

생성형 AI 기술의 현재와 비디오 생성 AI 모델 소라(Sora)

글. 한영주 한국방송통신전파진흥원 연구위원 / 언론정보학 박사

#생성형 AI의 이용 경험

#비디오 생성 AI 모델

#소라(Sora)

ChatGPT가 출시된 이후, AI 기술을 향한 관심이 생성형 AI로 옮겨가며 AI 열풍이 한껏 고조되었다. 올해 초 미국 라스베이거스에서 열린 세계가전전시회(The International Consumer Electronics Show, CES)와 지난 3월 미국 텍사스에서 개최한 사우스 바이 사우스 웨스트(South by Southwest, SXSW)에서도 핵심 주제를 통해 생성형 AI로 증폭된 AI 열풍을 확인할 수 있었다. 세계적인 국제 행사에서도 AI 기술이 핵심 주제로 화두에 올랐는데, 이전 AI 논의와 다른 점은 기술적 측면만이 아니라, 일상에 스며들고 있는 AI 기술, 즉 ‘콘텐츠에

녹아든 AI 기술’에 초점을 맞추고 있었다.

생성형 AI의 폭발적인 관심과 성장은 실제로 생성형 AI 기술에 투자된 벤처캐피탈 자금에서도 확인해 볼 수 있다. 2022년 4분기 45억 달러에서 2023년 1분기 129억 달러로 무려 3배 가까이 늘어났는데, 생성형 AI 기술을 통해 경쟁사보다 앞서나가기 위한 산업계의 움직임이 고스란히 반영된 것으로 볼 수 있다.

AI는 혁신적이지만 어려운 기술이었다면, 생성형 AI는 누구나 쉽게 접근하고 활용할 수 있는 대중적인 기술이다.

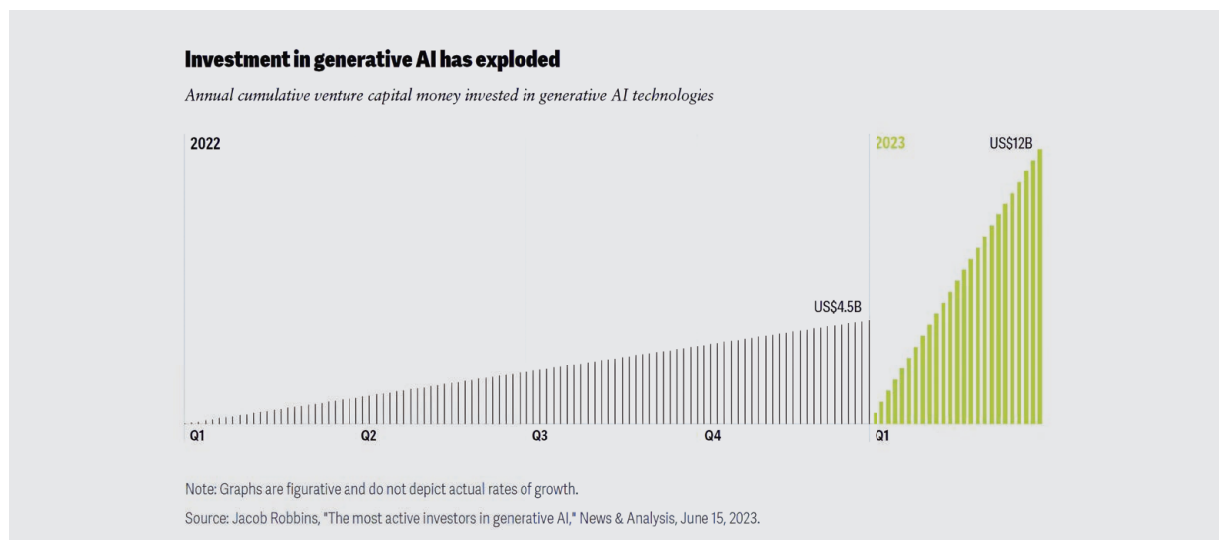


그림 1. 벤처캐피탈의 생성형 AI 투자 규모 / 출처 : Deloitte (2023).

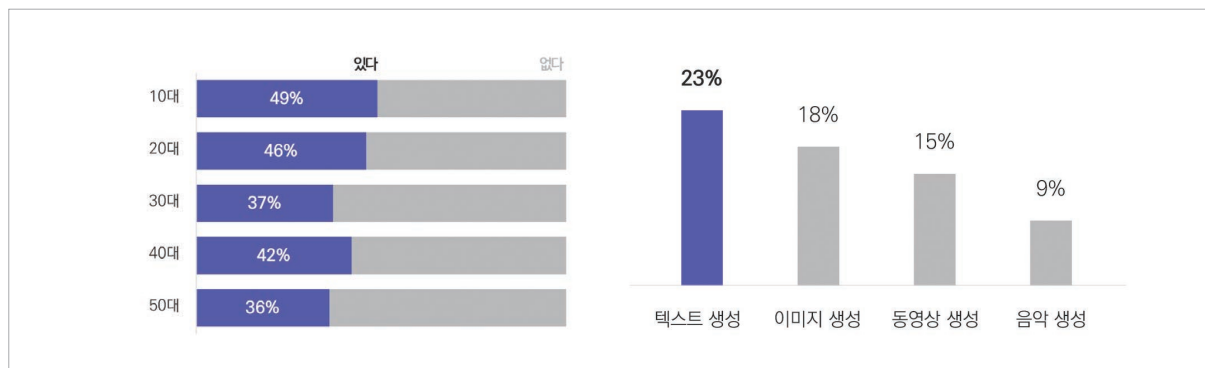


그림 2. 생성형 AI 이용 경험(왼쪽)과 이용 종류(오른쪽) / 출처 : 메조미디어 (2023).

생성형 AI의 이용 경험

생성형 AI가 주목받는 이유는 복잡한 연산과 수식이 필요한 머신러닝, 딥러닝 등 기술적 지식이 없어도 프롬프트에 간단한 명령어만 입력하면 누구나 이미지, 비디오, 오디오, 컴퓨터 코드를 쉽고 편리하게 생성할 수 있다는 점이다. 실제로 생성형 AI 이용 행태를 조사한 설문 결과에서 10명 중 4명이 생성형 AI를 이용한 경험이 있다고 응답했는데, 기존의 인터넷과 모바일을 통해 사용할 수 있어서 이질감이 없고 접근성은 높아 향후 생성형 AI의 이용 경험은 더욱 늘어날 것으로 보인다.

연령대별로는 신기술에 대한 친화력이 높은 10대와 20대가 다른 연령대에 비해 생성형 AI의 이용 비율이 높았다. 아무래도 디지털 미디어와 함께 성장해 온 10대와 20대는 디지털 네이티브로서 생성형 AI에 대한 거부감 없이 신기술을 자연스럽게 받아들이는 것으로 보인다. 이들이 주로 이용해 본 생성형 AI 종류는 텍스트 생성(23%)이 가장 많았고, 그다음은 이미지 생성(18%), 동영상 생성(15%), 음악 생성(9%) 순으로 나타났다.

또한, 생성형 AI 챗봇은 ChatGPT 이용 경험이 가장 많았다. 주로 생성형 AI 챗봇을 이용하는 목적은 호기심 때문(35%)이라는 답변이 가장 많았지만, 빠른 답변을 얻기 위해서(24%)와 과제나 업무 시 활용하기 위해서(19%)라고 답변한 비율도 적지 않아, 생성형 AI 챗봇 서비스를 기점으로 여러 형태의 전문화된 서비스로 확장해 볼 수 있다. 생성형 AI 챗봇의 만족도는 보통이다(45%), 만족한다(43%), 불만족한다(12%) 순으로 대체로 만족하는 편이었지만 보통이라고 응답한 비율이 높기 때문에 질문에 대한 이해력을 높이고 질문에 적합한 답변을 제공할 수 있는 기술 고도화가 계속될 필요가 있다.

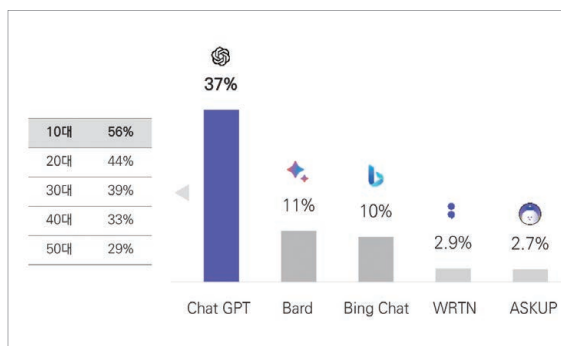


그림 3. 생성형 AI 챗봇 서비스의 이용 경험 / 출처 : 메조미디어 (2023).

비디오 생성 AI 모델 소라(Sora)

생성형 AI가 산업계를 주축으로 빠르게 확대될 것으로 전망되는데, 생성형 AI의 산업적 기대감은 지난 2월 15일, 오픈AI(OpenAI)에서 '텍스트 투 비디오(text-to-video)'를 위한 대규모 생성 모델(LLM) '소라(Sora)'를 발표하며 다시 한번 크게 상승하였다. 산업계 전반이 AI 기술을 두고 새로운 생태계를 구성하는 가운데, 소라의 소식은 ChatGPT 이후 또다시 목격한 충격을 안겨줬다. 분명 머지않은 시기에 AI로 영상물을 창작하게 되리라 누구도 의심치 않았지만, 그 시점이 이리도 빠르게 다가올 줄은 예상하지 못한 결과였다. 오픈AI에서 공개한 소라의 샘플 영상은 짧지만, 압도적인 퀄리티를 선보였다. 소라는 텍스트 명령으로 최대 1분 길이의 고품질 비디오를 생성할 수 있는데, 달리는 지하철 안에서 창에 비치는 그림자까지 디테일을 살려는 놀라운 실력을 보여줬다. 이는 소라가 물리적 법칙을 이해하고 동영상을 일관성 있게 만들어낼 수 있다는 것을 보여준다. 오픈AI 소라 카테고리에 방문하면, 제일 상단에 밤거리를 거닐고 있는 선글라스를 쓴 여성의 모습을 샘플 영상으로 볼 수



“스타일리시한 여성이 따뜻하고 빛나는 네온과 애니메이션 도시 간판으로 가득한 도쿄 거리를 걷고 있다. 그녀는 검은색 가죽 재킷, 빨간색 긴 드레스, 검은색 부츠를 신고 검은색 지갑을 들고 있다. 그녀는 선글라스와 빨간 립스틱을 착용한다. 그녀는 자신감 있고 자연스럽게 걷는다. 거리는 축축하고 반사되어 다채로운 조명의 거울 효과를 만들어낸다. 많은 보행자들이 걸어 다닌다”

그림 4. 비디오 생성 AI 모델 소라로 만든 샘플 영상과 프롬프트 / 출처 : OpenAI Official page.

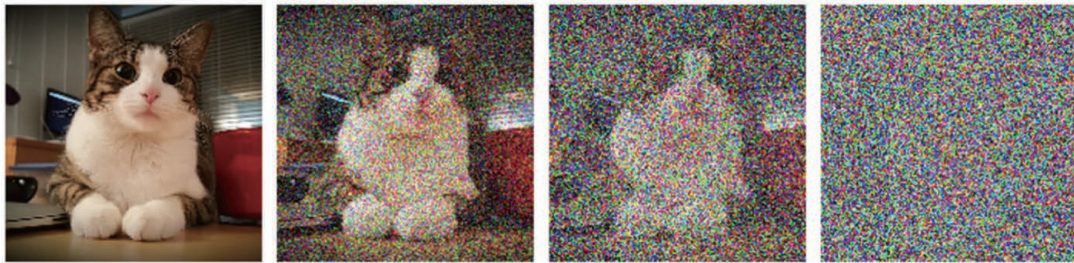
있다. 텍스트만으로 현실 같은 영상을 만들어낼 수 있으니, 정말 놀라지 않을 수가 없다.

소라는 생성형 AI 모델 중 하나인 확산 모델(Diffusion Model)을 기반으로 GPT, DALL-E와 마찬가지로 트랜스포머(Transformer) 알고리즘에 바탕하여 만들어졌다. 확산 모델은 입력한 이미지에서 여러 단계에 걸쳐 노이즈(Noise)를 추가한(encoder) 다음, 다시 여러 단계에 걸쳐 노이즈를 제거해서(decoder) 최초 입력한 이미지와 유사한 확률 분포를 나타내는 이미지를 생성한다. 확산 모델에서 노이즈는 이미지나 비디오를 생성하는 과정에서 무작위로 추가하는 데이터를 의미한다. 이미지에 노이즈가 추가될 때는 포워드 디퓨전 프로세스(Forward Diffusion Process)를 거쳐 고정된 정규분포로 생성된 노이즈가 더해지며, 반대로 이미지에서 노이즈가 제거될 때는 리버스 디퓨전 프로세스(Reverse Diffusion

Process)를 진행하여 학습된 정규분포로 생성된 노이즈가 빠지게 된다. 또한 데이터 학습은 트랜스포머에서 텍스트를 토큰(token) 단위로 치환하는 것처럼 이미지나 비디오를 패치(patches) 단위로 치환하는 방식이다.

소라 이외에도 구글의 루미에르(Lumiere), 메타의 에뮤 비디오(Emu Video), 스테빌리티AI의 스테이블 비디오 디퓨전(Stable Video Diffusion) 등 여러 IT 전문 기업에서 비디오 제작을 위한 생성 AI 모델을 선보이고 있다. 이 중에서 구글의 비디오 생성 AI 모델 루미에르(Lumiere)는 소라보다 한 달 먼저 개발자 커뮤니티 사이트 깃허브(GitHub)를 통해 공개되었다. 구글은 초보자도 루미에르를 이용해서 창의적인 콘텐츠를 생성하게 될 것이라고 설명하였다. 루미에르는 프롬프트에 사진과 명령어를 입력해서 영상을 제작하거나 영상 속의 스타일을 변경할 수 있다. 또한 화면 일부가 잘리거나 가려진

Forward Diffusion Process



Reverse Denoising Process

그림 5. 비디오 생성 AI 모델 소라의 이미지 노이즈 과정 / 출처 : 김수진 (2024. 2. 20).

영상을 복원하는 기능도 갖고 있다.

구글은 비디오 생성용 시공간 확산 모델(Space-Time Diffusion Model for Video Generation)을 사용해서 텍스트 투 비디오(text-to-video), 이미지 투 비디오(image-to-video), 특정 스타일로 변환, 스틸 이미지 애니메이션 기반으로 5초 분량의 클립 생성, 사진에서 선택 부분만 모션 가져오기 등 기능을 구현한다. 루미에르의 기능 중에서 풀프레임, 저해상도 비디오 생성에 탁월한 인페인팅(inpainting) 기능은 주목해 볼 만하다.

그런데도 단연 오픈AI 소라의 평가가 압도적으로 높은 편이다. 향후 소라를 통해 콘텐츠에 대한 아이디어나 기획을 직관적인 영상으로 시뮬레이션할 수 있다. 본격적인 콘텐츠 제작에 앞서, 소라를 이용한 사전시각화 프리비즈(previs)를 실행해 볼 수 있다. 즉 소라와 같은 생성형 AI를 통해 사전시각화를 진행하여 콘텐츠의 제작이 완성되기 전에 대략적인 내용과 흐름을 파악해서 투자금 등 재원확보와 흥행 가능성 예측에 활용해 볼 수 있다. 소라는 아직 복잡한 공간에 대한 시뮬레이션의 한계, 원인과 결과에서 일부분 사례 부족, 프롬프트 공간의 세부 정보 혼동 등이 한계이며 어떤 동영상으로 학습했는지 구체적으로 알 수 없다. 저작권 문제 등 정식 출시까지 해결할 과제가 남아있고 당장 기업 실적에 반영되지 않지만, 그간 오픈AI가 보여준 강력한 생성 기술은 소라를 기대하기에 충분해 보인다.

AI 기술 채택의 어려움

미디어 산업계에서 AI 기술은 콘텐츠 제작, 의사결정 지원, 프로세스 자동화까지 거의 모든 비즈니스를 장악했다고 말해도 과언은 아니다. 그러나 대형 기업을 제외한 대부분 사업자는 기본적인 AI 기술마저 온전히 자사 시스템에 탑재하지도 사업에 접목하지도 못한 상태이다. 또한 미디어 관련 종사자들이 AI 기술을 다루기 위한 이론과 실무를 이해하기도 전에 AI 기술은 하루가 다르게 발전하고 변화하며 생성형 AI 소라와 같은 새로운 모델과 기능을 선보이고 있어서, 이 흐름만을 따라가기도 만만치 않을 것이다.

AI는 높은 수준의 연구와 성공 사례로 인상적인 결과를 달성하고 있지만, 아이러니하게도 실제로는 경제성, 환경적 영향, 가치 부재와 같은 이유가 AI 채택에 걸림돌이 되고 있다. 현실은 우리의 예상과 달랐다.

이유는 다르지만, AI 기술을 수용한 대기업에서도 AI 기술 사용에 들어가는 엄청난 비용을 감당하기를 버거워한다. 이보다 규모가 작은 중소기업과 스타트업은 AI 기술을 활용할 엄두조차 내기 어려운 상황이다.

전 세계에서 AI 기술 수준 1순위를 차지하는 미국에서도 AI 채택 수준은 저조한 편이다. 미국 기업들의 AI 채택 비율은 두 자릿수를 넘지 못한다. AI 도입의 변화를 연구한 토론토 대학의 맥엘헤란(McElheran) 교수는 AI 기술은

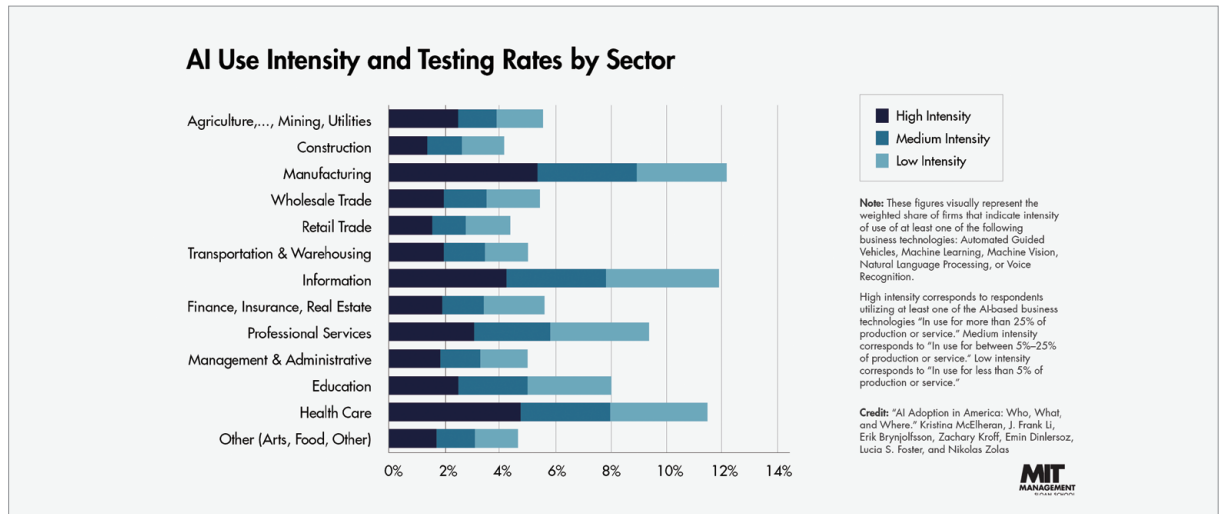


그림 6. AI 사용 강도와 테스트 비율 / 출처 : Eastwood, B. (2024. 2. 7).

어디에나 존재하지만, 실상 데이터를 살펴보면 사람들의 관심과는 다르게 실행하기가 어렵다며 디지털 시대가 도래했지만 전 영역에 고르게 도달한 것은 아니라고 말하였다. 그의 최신 논문에서 AI가 대기업과 제조, 의료 등의 특정 산업군에서만 사용되고 있고 산업 전반에서 고르게 채택되지 못한 것을 결과로 도출하였다. 2017년 미국 기업 6%만이 AI를 사용했으며 최근 2023년 11월 인구조사국이 실시한 조사에서는 기업 중 4% 미만만이 AI를 사용해서 상품과 서비스를 생산하는 것으로 나타났다. 이는 기존에 AI를 채택한 기업이 최근에는 상품과 서비스의 생산에 참여한 것으로 짐작해 볼 수 있다. AI 기술이 꾸준히 발전하며 생성형 AI 기술로 진화했지만, AI를 통해 얻는 기업 효과는 의미 있는 가치로 변환되지 못한 것이다.

벤처캐피탈 회사 GCP(General Catalyst Partners) CEO 헤먼트 타네자(Hemant Taneja)는 생성형 AI 기술은 대중적인 플랫폼에는 적합하지 않으며 사람과 프로세스를 생산적으로 만들고 대규모 조직을 변화시키는데 이상적이라고 설명하였다. 그간 생성형 AI를 향한 산업적 기대와 다르게 타네자처럼 일부 전문가들은 다른 견해를 제시하고 있다. 그러나 생성형 AI 기술은 이 순간에도 계속 진화하고 있어서 산업적인 적합성을 단정하기에는 아직 이른 감이 있다. 지금은 다양한 생성 기능으로 AI 기술이 어떠한 방향으로 흘러갈지 주목해 보자. 

참고문헌

- 김수진 (2024. 2. 20). 생성AI 게임체인저: 비디오 만들어주는 OpenAI의 Sora. 미래에셋증권 리포트.
- 뉴스1 (2024. 1. 27). 영상 속 의상도 바꿔 준다…구글, 영상 생성 AI ‘루미에르’ 공개[아무Tech].
- 메조미디어 (2023). 생성형 AI에 대한 소비자 인식조사. Insight M.
- Deloitte (2023). Tech Trends 2024. Deloitte Insights.
- Eastwood, B. (2024. 2. 7). The who, what, and where of AI adoption in America. MIT Sloan School of Management.