자를 포함한 소수의 외부 파트너에게만 우선 제공했다. 그리고 2025년에는 인스타그램과 왓츠앱 등 자사 소셜미디어 앱에 탑재할 계획이라고 발표했다. 동영상 생성형 AI 서비스의 미흡한 완성도를 고려할 때 단기적으로 일반인들에게 확산시키기 위해서는 소셜미디어, 즉 온라인·모바일 위주 동영상 제작에 적용하는 것이 합리적이다. 구글도 5월에 베오를 공개하면서 유사한 계획을 발표했다. 비디오FX(VideoFX)를 통해 소수의 영상 제작자가 참여하는 실험을 진행했으며 향후 서비스의 오류가 수정되고 생성 영상의 완성도가 개선될 경우 유튜브 쇼츠에 적용하겠다고 발표했다.

〈표 1〉 2024년도 주요 동영상 생성형 AI

| 서비스 | Veo | Gen-3 Alpha | Dream Machine | Kling AI | Movie Gen |
|-------|--|---|--|--|--|
| 개발사 | Google | Runway | Luma AI | Kuaishou | Meta |
| 공개시기 | 2024. 5 | 2024. 6. | 2024. 6. | 2024. 7. | 2024. 10. |
| 주요특성 | <ul style="list-style-type: none"> 영상, 이미지, 텍스트 입력으로 영상 생성 및 편집 마스크 기능으로 특정 영역 변경 가능 참조 이미지 스타일 활용 영상 제작 영상 클립 연속 확장 | <ul style="list-style-type: none"> Gen-2 대비 높은 실재감, 시간적 일관성 영상 생성 속도 빠름 다양한 스타일 적용 영상 생성 | <ul style="list-style-type: none"> 텍스트, 이미지 입력으로 고품질 영상 생성 영상 생성 속도 매우 빠름 사용하기 쉬운 UI 이미지 2개로 처음-끝 영상 생성 가능 다언어 지원 (한국어) | <ul style="list-style-type: none"> 텍스트, 이미지 입력으로 고품질 영상 생성 실재감 높고 비교적 긴 영상 생성 가능 다언어 지원 | <ul style="list-style-type: none"> 텍스트 입력으로 영상 생성 및 편집 사진 활용 맞춤형 영상 제작 음향효과, 사운드트랙 제작 |
| 상용화 | VideoFX 실험 중, Youtube Shorts 적용 계획 | 무료/표준(월\$12)/프로(월\$28)/무제한(월\$76)/기업용(별도) | 무료/표준(월\$23.99)/프로(월\$79.99)/프리미어(월\$399.99) 외 다양 | 무료/표준(월\$5)/프로(월\$18.5)/프리미어(월\$46) | 2025년 인스타그램 탑재 예정 |
| 영상 길이 | 60초 이상 | 10초 | 5초 | 최대 2분(HD) | 16초 |

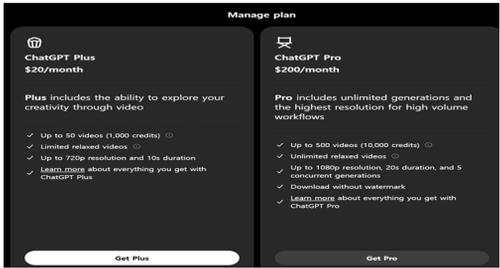
자료: 필자 작성

part
2

소라 터보(Sora Turbo) 상용화, 그리고 베오2(Veo2) 공개

오픈AI는 2024년 2월에 공개했던 소라를 개선한 '소라 터보(Sora Turbo)'를 12월 9일에 일반인도 사용할 수 있는 서비스로 출시했다.²⁾ 이 서비스는 챗GPT 플러스 또는 프로 이용자면 추가 요금 없이 이용할 수 있다. 챗GPT 플러스 이용자는 매월 480p 해상도 기준 최대 50개의 영상을 생성하거나 더 적은 수의 720p 해상도의 영상을 생성할 수 있다. 챗GPT 프로 이용자는 플러스보다 10배 많은 영상을 생성할 수 있으며 최대 1080p 해상도, 최대 20초 길이의 영상을 생성할 수 있다. 이용자는 텍스트(프롬프트) 입력으로 새로운 영상을 생성할 수도 있고 텍스트, 이미지, 영상을 입력해 새로운 영상을 생성할 수 있다. 그리고 소라 터보에서는 스토리보드 도구를 제공해 이용자가 각 프레임에 대한 입력(화면 비율, 해상도, 영상 길이, 변형 등)을 별도로 지정할 수 있으며 자신이 가진 영상을 확장, 편집할 수도 있다. 12월 17일에는 5월에 베오를 공개했던 구글 딥마인드가 이를 개선한 베오2(Veo2)를 발표했다.³⁾ 최대 2분 길이의 4K 해상도 영상을 생성할 수 있어 소라 터보보다도 더 길고 고화질의 영상을 만들 수 있으며, 이용자가 렌즈 유형, 카메라 효과 등을 지정하는 등 개선된 카메라 제어 기능을 제공했다.

〈그림 1〉 소라 터보 이용조건

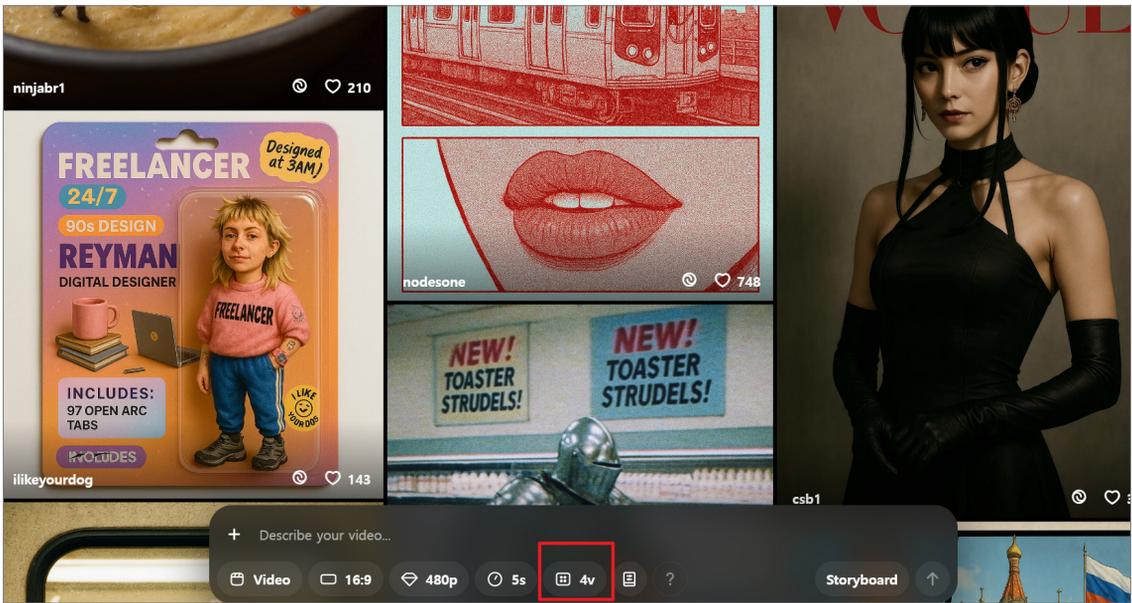
| | | | | | |
|---|--------------|-------------|-------------|--------------|--------------|
|  <p>(a) 챗GPT 요금제별 이용조건</p> | | 5s | 10s | 15s | 20s |
| | 480p square | 20 credits | 40 credits | 60 credits | 80 credits |
| | 480p | 25 credits | 50 credits | 100 credits | 150 credits |
| | 720p square | 30 credits | 75 credits | 150 credits | 225 credits |
| | 720p | 60 credits | 180 credits | 360 credits | 540 credits |
| | 1080p square | 100 credits | 300 credits | 650 credits | 1000 credits |
| | 1080p | 200 credits | 600 credits | 1300 credits | 2000 credits |

(b) 영상 해상도, 길이별 소요 크레딧

출처 : 오픈AI 홈페이지(2024)

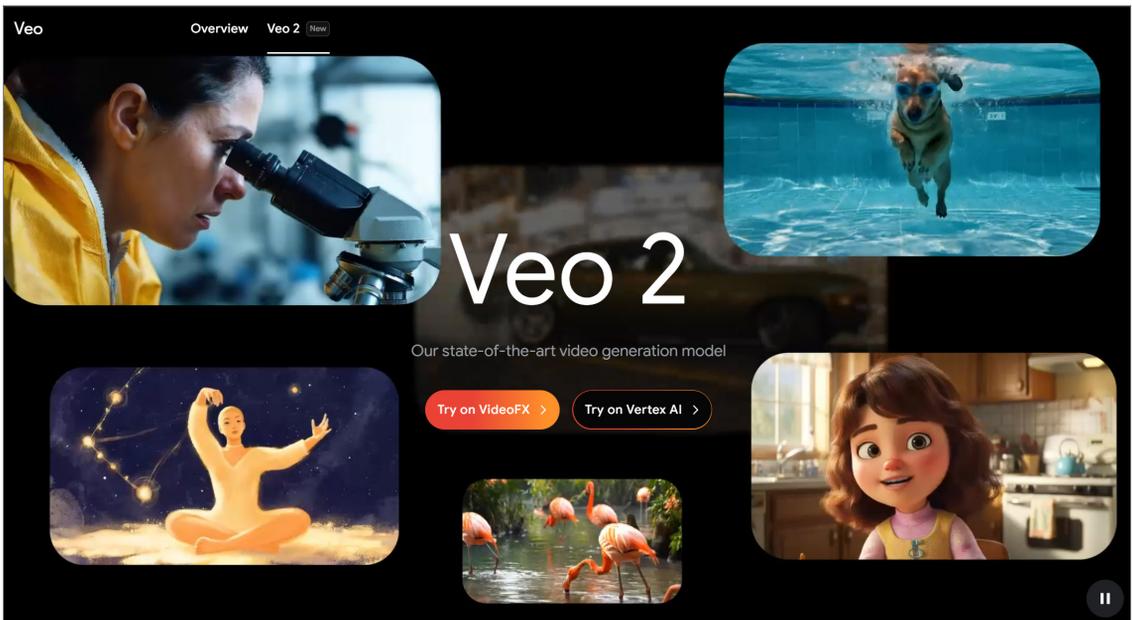
1. 2011년에 설립된 중국의 숏폼 및 라이브커머스 플랫폼 기업.
 2. 오픈AI 홈페이지 (2024)
 3. <https://deepmind.google/technologies/veo/veo-2/>

<그림 2> 소라 터보



출처 : 오픈시 홈페이지

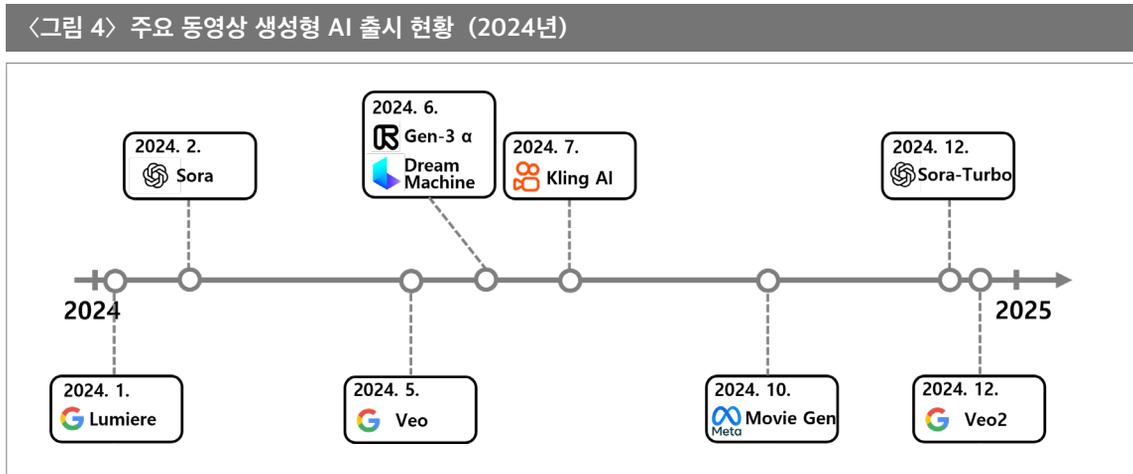
<그림 3> 베오2



출처 : 구글 딥마인드 홈페이지 (2024)

그러나 베오2는 아직 일반인들은 이용하지 못하며 비디오FX, 유튜브, 버텍스 AI(Vertex AI)를 통해 소수 영상 제작자에게만 이용이 허용되고 있다.

소라 터보와 베오2는 아직 이를 활용해 고품질의 영상을 제작하기에는 여러 가지 오류와 한계를 가지고 있지만 기존의 동영상 생성형 AI보다 물리현상을 더 정확히 시뮬레이션하고 영상의 일관성을 잘 유지하는 것으로 평가받고 있다.⁴⁾ 소라 터보는 타임라인, 키 프레임, 편집 기능 등 사용자 인터페이스와 영상 제어에 강점이 있는 반면 베오2는 생성 영상의 화질과 물리 시뮬레이션에 상대적으로 더 장점이 있다.⁵⁾ 그래서 소라 터보가 스포츠 콘텐츠 제작자의 수요에 적합하다면 베오2는 향후 보다 전문적인 영상을 제작하려는 수요에 적합할 것이라는 의견도 있다.⁶⁾



출처 : 필자 제공

4. Robert-Islam, 2025's AI Video Showdown: Comparing Google Veo 2 And OpenAI Sora. Forbes. 2025. 3. 6.

5. Sharma, Sora vs Veo 2: Which One Creates More Realistic Videos? Analytics Vidhya. 2025. 1. 15.

Robert-Islam, 2025's AI Video Showdown: Comparing Google Veo 2 And OpenAI Sora. Forbes. 2025. 3. 6.

6.Hosalikar, Veo2 vs. Sora AI: Which Is a Better Video Generation Model? Gumlet. 2025. 1. 5.

계속해서 다양한 기업들에 의해 동영상 생성형 AI 서비스가 개발되고 공개되면서 이들의 성능을 객관적으로 평가하고 서로 비교하는 것이 중요해졌다. 소프트웨어 및 서비스 분야에서 이와 같은 비교 및 평가를 벤치마크(또는 벤치마크 테스트)라고 한다. 벤치마크(benchmark)란 두 개 이상의 소프트웨어의 성능을 평가하고 비교하기 위해 동일한 조건과 환경에서 실행하는 테스트 또는 그러한 테스트를 하기 위해 만든 환경(조건, 평가 기준)을 의미한다.⁷⁾

메타는 무비젠을 공개하면서 새로운 벤치마크로 무비 젠 벤치(Movie Gen Bench)를 소개했다.⁸⁾ 이 벤치마크는 동영상 생성형 AI에 대한 평가의 기준으로 텍스트 정합성(text-alignment), 시각적 품질(visual quality), 사실감과 미적 요소(realness & aesthetics)라는 3가지 평가 기준(evaluation axes)을 설정했으며 각 평가 기준별로 세부 평가항목(sub-axes)을 정의했다. 생성된 영상의 성능은 위 평가 기준과 세부 평가항목에 대해 인간 평가(human evaluation)로 판정했다.⁹⁾ 전문가 평가자들이 두 서비스에서 생성된 영상을 비교하게 해 더 우수한 영상을 선택하거나 명확히 우수한 영상이 없을 경우 동률로 판정하도록 한다. 이 벤치마크는 복수의 동영상 생성형 AI 서비스 비교를 위해 1,003개의 프롬프트를 포함하고 있다. 이 프롬프트는 인간 활동, 동물, 자연 및 풍경, 물리법칙, 비정형적인 주제 및 특이한 행동 등 다양한 영역의 내용으로 제작되었다.

메타는 이 벤치마크를 소개하면서 무비젠과 당시 공개된 다른 서비스들(런웨이 젠-3, 루마 AI 드림머신, 오픈AI 소라, 콰이쇼우 클링 1.5)¹⁰⁾과 성능을 비교했다. 성능 비교를 위해 각 서비스에서 생성된 영상의 해상도와 영상 길이는 유사한 수준이 되도록 조정했다. 평가 결과, 전체 품질을 기준으로 무비젠은 런웨이

7. 우리말샘 홈페이지 (2025).

8. 메타는 영상을 위한 무비젠 비디오 벤치(Movie Gen Video Bench)와 음향을 위한 무비젠 오디오 벤치(Movie Gen Audio Bench) 두 개의 벤치마크를 공개했다. 본 글에서는 이 중 무비젠 비디오 벤치를 다룬다.

9. 기존에 개발된 많은 벤치마크에서는 인간 평가가 아니라 공학적 계산을 통한 자동화된 평가 지표를 적용했다(Singer et al., 2023; Girdhar et al., 2024). 그러나 메타 기술자들은 텍스트 정합성, 시각적 품질, 사실감 및 미적 요소 등 평가 기준의 복잡성으로 인해 기존의 자동화된 평가 지표가 신뢰할 만한 결과를 제공하지 못한다고 했평가했다(메타, 2024. 10. 16.).

10. 2024년 9월 콰이쇼우에서 기존 클링 AI를 개선해 내놓은 후속 버전 서비스.

〈표 2〉 무비젠 벤치 평가 기준 및 항목

| 평가 기준 (Evaluation Axes) | 세부 평가항목 (Sub-axes) | 내용 |
|----------------------------|--------------------------------------|---|
| 텍스트 정합성 | 주제 일치 모션 일치 | <ul style="list-style-type: none"> • 생성된 영상이 입력된 프롬프트와 얼마나 잘 일치하는지 정도 • 주제의 외형, 배경, 조명, 스타일 등의 정합성 • 동작 관련 설명과 생성된 영상의 움직임이 일치하는지 정도 |
| 시각적 품질 | 프레임 일관성 모션 완전성 모션 자연성 전체 품질 | <ul style="list-style-type: none"> • 생성된 영상이 일관되고 자연스럽게 충분한 양의 움직임을 묘사하는 정도 • 생성된 영상의 시간적 일관성 • 영상에 충분한 움직임이 포함되어 있는지 정도 • 생성된 영상의 움직임이 현실적이며 물리법칙을 준수하는지 정도 • 전체적인 시각적 품질을 기준으로 더 나은 영상을 선택 |
| 사실감과 미적 요소 | 사실감 미적 요소 | <ul style="list-style-type: none"> • 생성된 영상이 사실적인지, 그리고 미적으로 우수하고 매력적인 정도 • 생성된 영상이 실제 영상과 유사한 정도 • 생성된 영상의 구성, 조명, 색상, 카메라 효과 등이 매력적인 정도 |

출처 : 메타 (2024. 10. 16.)

〈표 3〉 무비젠 비디오와 다른 서비스 간 벤치마크 결과 (무비젠 비디오의 순승률)

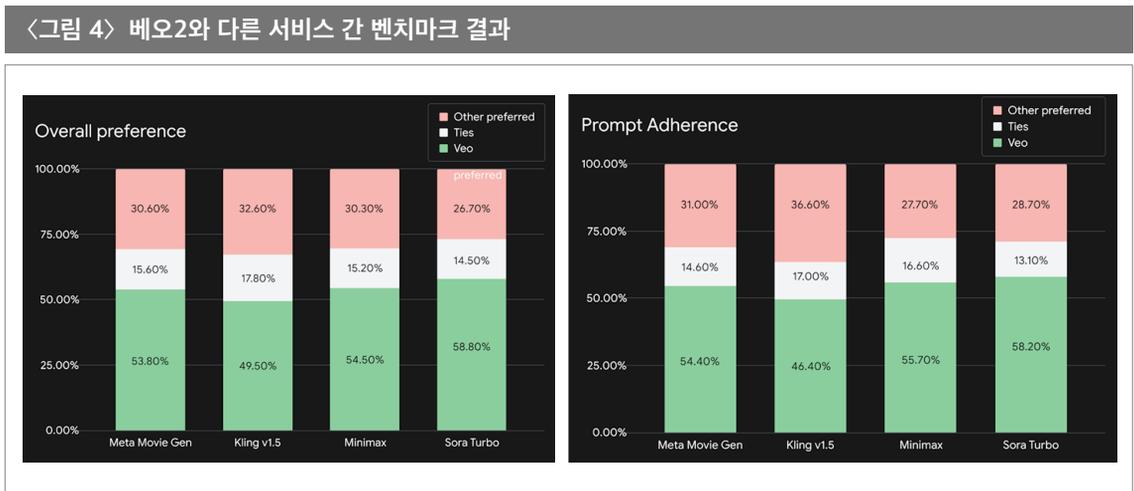
| 평가항목 | Runway Gen-3 | Dream Machine | Sora | Kling 1.5 | 표준편차 |
|---------|--------------|---------------|-------|-----------|-------|
| 텍스트 정합성 | 10.45 | 12.23 | 17.72 | -1.99 | ±3.74 |
| 전체 품질 | 35.02 | 60.58 | 8.23 | 3.87 | ±5.07 |
| 프레임 일관성 | 33.10 | 42.14 | 8.22 | 13.50 | ±4.08 |
| 모션 완전성 | 19.27 | 29.33 | 4.43 | 0.52 | ±3.98 |
| 모션 자연성 | -1.72 | 23.59 | 8.86 | -10.04 | ±1.68 |
| 사실감 | 48.49 | 61.83 | 11.62 | 37.09 | ±2.52 |
| 미적요소 | 38.55 | 48.19 | 6.45 | 26.88 | ±4.84 |

출처 : 메타 (2024. 10. 16.)

이 젠-3(순승률 35.02%), 루마 AI 드림머신(60.58%)보다 성능이 우수했으며, 소라에게는 약간의 우위를 (8.23%), 클링 1.5와는 비슷한 성능(3.87%)을 보였다(표준편차 ±5.07%). 모션 자연성에 있어서는 무비젠 이 플링 1.5보다 더 열위에 있는 것으로 평가되었다(순승률 -10.04%).

구글 딥마인드도 작년 12월 베오2를 공개하면서 메타의 무비젠 벤치를 활용해 자사의 서비스와 다른 주요 서비스들을 비교했다. 1,003개의 프롬프트 입력에 대해 베오2와 메타의 무비젠, 콰이쇼우의 클링 1.5, 미니맥스의 하이루오 AI(Hailuo AI)¹¹⁾, 오픈AI의 소라 터보가 생성한 영상들을 비교했는데, 모든 영상의 해상도는 720p로 통일했고 영상의 길이도 5초~10초로 서로 차이가 크지 않게 했다. 테스트에서 평가자의 절반 이상이 베오2를 무비젠, 하이루오 AI, 소라 터보보다 더 선호하고 프롬프트 준수(prompt adherence)에서도 더 우수한 것으로 평가했다. 클링 1.5와의 비교에 있어서는 49.5%가 베오2를 더 선호하고 32.6%는 클링 1.5를 더 선호한다고 했다. 베오2가 이렇게 개선된 성능을 발휘하게 된 이유 중 하나는 유튜브가 저작권을 가지고 있는 충분한 양의 학습데이터로 생성 모델을 학습시켰기 때문이라는 분석도 있다.¹²⁾

유튜브는 2023년에 이미 크리에이터들이 '쇼츠' 제작을 위해 생성형 AI 기능인 '드림 스크린'을 활용할 수 있도록 했다. 드림 스크린은 텍스트 설명을 입력하면 쇼츠용 이미지 또는 영상을 생성할 수 있는 기능이다. 최근 유튜브는 미국, 캐나다, 호주, 뉴질랜드에서 드림 스크린에 베오2를 통합해 사용자가 독립적으로 쇼츠용 영상 클립을 생성할 수 있도록 했다고 발표했다.¹³⁾



출처 : 구글 딥마인드 홈페이지 (2024)

두 기업의 벤치마크 테스트 결과 공개는 자사가 새로 공개하는 서비스가 다른 경쟁사의 서비스보다 더 완성도가 높은 영상을 생성할 수 있다는 것을 알리기 위한 의도로 보인다. 그러나 클링 1.5의 경우

메타(무비젠)의 벤치마크 테스트 결과와 구글 답마인드(베오2)의 벤치마크 테스트 결과에서 모두 우수한 결과를 내는 것을 보면 벤치마크 테스트가 어느 정도 일관된 결과를 산출하고 있다고도 볼 수 있다. 그리고 여러 기업으로부터 새로운 서비스가 빠르게 출시되는 상황에서 이러한 벤치마크 테스트 결과는 이후에 나오는 서비스가 이전에 출시된 서비스보다 더 개선된 영상을 생성하고 있다는 것을 유추할 수 있도록 해주며 이 분야의 기술 발전 속도가 매우 빠르다는 것을 알려준다.

동영상 생성형 AI에 대한 벤치마크는 학계에서도 활발하게 개발되고 있다.¹⁴⁾ 다양한 동영상 생성형 AI 서비스가 출시되고 있는 상황에서 모든 서비스에 동일한 조건을 부여해 성능을 비교, 평가하는 것에 대한 연구는 앞으로도 더욱 많이 수행될 것으로 보인다.



동영상 생성형 AI의 적용 사례

현재 출시된 동영상 생성형 AI 서비스들은 드라마, 영화 등 고화질의 영상을 제작하기에 아직 많은 기술적 한계를 가지고 있다. 그러나 스포츠 동영상, 온라인 광고, 단편 영상 제작 분야에서는 활용이 시작되고 있다.

스튜디오 프리윌루전의 권한슬 감독은 2024년 2월 제1회 두바이 국제 AI 영화제에서 《One more pumpkin(원 모어 펌킨)》으로 대상과 관객상을 수상했다. 당시에는 T2I(Text-to-Image)와 I2V(Image-to-Video)를 사용해서 영상을 생성했는데 스테이블 디퓨전(Stable Diffusion)으로 텍스트에서 이미지를 생성한 후 피카랩스(Pika Labs)로 이미지에서 영상을 생성했다. 이 작품은 당시에 실험적인 작품으로 호평을 받았지만, 등장인물과 배경의 일관성이 다소 떨어지고 등장인물들의 표정이나 동작이

11. 중국의 시기업 미니맥스가 2024년 3월 출시한 LLM 서비스이며, 2024년 9월 T2V 모델을 공개했다.

12. Hosalikar, Veo2 vs. Sora AI: Which Is a Better Video Generation Model? Gumlet. 2025. 1. 5.

Roberts-Islam, 2025's AI Video Showdown: Comparing Google Veo 2 And OpenAI Sora. Forbes. 2025. 3. 6.

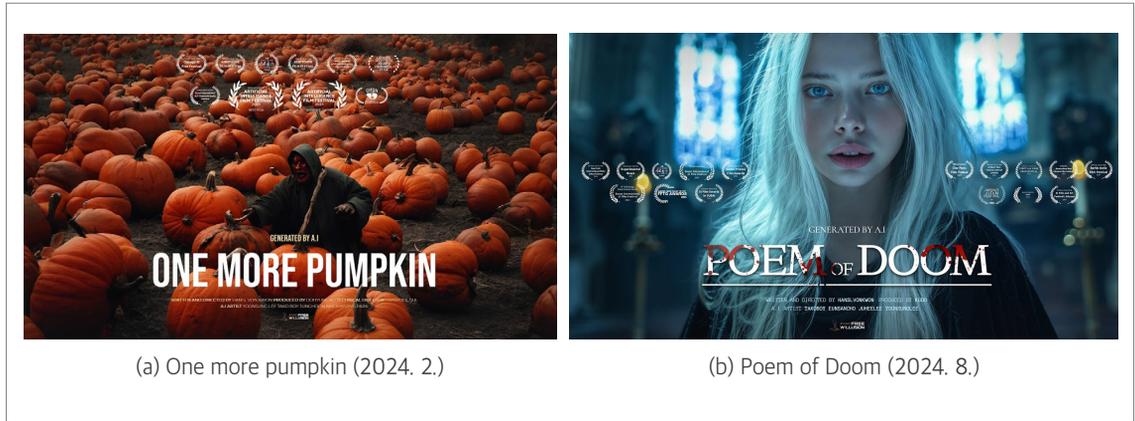
13. Berrada, D. Imagine it, create it: Veo 2 is coming to YouTube Shorts. YouTube Official Blog. 2025. 2. 13.

14. Qi, Z., Shi, P., Wang, S., Zhang, C., Zhao, F., Ying, Z., Pan, D., Yang, X., He, Z., & Dai, T. T2VEval: Benchmark Dataset and Objective Evaluation Method for T2V-generated Videos. arXiv:2501.08545. 2025

Sun, K., Huang, K., Liu, X., Wu, Y., Xu, Z., Li, Z., & Liu, X. T2V-CompBench: A Comprehensive Benchmark for Compositional Text-to-video Generation. arXiv:2407.14505. 2024

어색하다는 평을 받았다.¹⁵⁾ 스튜디오 프리윌루전은 8월에 《Poem of Doom(포엠 오브 돔)》이라는 단편 영상을 공개했는데 이 작품은 이전의 작품보다 상당 부분 발전되고 마치 실사에 가까운 영상을 생성했다는 평가를 받고 있다.¹⁶⁾

<그림 5> 생성형 AI 활용 단편 동영상 제작 사례



(a) One more pumpkin (2024. 2.)

(b) Poem of Doom (2024. 8.)

출처 : 스튜디오 프리윌루전 유튜브 채널¹⁷⁾

일부 온라인, 모바일용 광고영상에도 생성형 AI가 활용되기 시작했다. 현대자동차는 2024년 6월에 생성형 AI를 활용해 제작한 온라인 광고영상 「영원히 달리는 자동차」 총 3부작을 유튜브에 공개했다. 이 영상은 별도의 촬영 없이 시로만 생성된 영상과 음악을 활용했으며 등장인물인 현대차 연구원 등도 모두 시로 제작됐다. 이 영상을 제작하는데 총 15개가 넘는 소프트웨어가 활용된 것으로 알려져 있다.¹⁸⁾ 미국에서는 장난감 회사인 토이저러스(ToysRus)가 오픈AI 소라와 VFX 기술을 활용해 약 66초 길이의 광고영상을 제작했다. 이 영상은 토이저러스의 창업자 찰스 라자루스의 어린 시절과 브랜드 마스코트인 기린 제프리 이야기를 영상으로 제작한 것이었다. 공식적인 작품은 아니지만 2024년 7월에는 한 누리꾼(hellolaco)이 런웨이의 젠-3 알파를 이용해 볼보 자동차의 광고영상을 만들어서 올렸다. 이 누리꾼이 약 49초의 광고영상을 제작하는 데는 딱 하루가 소요됐다. 로고나 번호판의 오류와 두 대의 자동차가 등장하는 경우 등 세부적인 오류는 어도비의 애프터이펙트로 수정해 영상의 완성도를 높였다.¹⁹⁾

현재 동영상 생성형 AI 서비스로 제작된 영상들에는 공통점이 있다. 첫째 짧은 영상을 제작하더라도 하나의 동영상 생성형 AI 서비스만으로는 영상 제작을 하는데 한계가 있어서 LLM, T2I, 음성 더빙,

<그림 6> 생성형 AI 활용 광고영상 제작 사례



(a) 현대자동차 광고 「영원히 달리는 자동차」(2024. 6.)



(b) 미국 토이저러스 광고(2024. 6.)

출처 : 현대자동차 유튜브 채널²⁰⁾, 토이저러스 유튜브 채널²¹⁾

음향·음성 생성 등 복수의 생성형 AI 서비스를 함께 이용하거나, 생성 영상을 만든 이후에 추가적인 후반작업을 수행했다는 것이다. 이는 비록 많은 기업이 동영상 생성형 AI 시장에 참여해 이 분야에서 빠른 기술적 발전을 이루고 있지만, 영상제작자가 프롬프트 입력만으로 영상을 제작하기에는 아직 기능적 미흡함이 있다는 것을 의미한다.

둘째 동영상 생성형 AI 서비스를 활용한 영상의 제작은 아직 모바일 또는 온라인에서 유통되는 영상에 머물러 있다. 현재 출시된 생성형 AI 서비스로는 등장인물(배우)에 대해 표정, 뉘앙스, 타이밍 등 세밀한 디렉팅이 어려우며 시간적 일관성을 유지한 긴 영상의 생성이 어렵다. 영상은 약 3, 4초가 지나면 화면이 무너지는 현상이 나타나고 등장인물의 동작이 부자연스러운 부분이 나타난다. 생성형 AI 기술을 활용해 생성한 영상은 아직 화질, 시간적 일관성, 물리법칙의 준수 등에 있어서 완성도가 다소 낮기 때문에 대형 화면에서 긴 시간 시청하는 영상보다는 소형 화면에서 짧은 시간 시청하는 영

15. No, K. AI a tool, not a threat, says 31-year-old AI filmmaker. The Korea Herald. 2024. 8. 17.

16. 조유빈, 2025. 1. 1. '위협'인가 '공생'인가...AI 콘텐츠의 명과 암. 시사저널.

17. 스튜디오 프리월루션, One More Pumpkin, Poem of Doom

18. 김지연, 2024. 6. 14. "100% 생성형 AI로 만들었다"...현대차 신규 광고 '화제'. 매일경제.

19. 유다정, 2024. 7. 15. "볼보, 보고 있나?"... '런웨이 젠3'로 하루만에 만든 100% AI 광고에 업계 충격. 뉴데일리 경제.

20. <https://www.youtube.com/watch?v=vL1F3JJvi78>

21. https://www.youtube.com/watch?v=F_WflzYGlg4

상 위주로 제작이 되고 있다.

이용자 측면에서는 어설피게 닮은 가짜를 보면서 불쾌감을 느끼는 ‘불쾌한 골짜기(uncanny valley)’가 생성형 AI로 만들어진 영상에 대한 수용을 늦출 수 있다.²²⁾ 그리고 이런 이용자의 거부감에 대한 우려로 인해 영상 제작자들이 동영상 생성형 AI를 도입하기를 주저하게 될 수도 있다.

그럼에도 불구하고 향후 동영상 생성형 AI가 단기간에 어느 분야에 활용될 수 있을지 다양한 전망이 나오고 있다. 지금도 스포츠 콘텐츠 제작이나 제한적이거나 드라마, 영화의 짧은 장면을 제작하는 데 활용되고 있지만 향후에 다른 분야에도 활용이 확대될 것으로 예상된다. 전통적인 영상 제작 방식보다 비용과 시간을 크게 줄일 수 있는 장점을 활용할 수 있으면서 영상의 품질과 시간적 일관성이 다소 떨어지는 단점이 허용되는 분야부터 단계적으로 확산될 것이다. 영상 제작 분야에서는 영상 기획 단계에 활용할 수 있는 프리비즈, 스토리보드 등을 제작하는 데 적용할 수 있을 것으로 전망된다.²³⁾ 영상의 품질보다는 명확한 컨셉을 전달하는 것으로 충분한 콘텐츠를 제작하는 데 활용이 가능할 것이라는 예상이다. 이런 관점에서 영상의 완성도에 대한 기대치가 비교적 낮은 교육 콘텐츠를 제작하는 데에도 활용될 수 있을 것으로 예상된다.²⁴⁾

비록 생성형 AI로 제작한 동영상이 실사 영상보다 품질은 다소 떨어지지만 교육 효과에 있어서 큰 차이가 나지 않는다면 교육 분야에서 동영상 생성형 AI의 이용은 확대될 수 있다. 실제로 많은 연구에서 생성형 AI로 제작한 동영상이 실사 영상과 비교해 교육 효과 측면에서 더 우수하거나 둘 간에 큰 차이를 보이지 않는다는 것을 보여주고 있다.²⁵⁾

part 5 — 마치며

동영상 생성형 AI 서비스와 관련해서는 기술적 한계뿐만 아니라 제도적으로도 해결해야 할 점이 많다.

생성형 AI 서비스의 학습데이터에 대한 저작권 문제는 아직 제도적으로 명확한 기준이 확립되어 있지 않은 상황이다. 동영상 생성형 AI 모델의 학습을 위해서는 텍스트 기반 LLM, 이미지 생성형 AI 모델보다 더 많은 양의 학습데이터가 필요하다.²⁶⁾ 학습데이터(동영상-텍스트 쌍)의 양이 적으면 생성된 영

상에서 등장인물의 움직임이 크거나 긴 동작을 할 때 의도하지 않은 비현실적인 변화가 발생하거나 깜빡임(flickering) 현상이 발생할 수 있다. 그리고 개별 케이스에 대해 정확한 인과 관계가 반영된 물리 세계를 구현하기가 어렵다.

AI 도구를 사용해 허구의 이야기를 영상으로 만드는 데 비용과 시간을 획기적으로 줄일 수 있게 되었기 때문에 불법적인 정보를 담은 영상을 제작해 유통하는 것도 쉬워졌다. 따라서 이를 어떻게 규제할 것인지에 대해서도 고민이 필요하다. 오픈AI는 소라 터보를 공개하면서 아동 성적 학대 자료 및 성적 뎀페이크를 생성하는 것을 차단하고 뎀페이크 악용을 방지하기 위해 생성 영상에 대해 특히 엄격한 심사 기준을 설정했다고 밝혔다. 그리고 허위 정보, 불법 콘텐츠 및 안전 문제를 평가하는 전문가 그룹(레드팀)을 운영하면서 서비스 운영 상태를 테스트하고 있다. 그러나 기업의 이런 노력이 있음에도 불구하고 가짜뉴스, 선정적 콘텐츠 제작에 대한 우려는 여전히 큰 상황이다.

최근 1~2년 동안 동영상 생성형 AI 기술은 매우 빠르게 발전하고 있다. 향후 이 기술이 스포츠 콘텐츠 제작뿐만 아니라 교육, 게임, 광고·마케팅 등 다른 영역에서 활용될 가능성이 커 보인다. 산업적 활용 방안뿐만 아니라 제도적 보완점에 대한 연구와 대비가 필요한 시점이다.

-
22. 강지영, 생성형 인공지능 기반 버추얼 휴먼의 언캐니 밸리 효과에 대한 연구. 융복합지식학회논문지, 12(2), 37-49. 2024
 23. 박수호, AI가 만든 영화로 국제영화제 대상 ... 광고도 AI로 만드는 '이 남자'. 매일경제. 2024. 7. 6.
 24. Liu, Y., Zhang, K., Li, Y., Yan, Z., Gao, C., Chen, R., Yuan, Z., Huang, Y., Sun, H., Gao, J., He, L., & Sun, L. Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models. arXiv:2402.17177. 2024
 25. LLeddo, J., & Garg, K. Comparing the effectiveness of AI-powered educational software to human teachers. International Journal of Social Science and Economic Research, 6(3), 953-963. 2021
 - Netland, T., von Dzengelevski, O., Tesch, K., & Kwasnitschka, D. Comparing human-made and AI-generated teaching videos: An experimental study on learning effects. Computers & Education, 224, 105164. 2025
 - Pi, Z., Deng, L., Wang, X., Guo, P., Xu, T., & Zhou, Y. The influences of a virtual instructor's voice and appearance on learning from video lectures. Journal of Computer Assisted Learning, 38(6), 1703-1713. 2022
 26. 이기석·이승욱·윤민성·유정재·오아름·최인문·김대욱, 디지털 에셋 창작을 위한 생성형 AI 기술 동향 및 발전 전망. 전자통신동향분석, 39(2), 33-42. 2024

레퍼런스



참고문헌

- 강지영 (2024). 생성형 인공지능 기반 버추얼 휴먼의 언캐니 밸리 효과에 대한 연구. 융복합지식학 회논문지, 12(2), 37-49.
- 김지연 (2024. 6. 14.). "100% 생성형 AI로 만들었다"...현대차 신규 광고 '화제'. 매일경제.
- 박수호 (2024. 7. 6.). AI가 만든 영화로 국제영화제 대상 ... 광고도 AI로 만드는 '이 남자'. 매일경제.
- 우리말샘 (2025).
- 유다정 (2024. 7. 15.). "불보, 보고 있나?"... '런웨이 젠3'로 하루만에 만든 100% AI 광고에 업계 충격. 뉴데일리 경제.
- 이기석·이승욱·윤민성·유정재·오아름·최인문·김대옥 (2024). 디지털 에셋 창작을 위한 생성형 AI 기술 동향 및 발전 전망. 전자통신동향분석, 39(2), 33-42.
- 조유빈 (2025. 1. 1.). '위협'인가 '공생'인가...AI 콘텐츠의 명과 암. 시사저널.
- Berrada, D. (2025. 2. 13.). Imagine it, create it: Veo 2 is coming to YouTube Shorts. YouTube Official Blog.
- Girdhar, R., Singh, M., Brown, A., Duval, Q., Azadi, S., Rambhatla, S. S., Shah, A., Yin, X., Parikh, D., & Misra, I. (2024). Emu video: Factorizing text-to-video generation by explicit image conditioning. In ECCV.
- Hosalikar, S. (2025. 1. 5.). Veo2 vs. Sora AI: Which Is a Better Video Generation Model? Gumlet.
- Leddo, J., & Garg, K. (2021). Comparing the effectiveness of AI-powered educational software to human teachers. International Journal of Social Science and Economic Research, 6(3), 953-963.
- Liu, Y., Zhang, K., Li, Y., Yan, Z., Gao, C., Chen, R., Yuan, Z., Huang, Y., Sun, H., Gao, J., He, L., & Sun, L. (2024). Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models. arXiv:2402.17177.
- Meta (2024. 10. 16.). Movie Gen: A Cast of Media Foundation Models.
- Netland, T., von Dzengelevski, O., Tesch, K., & Kwasnitschka, D. (2025). Comparing hu-

man-made and AI-generated teaching videos: An experimental study on learning effects. *Computers & Education*, 224, 105164.

- No, K. (2024. 8. 17.). AI a tool, not a threat, says 31-year-old AI filmmaker. *The Korea Herald*.
- Qi, Z., Shi, P., Wang, S., Zhang, C., Zhao, F., Ying, Z., Pan, D., Yang, X., He, Z., & Dai, T. (2025). T2VEval: Benchmark Dataset and Objective Evaluation Method for T2V-generated Videos. *arXiv:2501.08545*.
- Roberts-Islam, M. (2025. 3. 6.). 2025's AI Video Showdown: Comparing Google Veo 2 And OpenAI Sora. *Forbes*.
- Sharma, N. (2025. 1. 15.). Sora vs Veo 2: Which One Creates More Realistic Videos? *Analytics Vidhya*.
- Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O., Parikh, D., Gupta, S., & Taigman, Y. (2023). Make-A-Video: Text-to-video generation without text-video data. In *ICLR*.
- Sun, K., Huang, K., Liu, X., Wu, Y., Xu, Z., Li, Z., & Liu, X. (2024). T2V-CompBench: A Comprehensive Benchmark for Compositional Text-to-video Generation. *arXiv:2407.14505*.
- Pi, Z., Deng, L., Wang, X., Guo, P., Xu, T., & Zhou, Y. (2022). The influences of a virtual instructor's voice and appearance on learning from video lectures. *Journal of Computer Assisted Learning*, 38(6), 1703-1713.



참고 사이트

- 구글 딥마인드 : <https://deepmind.google/>
- 메타 : <https://www.meta.com/kr/>
- 오픈AI 홈페이지 : <https://openai.com/>
- 우리말샘 홈페이지 : <https://opendict.korean.go.kr/main>
- 스튜디오 프리윌루전 <https://www.studiofreewillusion.com/>
- 현대자동차그룹 유튜브 : <https://www.youtube.com/@HyundaiLive>
- 토이저러스 유튜브 : <https://www.youtube.com/@toysrus>