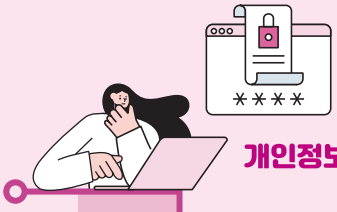
생성형

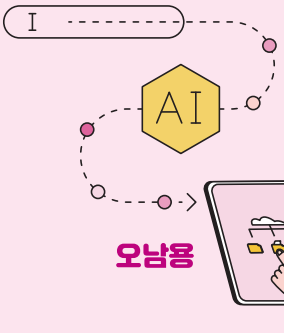
AI 윤리 가이드북



개인정보·인격권

허위조작정보

저작권



오남용

책임성

방송미디어통신위원회 NIA 한국지능정보사회진흥원

작성 (가나다순)
김진욱 변호사, 백경태 변호사(법무법인 신원)
전창배 이사장(IAAE 국제인공지능 & 윤리협회)
정필운 교수(한국교원대학교)
주윤경 수석(한국지능정보사회진흥원)

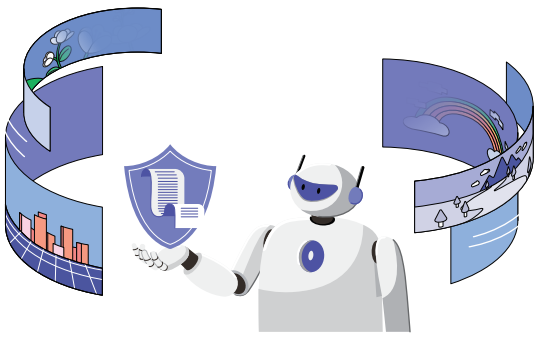
문의
주윤경 수석(한국지능정보사회진흥원, jyk@nia.or.kr)

방송미디어통신위원회 NIA 한국지능정보사회진흥원

Contents

PART 03

책임성



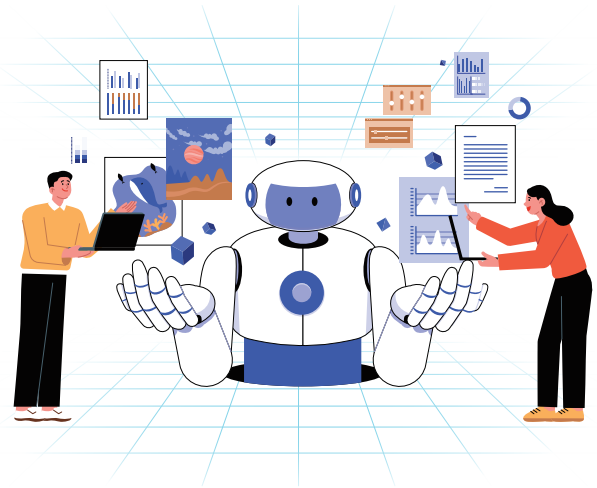
Contents

PART 04 허위조작정보

PART 05 개인정보·인격권

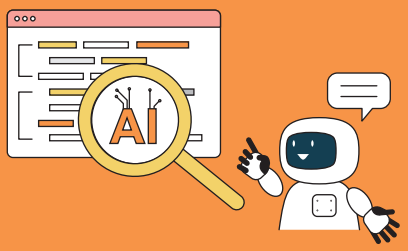
PART 06

오남용



생성형

윤리 가이드북



생성형 AI란?

01

PART 01

생성형 AI란?

생성형 AI가 무엇일까요?

생성형 AI(Generative AI)란, 대규모 데이터와 패턴을 학습하고 기존의 데이터를 활용하여 이용자의 요구에 따라 텍스트, 이미지, 비디오, 음악, 코딩 등 새로운 결과를 만들어 내는 인공지능 기술입니다. 생성형 AI는 누구나 쉽게 활용할 수 있도록 설계되었고, 굉장히 다양한 영역에서 유용하게 활용될 수 있기 때문에 인공지능 분야의 새로운 화두가 되었습니다.

지난 2016년에 구글 딥마인드(Google DeepMind)가 개발한 인공지능 바둑프로그램인 알파고(AlphaGo)와 이세돌 9단의 바둑 대결은 인공지능과 인간의 대결이라는 타이틀로 모두에게 놀라움을 안겼습니다. 그런데 실제로 많은 사람들이 직접 경험하거나 체감할 수 없는 부분이기 때문에 차츰 잊혀져 갔습니다. 그러다가 2022년 11월 OpenAI의 생성형 AI 모델 ChatGPT의 등장으로 사람들은 마치 가까운 친구와 편하게 대화를 나누듯 쉽고, 빠르고, 편리하게 인공지능을 활용할 수 있게 되자 다시금 인공지능에 대한 관심과 기대가 높아졌습니다.

스태티스타(Statista)의 조사결과 ChatGPT가 사용자 100만명을 확보하는데 걸린 시간은 단 5일이었다고 합니다. 넷플릭스가 3.5년, 인스타그램이 2.5개월이 걸린 것과 비교하면 굉장히 빠르게 확산된 것이란 걸 알 수 있습니다.

생성형 AI는 텍스트(ChatGPT, Bard, Bing, 뤼튼, AskUp 등), 이미지(DALL·E 2, Midjourney, Stable Diffusion 등), 영상(D-ID, VCAT, Runway, InVideo 등), 음성(클로바더빙, AI보이스 스튜디오, Typecast. VoxBox 등), 글쓰기(Notion AI, Magic Write 등), 프로그래밍(GitHub Copilot, Tabnine 등) 등 활용도 무궁무진합니다.

생성형 AI가 가져온 변화는 무엇일까요?

생성형 AI는 우리 일상에 빠르게 확산되며 개인, 그리고 산업에도 많은 변화를 가져왔습니다. 먼저 개인 누구나 생성형 AI를 활용해 외국어 문서를 번역할 수 있고, 방대한 양의 문서를 쉽게 요약할 수 있습니다. 음성 기능을 활용하면 회의록을 직접 쓰지 않아도 텍스트로 기록해 주고, 전화 영어를 하듯 회화 공부를 할 수도 있습니다. 또 전문적인 영역이라 여겼던 소프트웨어 코딩, 작곡, 원하는 컨셉의 이미지 제작도 해줍니다. 누구라도 필요한 일에 도움을 받을 수 있습니다. 이렇게 활용도가 높다보니 연령, 직업에 관계없이 생성형 AI를 활용하게 되면서

인공지능 활용이 일상화가 되었고, 개인이 필요로 하는 학업, 일하는 방식 등에 생산성과 효율성을 한층 높이는 변화를 가져왔습니다.

산업적 측면에서도 생성형 AI를 도입하려는 움직임이 활발합니다. 시장분석기관 한국 IDC에 따르면 2023년 전 세계 생성형 AI 시장 규모는 약 160억 달러(약 20조 9천억원)에 달하며, 2027년에는 1,430억 달러(약186조 8천억) 규모로 성장할 것으로 보입니다. 생성형 AI가 디지털 기반 기업의 주요 전략 기술로 자리매김한 것입니다.

미국 벤처 캐피털 회사 세쿼이아(SEQUOIA)가 2022년부터 업데이트해 온 생성형 AI 시장 지도를 보면 관련 분야와 기업이 점점 더 다양해지고 확대되고 있다는 것을 알 수 있습니다.

출처 : SEQUOIA, Generative AI's Act Two, 2023. 9. 20.

빅테크 기업부터 스타트업까지 많은 기업들이 생성형 AI 시장에 뛰어들고 있는 이유는 생성형 AI의 발전 가능성 때문입니다. 생성형 AI는 계속해서 더 많은 데이터를 학습하기 때문에 결과물이 지속적으로 개선될 수 있습니다. 한 예로 OpenAI의 GPT-3.5는 미국 통합 변호사 시험(Uniform Bar Exam) 하위 10% 정도였는데, GPT-4는 상위 10% 수준에 해당하는 점수를 얻었고, 미국 생물올림피아드 시험에서도 GPT-3.5는 하위 30%의 성적이었으나, GPT-4는 상위 1% 수준의 성적을 냈다고 합니다.

이처럼 빠른 속도로 꾸준히 성장하는 생성형 AI는 지난 몇 년 사이 인공지능의 엄청난 발전을 보여준 증거라고도 할 수 있습니다. 또한 인간의 창의적 상상력에 불씨가 되고, 기업의 생산성 향상에 혁신적 역할을 할 것으로 기대가 되는 잠재력을 지닌 기술이라 할 수 있습니다.

생성형 AI를 왜 윤리적으로 활용해야 하나요?

앞에서도 살펴봤듯이 생성형 AI는 기본 단계에 머무르는 것이 아니라 시간을 거듭할수록 더 많은 데이터를 학습하고 진화합니다. 혁신적인 기술로서 전에 없던 새로운 서비스를 만들어 내고 가치 창출을 극대화하며, 우리 경제와 사회에 미치는 영향이 점차 커지고 있습니다.

그러나 모든 혁신적이고 새로운 기술에도 한계는 있기 마련입니다. 기술은 긍정적인 영향과 부정적인 영향을 동시에 갖고 있는 양면적인 특성을 가졌습니다. 생성형 AI도 효율성, 생산성, 혁신성, 창조성 등의 긍정적인 특성 이외에 모호한 저작권, 개인정보유출, 허위조작정보, 정보편향, 오남용 등 부정적인 문제를 내포하고 있어서 심각한 사회적 문제를 가져올 것에 대한 우려도 있습니다.

각국 정부는 생성형 AI를 포함한 인공지능의 역기능 대응을 위한 각종 제도적 보완책 마련에 집중하고 있고, 기업도 생성형 AI의 경쟁이 본격화 되면서 자칫 역기능에 대한 우려로 이용자가 서비스를 회피하거나 피해를 입지 않도록 하기 위해 저작권 등 분쟁으로 소송이나 손해배상을 해야 할 때 서비스 제공자인 기업이 책임있는 조치를 하겠다고 앞다퉈 발표하고 있습니다.

법제도적 조치와 피해지원 등은 역기능 대응에 있어서 매우 중요합니다. 하지만 이것은 역기능 발생 후의 조치, 즉 사후적인 대응방안에 가깝다고 할 수 있습니다. 그렇다면 우리가 사전에 역기능을 예방 할 수 있는 방법은 무엇일까요? 바로 이용자의 역기능 대응 역량 강화입니다. 어떤 것이 왜 문제가 될 수 있는지 생성형 AI의 한계와 역기능이 무엇인지를 조금이라도 인식하고 있다면 그렇지 않은 것에 비해 보다 긍정적이고 생산적으로 활용할 수 있게 될 것입니다.

생성형 AI의 역기능은 무엇인가요?

최근 이슈가 되고 있는 생성형 AI의 역기능에 대해 살펴보겠습니다.

생성형 AI는 수많은 데이터를 조합해 기존 콘텐츠를 바탕으로 결과물을 생성합니다. 그래서 생성형 AI가 기존 저작물을 바탕으로 만든 콘텐츠의 소유권이 누구에게 있는지 등 저작권에 대한 부분이 모호합니다. 생성형 AI가 만든 창작물의 저작권을 인정하느냐, 마느냐에 대한 법적 기준도 현재로서는 명확하지 않아 해외에서는 작가단체, 예술가 등의 소송이 빗발치고 있습니다. 생성형 AI가 제시한 결과물이 누군가의 노력이 담긴 지식재산권일 수 있다는 생각을 하며, 생성형 AI가 만든 창작물임을 밝히는 등 책임있는 활용이 중요합니다.

저작권 문제 못지않게 개인정보 유출에 대한 우려도 생성형 AI의 역기능으로 지적되고 있습니다. 개인정보가 포함된 데이터를 학습한 후 여과 없이 노출되는 문제도 있을 수 있지만 가장 주의해야 하는 것은 내가 생성형 AI에 결과물을 얻기 위해 입력한 텍스트, 파일 등의 정보 또한 그대로 학습되어 타인에게 보여질 수 있다는 것입니다. 더욱이 최근에는 문서 요약, 번역, 제안서 작성 등 업무 효율성을 높일 수 있는 생성형 AI를 직장인이 많이 활용하면서 회사 자료 중 외부로 공개해서는 안 될 회의록, 내부 자료까지 유출되는 문제가 발생하기도 합니다. 생성형 AI를 활용할 때는 입력하는 모든 정보가 외부에 공개될 수 있기 때문에 개인정보나 민감정보가 없는지 꼭 한번 살펴봐야 합니다.

또 여기저기 퍼져있는 정보를 학습해 결과물을 제시하기 때문에 허위조작정보나, 성차별, 인종차별 등 사회적 편견이 내포된 편향된 정보를 포함될 수 있습니다. 무엇보다도 할루시네이션(Hallucination, 환각) 문제가 심각합니다. 사실이 아닌데도 정답처럼 보이는 그럴싸한 생성형 AI의 답변은 정보획득에 혼란을 가져옵니다. 생성형 AI도 완벽할 수 없다는 것을 잊지 말고 생성된 정보에 대한 진실성과 정확성을 꼼꼼하게 확인하는 교차 검증이 꼭 필요합니다.

요즘은 생성형 AI가 무엇이든 쉽고 빠르게 답을 해주니 자신이 먼저 고민해 보지 않고 무조건 생성형 AI에 요구하는 경우도 많다고 합니다. 생성형 AI에 대한 의존도가 심해지면 보면 생각하는 사고력이 떨어질 수 있고, 글을 쓰는 능력, 그림을 그리거나 악기를 연주하는 재능 등 스스로 할 수 있는 역량도 점차 낮아질 수 있습니다. 생성형 AI는 인간이 가진 훌륭한 능력인 감성과 창의성을 지원하고 효율성을 높이는 도구로서 활용해야 한다는 것을 기억해야 합니다.

마지막으로 딥페이크(첨단조작기술)를 악용해 사진, 영상 등 가짜 콘텐츠를 생성하기도 하는데, 기술이 빠르게 고도화되면서 점점 더 진짜와 가짜를 구별할 수 없을 정도로 정교해지고 있습니다. 인간중심의 가치를 무시한 딥페이크 콘텐츠의 오남용은 사회적으로나 경제적으로 매우 위험합니다. 최근에는 유명인뿐만 아니라 일반인까지 합성 대상이 확대되고 있고, 음성 데이터를 조작해 보이스피싱 등의 범죄에 악용될 소지가 있어서 문제가 더욱 심각합니다. 장난으로 시작한 딥페이크 합성물이 디지털 공간 내에 퍼질 경우 여론조작, 명예훼손 등을 유발할 위험이 있다는 것을 잊지 말고 올바르게 판단하여 활용하는 것이 중요합니다.

생성형 AI의 윤리적 활용을 위해 어떤 노력이 필요 할까요?

생성형 AI를 잘 못 쓰면 역기능이 발생할 수 있다고 해서 생성형 AI의 사용을 제한하는 것은 좋은 해법이 아닙니다. 기술은 우리가 직접 하기 어려운 것을 도와주고, 시간을 절약해 주고, 더 나은 결과를 제시해 주는 등 이로운 점도 충분하기 때문입니다.

생성형 AI를 긍정적으로 활용하기 위해서는 정부, 기업, 이용자 모두의 노력이 필요합니다. 서비스 제공 기업은 생성형 AI의 이면을 냉철하게 진단하고, 누구보다 윤리적으로 책임감 있게 기술을 설계해야 합니다. 정부도 모든 국민이 안전하게 인공지능을 활용할 수 있도록 이용자 보호를 위한 제도 보완 노력이 필요합니다.

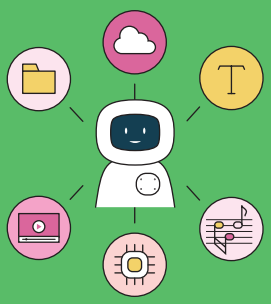
또한 학교, 기업 등은 구성원들에게 생성형 AI의 위험에 대한 교육, 가이드라인 등을 통해 생성형 AI의 기본 활용 방향, 사용할 때 지켜야 할 윤리 매뉴얼 등 “생성형 AI를 올바르게 잘 활용하는 방법”을 알려주고 공감대를 형성할 필요가 있습니다.

가장 중요한 것은 이용자 스스로 판단과 검토를 통해 생성형 AI를 안전하고 책임감 있게 활용하는 윤리적 마인드를 갖추는 것입니다. 이용자가 생성형 AI를 활용하기 전에 윤리적인 책임의 중요성과 올바르게 사용해야 할 필요성에 대해 공감하고 이해한다면 역기능은 최소화하면서, 긍정적이고 생산적으로 활용할 수 있을 것입니다.

이 가이드북은 생성형 AI를 활용하면서 발생할 수 있는 윤리적인 문제에 대해 알아보고, 이용자 스스로 예방하고 해결할 수 있는 역량을 강화하기 위해 준비되었습니다. 여러분이 생성형 AI를 윤리적이고 생산적으로 활용하는데 참고가 될 것입니다.

생성형

윤리 가이드북

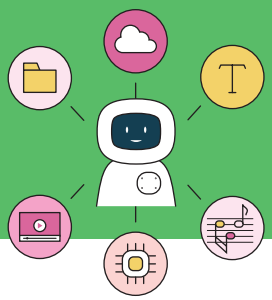


저작권

02

PART

02 저작권



챗GPT가 쏘아 올린 생성형 AI 열풍

2022년 11월 30일 오픈AI가 공개한 대화형 AI 챗봇 서비스 '챗GPT'는 사람들의 생활에 일대 변혁을 일으켰습니다. 고작 두 달 만에 사용자가 1억 명을 넘길 만큼 그야말로 챗GPT 열풍이 일었죠. 직장인들은 업무 보고서 같은 서류 작업에 활용하기 시작했고 학생들은 챗GPT가 써 준 과제를 그대로 제출하기도 합니다. 챗GPT는 소설 창작, 번역, 교정 등 단순한 챗봇의 기능을 넘어 다양한 언어 콘텐츠 제작 능력을 보여주고 있어요.[1]

사실 챗GPT의 영문명 'Chat Generative Pre-trained Transfomer'를 살펴보면 어떤 능력을 가지고 있는지 고스란히 드러납니다. 즉 대화를 통해(Chat) 의미 있는 텍스트나 코드 등을 생성한(Generative) 것을 전달하는데 이는 사전 학습(Pre-trained)을 통해 습득한 정보들을 바탕으로 하며 그 기반 기술은 구글의 인공지능 모델인 트랜스포머(Transfomer)라는 것이에요. 또한 챗GPT와 같은 자연어 생성형 AI는 머신러닝을 통해 아주 방대한 데이터를 학습하기 때문에 거대 언어 모델(Large Language Model : LLM)에 해당합니다.

사람들이 챗GPT에 열광하는 이유는 그 우수한 능력도 있지만 개발사인 오픈AI가 웹사이트 방문객들에게 누구나 무료로 사용할 수 있도록 열어 준 것도 있습니다. 한편 이런 챗GPT, 즉 GPT3.5의 대성공 이후 오픈AI는 2023년 3월 성능을 보다 강화한 유료 버전인 GPT-4를 출시하기도 했습니다.

AI 생성물의 저작권은 누구에게? [2]

그렇다면 이러한 AI가 생성한 결과물의 저작권은 누구에게 있을까요? 우리나라를 비롯한 대부분의 국가에서는 AI가 만들어 낸 것을 인간의 창작물로 보기 어렵다는 이유로 생성형 AI의 저작권을 인정하지 않고 있습니다. 저작권법상 저작물로서 보호받기 위해서는 인간의 창의성, 기술, 노력의 결과로써 만든 고유한 창작물이어야 하기 때문이에요. 게다가 AI는 어떤 결과물을 내놓을지 사용자가 정확하게 예측하기도 어렵다는 리스크가 있습니다.

더욱이 AI가 생성한 텍스트, 이미지 등의 결과물은 세상에 없던 것을 창조한 것이 아니라 기존에 누군가가 이미 만든 저작물을 학습한 것이라는 점에서, 자칫 원저작자의 저작권을 침해할 수 있어 주의가 필요해요. 생성형 AI의 원리상 저작권이 있는 저작물이 학습되는 것은 불가피하므로, 결국 알고리즘 학습에 기여한 자들의 권리에 대해 지속적인 사회적 합의가 진행되어야 할 것입니다.

---

1) 'THE AI REPORT : 공공분야 생성형 AI 활용 방안', NIA, 2023
2) 'THE AI REPORT : 공공분야 생성형 AI 활용 방안', NIA, 2023

1 생성형 AI를 활용해 새로운 콘텐츠를 만들었어요! 그럼 저는 저작권을 가질 수 없는 건가요?

NO!

생성형 AI가 자동적으로 생성한 음악이나 그림, 소설과 같은 결과물은 현행 저작권법의 보호 대상이 아니에요. 하지만 인간이 AI가 생성한 결과물에 창작적 표현을 추가했다면 그 기여 부분에 대해서는 저작권을 가질 수 있어요!

'생성형 AI'란 이용자의 요구에 맞추어 기존에 학습한 콘텐츠나 사물의 특징들을 결합해 이와 유사하게 보이는 결과물을 만들어 내는 인공지능을 의미합니다.

생성형 AI 등장 이전까지는 사람이 컴퓨터 프로그램을 창작의 도구로 활용하거나, 구체적인 지시·명령을 해서 창작 활동을 했어요. 그러나 최근의 생성형 AI는 사람의 관여가 거의 없는 상황에서도 알고리즘 학습을 통해 인간의 창작과 유사한 결과물을 만들어 내고 있어요.[3] 즉 생성형 AI가 만들어 낸 음악, 그림, 소설과 같은 콘텐츠에는 인간의 창작적 기여가 거의 필요하지 않게 되었죠.

그런데 현행 저작권법의 해석에 따르면 '인간의 창작물'만이 저작물이에요. '권리능력을 가진 자연인 또는 법인'만이 저작자로 인정될 수 있습니다(대법원 2011. 2. 10. 선고 2009도291 판결, 대법원 2014. 9. 4. 선고 2012다115625 판결).

저작권법 제2조

1. '저작물'은 인간의 사상 또는 감정을 표현한 창작물을 말한다.
2. '저작자'는 저작물을 창작한 자를 말한다.

특히 '창작성'은 '저작자 나름대로의 정신적 노력의 소산으로서의 특성이 부여되어 있어야 인정될 수 있으므로(대법원 1995. 11. 14. 선고 94도2238 판결)', 정신적 활동이 불가능한 AI에게 창작성이 있다는 것을 인정할 수 없습니다. 따라서 '인간의 사상 또는 감정을 창작적으로 표현한 것'이 아닌 생성형 AI가 만들어 낸 결과물은 저작권법의 보호 대상으로 볼 수 없고,[4] 생성형 AI를 이용해 결과물을 산출한 자를 저작권자로 인정하기도 어려울 거예요.

3) 김원오, 'AI 자동 생성작품에 대한 저작권 부여 한계와 대안적 보호방법론', 계간 저작권 142호(36권 2호), 2023
4) '인공지능 창작물', '인공지능 저작물'이라는 용어는 법적으로 정확한 용어가 아닐 수 있으므로 사용에 유의할 필요가 있음

그렇지만 생성형 AI가 만들어 낸 결과물에 이용자가 나름대로 창작적 표현 형식을 추가했다면 얘기가 달라져요. 그 창작적 기여 부분은 이용자의 독자적 저작물로 인정받을 수 있으니까요.

예를 들어 이용자가 생성형 AI에게 창작적 기여를 인정할 만한 구체적인 지시를 내린 경우 이용자를 저작권자로 인정해야 한다는 견해가 존재합니다.[5] 다만 이 경우에도 이용자의 창작적 기여가 저작물로 인정받기 위한 기준 또는 창작성의 정도에 대해서는 또 다른 논의가 요구될 것으로 보여요.

관련사례

▶ **인공지능이 발명자가 되어야 한다고 다툰 사례**

AI가 특허법상 발명자로 인정될 수 있는지 여부에 대해서 법원은 "기술적 사상이란 결국 인간의 사유를 전제로 하는 것이고, 창작 역시 인간의 정신적 활동을 전제로 하는 것이다. (중략) 발명자의 지위는 원칙적으로 권리능력이 전제가 되어야 한다. (중략) 인공지능은 법령상 자연인과 법인 모두에 포섭되지 않으므로 현행 법령상으로 인공지능에게 권리능력을 인정할 수도 없다"라고 판시함(서울행정법원 2023. 6. 30. 선고 2022구합89524 판결).

특허법상 권리 보호 대상이나 규정 형식은 저작권법과 유사하고, 법원의 위와 같은 해석 방법은 저작권법상 '창작'에도 그대로 적용될 수 있음. 따라서 생성형 AI를 '저작물을 창작한 자', 즉 저작자로 인정하기는 어려워 보임.

2 생성형 AI가 작성한 거라길래 재가공해도 괜찮을 것 같아서 다운받아 사용했어요. 그런데 콘텐츠 게시자가 저작권 위반이라고 항의를 해왔네요. 생성형 AI가 만든 콘텐츠의 재사용도 법적으로 문제가 되나요?

생성형 AI 작성 콘텐츠에는 현행 저작권법상 저작권이 인정되지 않기 때문에 콘텐츠 게시자가 저작권 위반을 주장하는 것은 타당하지 않아요. 그러나 생성형 AI가 학습한 원저작물에 대해서는 여전히 저작권 침해 문제가 제기될 수 있어요!

단순히 생성형 AI 결과만으로 만든 콘텐츠는 저작권법의 보호 대상이 아니기 때문에 생성형 AI 작성물을 그대로 게시한 사람이 저작권을 주장하는 것은 법적으로 타당하지 않아요.

5) 박성오, "[박성호의 지재 공방] AI 생성물의 저작물성 및 저작권 귀속 문제", 법률신문, 2023. 3. 16.

오히려 콘텐츠 게시자는 생성형 AI 작성 콘텐츠를 이용하기 전에 생성형 AI 제공 회사의 '이용 약관'을 확인할 필요가 있어요. 예를 들어 챗GPT 개발사인 오픈AI(Open AI)는 이용자들에게 산출물(Output)에 대한 모든 권리를 양도하고 상업적으로 판매, 이용할 수 있도록 하고 있는데,[6] 이 경우에도 AI 이용자가 챗GPT를 활용해 제작한 산출물의 저작권을 취득하는 것은 아닙니다. 즉 타인이 이를 재사용하는 것을 금지할 수 있는 권리(배타적 권리)까지 가진다고 보기는 어려워요.

한편 최근에는 생성형 AI의 결과물을 공공의 영역(Public Domain)에 남겨 두고, 누구나 자유롭게 이용하도록 해야 한다는 견해도 제시되고 있어요.[7]

물론 그렇다고 해서 생성형 AI 콘텐츠는 누구나 자유롭게 이용할 수 있다는 얘기는 아닙니다. 예를 들어 생성형 AI가 기존에 존재하는 저작물이나 캐릭터 등과 동일 또는 유사한 그림을 그려 냈다면, 과연 그 그림은 생성형 AI가 작성했다는 이유만으로 자유롭게 이용할 수 있을까요? 그렇지 않습니다.

생성형 AI는 필연적으로 학습 과정에서 기존에 존재하던 다른 사람의 저작물을 알고리즘을 통해 학습하게 돼요. 이에 따라 생성형 AI는 기성 작가의 화풍을 그대로 모방하기도 하고, 기존 저작물에 존재하는 특징을 새로운 콘텐츠에 그대로 드러내기도 합니다. 즉 생성형 AI는 학습한 내용을 무작위로 조합하기 때문에, 개발자나 이용자 입장에서도 생성형 AI의 결과물이 어떤 저작물을 참조했는지 정확히 알기 어려워요.

만약 생성형 AI가 작성한 콘텐츠에서 기존 저작물과 동일 또는 유사한 특징이 감지된다면, 이는 AI가 원저작물을 복제·전송 또는 2차적 저작물(원저작물을 번역·편곡·변형·각색·영상 제작 그 밖의 방법으로 작성한 창작물) 작성의 방법으로 이용한 것이기 때문에 원저작권자의 허락이 없는 한 저작권을 침해한 것으로 볼 수 있습니다.

즉 '콘텐츠 게시자의 저작권'이 인정되기 어렵더라도 원저작권자에 대한 저작권 침해는 문제될 수 있는 거죠. 따라서 생성형 AI 콘텐츠를 이용하는 것에도 주의가 필요합니다.

관련사례

▶ **[AI 시대의 저작권] 기존 화가 화풍 모방해 그린 AI 그림, 저작권 침해일까 아닐까(조선비즈 2023. 7. 4.)**
세계 최대의 사진·이미지 제공업체인 게티이미지(Getty Images)는 2023년 1월과 2월경 영국의 AI 개발사 스테빌리티AI(Stability AI Inc.)를 상대로, 스테빌리티AI가 개발한 이미지 생성 AI '스테이블 디퓨전(Stable Diffusion)'이 1,200만 장 이상의 사진을 복제하고 자신의 서비스와 경쟁하고 있다고 주장하면서 영국 런던 및 미국 델라웨어주 연방법원에 저작권 침해 소송을 제기함.
이외에도 화가, 코드 개발자 등 저작권자들이 AI 개발사를 상대로 소송을 제기하는 등 저작권 분쟁이 일어나고 있음.

6) https://openai.com/policies/terms-of-use
7) 김원오, 'AI 자동 생성작품에 대한 저작권 부여 한계와 대안적 보호방법론', 계간 저작권 142호(36권 2호), 2023

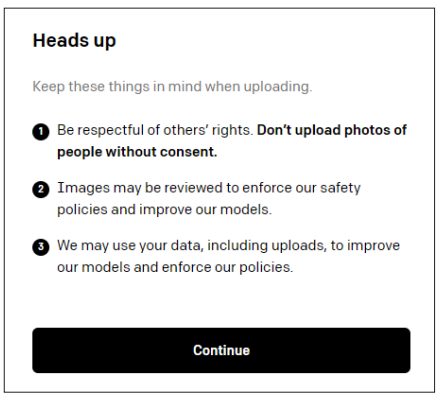
3 이미지(영상) 생성형 AI를 활용해서 연예인과 유명인의 얼굴이 나오는 콘텐츠를 만들었어요. 친구들과 함께 보려고 개인 블로그, SNS에 올릴까 하는데 아무래도 초상권이 문제되겠죠?

Yes! 생성형 AI가 제작한 결과물이라고 해도 연예인, 유명인처럼 공인의 초상임을 알 수 있는 콘텐츠를 무단으로 게시해서는 안 돼요. 공인뿐 아니라 국민 누구의 초상이라도 허락 없이 게시한다면 법적으로 문제가 될 수 있습니다!

현재 '초상권'을 직접적으로 규율하는 법은 존재하지 않습니다. 하지만 헌법 제10조 제1문에 따라 헌법적으로 보장되는 권리가 있어요. '사람은 누구나 자신의 얼굴 그 밖에 사회통념상 특정인임을 식별할 수 있는 신체적 특징에 관해 함부로 촬영되거나 그림으로 묘사되지 않고 공표되지 않으며 영리적으로 이용되지 않을 권리'를 가질 수 있습니다(대법원 2021. 4. 29. 선고 2020다227455 판결).

따라서 연예인이나 유명인뿐 아니라 국민이라면 누구나 그의 초상이 허락 없이 촬영, 묘사, 영리적으로 이용되는 것을 거부할 수 있어요. 무단 이용자에 대해서는 손해배상 청구 및 게시 중단(초상권 침해 금지) 청구를 할 수 있고요. 특히 법원은 초상권의 내용에 '그림으로 묘사되지 않을 권리'가 있다는 점을 명시하고 있으니, 이미지나 영상 생성형 AI를 통해 특정인의 초상을 드러내는 것은 초상권 침해에 해당할 수 있음을 명심해야 합니다.

한편 생성형 AI를 통해 특정인의 얼굴을 식별할 수 있을 정도의 결과물을 만들어 내기 위해서는 ① 프롬프트 창에 연예인이나 유명인의 이름을 직접 입력해 공표된 사진을 참조하도록 하거나 ② 생성형 AI가 참고할 만한 사진을 직접 업로드하는 방법이 있는 것으로 알려져 있습니다. 이 두 방법 중 어느 것을 사용하더라도 결국은 이용자가 해당 연예인 또는 유명인을 지목해야 한다는 점에서 초상의 주인을 '특정'한 것이라고 볼 수 있어요. 실제로 오픈AI의 이미지 생성형 AI인 달리2(DALL·E 2)에서 이 중 두 번째 방법으로 이미지를 생성하려고 하면 아래와 같은 경고창이 나옵니다.



출처 : ['다른 사람의 동의 없이 사진을 업로드하지 마세요'라는 문구를 포함한 달리2(DALL·E 2)의 경고창(https://labs.openai.com/)]

이처럼 생성형 AI에 다른 사람의 초상이 담긴 사진을 참조하라는 명령을 입력하거나, 직접 업로드하는 방법으로 콘텐츠를 작성하고자 할 경우, 그 당사자의 동의를 얻는 것이 필요해요.

이제 이미지나 영상이 포함된 생성형 AI 콘텐츠는 점점 실제와 구분하기 어려워지고 있죠. 가짜 뉴스 또는 음란물, 명예 훼손 표현에 이용되는 사례도 많아지고 있고요. 하지만 개인 블로그 또는 SNS라 하더라도 이러한 불법 정보를 게시하는 경우 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 등에 따라 형사처벌까지 될 수 있다는 점(제74조, 제44조의7)을 유의해야 합니다.

관련사례

▶ 인격표지영리권 도입 가능성

연예인이나 유명인이 자신의 성명, 초상, 음성 등 인격적인 권리를 '재산'처럼 활용해 경제적 이익을 창출할 수 있도록 하는 권리를 영미법상 '퍼블리시티권'이라고 하는데, 최근 우리 법제에도 퍼블리시티권 도입 논의가 이루어지고 있음.

일례로 ① 2021년 12월 7일 개정되어 2022년 6월 8일부터 시행 중인 부정경쟁방지법 제2조 제1호 타목은 "국내에 널리 인식되고 경제적 가치를 가지는 타인의 식별표지를 자신의 영업을 위해 무단으로 사용하는 행위"를 부정경쟁행위로 규정했는데, 법원은 이 권리가 곧 퍼블리시티권을 의미하는 것은 아니라고 판시한 바 있음(서울중앙지방법원 2022. 7. 8. 선고 2021나38453 판결).

② 한편 정부는 2022년 12월 27일경 '인격표지영리권' 도입을 위한 민법 전부개정법률안을 입법 예고했음. 개정안에 따르면 유명인이 아니더라도 모든 국민은 자신의 인격을 영리적으로 이용할 수 있고, 재산권처럼 상속의 대상이 될 수 있음.

4 음성 생성형 AI를 활용해 인기 있는 가수의 음성으로 AI 커버곡을 만들었어요. 온라인에 공유하면 문제가 될까요?

Yes!

물론입니다. 커버곡의 원저작물에는 원저작권자의 저작권이 존재하고, AI로 재생된 가수의 음성 역시 별도의 음성권으로 보호돼요. 따라서 이와 관련된 모든 권리자에게 동의를 받은 후 이용해야 해요!

흔히 말하는 '커버곡'이란 이미 공표된 음악 저작물을 다시 연주·가창하는 것을 의미합니다. 그런데 음악 저작물에는 작사·작곡가의 복제권, 전송권, 2차적저작물작성권을 비롯해 가수와 연주자, 앨범 제작자의 저작인접권이 존재하죠. 따라서 커버곡을 만드는 경우에도 개별적인 권리자들에게 모두 동의를 받고 이용해야 해요.

최근 음성 생성형 AI를 통해 공개되는 음악 저작물들은 이미 많은 인기를 얻은 대중가요가 대다수인 것으로 보이는데요, AI가 새롭게 가창을 하더라도 원저작물에 대한 작사·작곡·편곡자의 권리는 여전히 그대로 존재합니다. 그러니 원곡의 가수만 변경해 그대로 이용하면 저작권자에 대한 복제권 침해가, 편곡해 이용하면 2차적저작물작성권 침해가 문제될 수 있고, 그 결과물을 온라인에 공유하면 전송권까지 침해될 소지가 있어요.

음성 생성형 AI의 또 다른 문제점은 '인기 있는 가수'의 목소리를 복제해 동의 없이 이용한다는 데 있습니다. 생성형 AI가 창작한 커버곡 영상을 보면 실제로 존재하는 유명 가수의 음색이나 창법까지 그대로 느껴지는 경우가 대부분이니까요.

사람은 누구나 자신의 음성이 함부로 녹음되거나 재생, 방송, 복제, 배포되지 않을 권리를 가지는데, 이러한 음성권은 헌법 제10조 제1문에 의해 법적으로 보장되고 있는 권리이므로, 음성권에 대한 부당한 침해는 불법행위에 해당해요(서울중앙지방법원 2019. 7. 10. 선고 2018나68478 판결).

아직까지 실존 인물의 목소리를 AI로 재현하기 위해서는 가수의 목소리를 샘플로 참조하고 학습시켜야 할 텐데, AI 결과물을 생성하는 과정에서 가수의 음성을 '복제'하고 있다는 사실은 충분히 인정될 수 있을 것으로 보입니다. 따라서 유명 가수의 동의 없이 그의 음성으로 AI 커버곡을 작성해 온라인에 공유하는 경우 민사상 불법행위 책임을 부담하게 될 수 있음을 명심해야 해요.

관련사례

▶ **뉴진스 '하입 보이' 브루노 마스가 부르면… 일반인도 AI 커버송(IT조선 2023. 5. 17.)**

AI를 통해 만들어진 커버송이 최근 주목을 받고 있음. 피프티피프티(FIFTY FIFTY)의 '큐피드'를 마이클 잭슨이 부르거나, 뉴진스의 '하입 보이(Hype Boy)'를 브루노 마스가 가창하기도 함.

이와 관련 한국음악저작권협회 관계자는 "AI 커버송이 등장한 지 얼마 되지 않아 아직 해당 이슈가 구체적으로 어떤 문제가 있는지 정리가 안 된 상태", "원저작권자가 어떻게 생각하는지 중요"하다고 말함.

5 내가 만들어서 온라인에 공유한 콘텐츠(글, 이미지, 음원 등)를 생성형 AI가 무단으로 학습해 활용하고 있음을 발견했어요. 내 저작권을 지킬 수 있을까요?

내 콘텐츠의 저작권을 생성형 AI가 무단으로 복제해 이용한 것을 발견했다면, AI 개발·서비스 회사에 학습 데이터에서 해당 콘텐츠를 즉시 제외하고 저작권 침해 행위를 중단하도록 요청할 수 있어요!

생성형 AI는 생성물의 형태에 따라 텍스트, 그림, 음성, 영상 등으로 구분되지만 그 작동 방식은 거의 유사합니다. AI 학습 중 대부분을 차지하는 '텍스트 데이터'는 대개 웹 크롤링(Web Crawling) 방식이나 웹 스크래핑(Web Scraping) 방식으로 수집되는 것으로 보입니다.

웹 크롤링 : 인터넷에서 웹페이지를 자동으로 탐색하고 수집하는 과정을 의미하며, 여러 사이트의 주소에 접근하고 관련된 링크를 찾아 수집한 다음 색인(Index)을 만들어 데이터베이스에 저장하는 작업. 검색 엔진에 주로 사용됨.

웹 스크래핑 : 웹페이지에서 필요한 정보를 추출하는 과정을 의미하며, HTML 소스 코드를 읽어 필요한 데이터를 원하는 형식으로 데이터베이스에 저장하는 작업. 다양한 목적으로 사용됨.

이와 같은 수집 방법에 따라 자동으로 수집된 웹페이지 내에 콘텐츠가 포함되어 있다면 AI는 해당 콘텐츠를 학습해 새로운 창작에 이용하게 되는데, 이러한 데이터의 수집 이용은 현행 법령상 적법하다고 보기 어려워요. 따라서 AI가 허락 없이 무단으로 콘텐츠를 이용하고 있는 것을 발견했다면 원저작물의 복제권·전송권·2차적저작물작성권 등 저작권법상 권리가 침해되었음을 주장할 수 있어요.

우선 AI 개발·서비스 회사에 대해 저작권 침해 금지를 요청할 수 있고(저작권법 제123조 제1, 2항), 자신의 저작물을 학습 데이터에서 제외함으로써 추가적인 저작권침해행위를 예방해 줄 것을 청구할 수도 있어요.

그런데 이 경우 AI 개발·서비스 회사가 저작자의 요구를 무조건 받아들이지 않을 수는 있습니다. 현행 저작권법상 AI가 기존 저작물을 완전히 동일하게 이용하는 것이 아니라면 '공정한 이용'(저작권법 제35조의5)에 해당한다고 판단할 수 있거든요. 즉 데이터 마이닝(Data Mining : 수집된 대용량 데이터의 패턴을 분석하고 이에 따른 유용한 정보를 찾아내는 것)을 위해 저작물을 이용하는 것은 적법하다는 유력한 견해가 존재합니다. 더욱이 AI 시대에 데이터 마이닝 기술을 보다 적극적으로 허용해야 할 필요성이 있다는 견해가 많은 지지를 받고 있고요.

실제로 국회에 발의되어 있는 저작권법 전부개정안(도종환 의원 대표 발의, 의안번호 7440)에는

AI와 빅데이터 기술의 발전으로 인해 저작물이 포함된 대량의 정보를 활용할 필요성이 있으나 현행 저작권법상 '공정이용 조항'(제35조의5)만으로는 이용이 어렵다는 점을 지적하면서, '데이터 마이닝' 과정에서 저작물을 이용하는 경우 면책되는 규정(안 제43조)을 신설한다는 내용이 포함되어 있습니다.

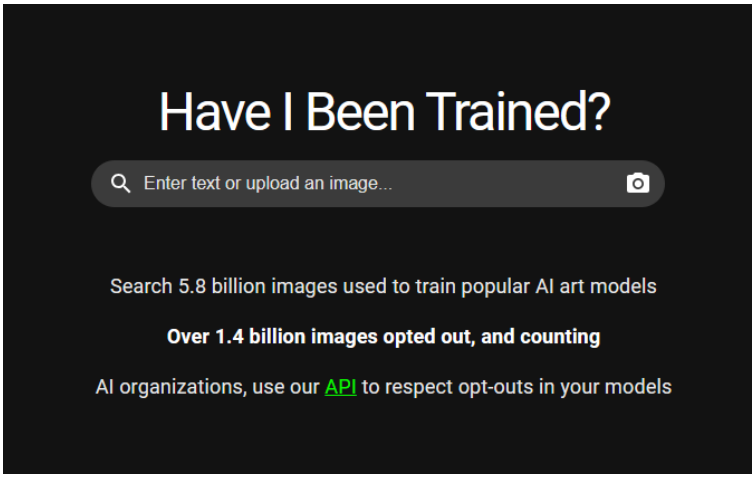
또한 국회는 최근 저작물이 포함된 정보를 데이터 마이닝 기법 등을 활용해 분석·활용하는 경우 저작권법상 공정이용 조항이 적용될 수 있는지를 재차 검토하면서, '정보분석'에 저작물을 자유롭게 이용할 수 있도록 하는 조항을 신설하는 안을 발의하기도 했어요.

요약해 보면 ① 내 콘텐츠를 AI가 무단으로 사용하고 있다면 현행 저작권법에 의해 저작권 침해에 해당함을 주장할 수 있습니다. 다만 AI 개발사가 데이터 마이닝과 같은 방법으로 꼭 필요한 한도에서 공정하게 이용했다고 항변할 가능성은 있어요. ② 또한 앞으로 AI가 콘텐츠를 수집해 학습하는 행위를 적극적으로 허용하도록 저작권법을 개정하자는 논의도 진행 중입니다.

관련사례

▶ 원작 이용에 대한 동의 여부는 아티스트 본인이 결정하도록

AI 관련 기업인 스포닝(Spawning, Inc.)은 'Have I Been Trained?'라는 사이트(https://haveibeentrained.com/)를 운영하면서, AI 이미지 생성에 사용되는 약 58억 개의 그림을 분석해 자신의 작품과의 유사성을 분석해 주고 있음.



그리고 각 아티스트가 자신의 데이터 사용 방식을 스스로 결정할 수 있도록 원작 이용에 대한 동의 여부를 결정할 권한을 주어야 한다고 주장함.

6 생성형 AI에 별도로 출처 표기가 없더라고요. 상업적 이용은 아니고 개인 SNS에 올릴 건데, 그래도 출처를 표기하지 않으면 문제되나요?

생성형 AI 결과물을 이용할 때 출처를 표기해야 한다는 법령은 아직까지 존재하지 않아요. 그렇지만 저작권 문제를 겪지 않으려면 생성형 AI에 이용된 원저작물이 무엇인지 표기해 주는 것이 바람직해요. 콘텐츠를 재사용하는 사람들이 정보의 정확성을 검증할 수 있도록 해당 콘텐츠가 생성형 AI로 작성되었음을 알려주는 것도 필요하고요!

생성형 AI 결과물 이용 시 출처를 표기해야 한다는 법령은 아직까지 존재하지 않습니다. 다만 현행 저작권법에 따르면 저작물을 이용하는 자는 저작자의 실명 또는 이명을 표시해야 하고(저작권법 제12조 제1항), '공정 이용' 조항처럼 저작권자의 저작재산권이 제한되는 경우에도 반드시 출처를 명시해야 합니다(저작권법 제37조 제1항).

이때 개인적으로 이용하는 경우와 상업적으로 이용하는 경우에는 큰 차이가 없어요. 특히 상업적으로 이용하는 경우에는 '해당 콘텐츠가 AI로 만들었음'을 소비자들이 알 수 있도록 고지하는 것이 바람직합니다. 자칫 사기죄의 기망 행위가 될 수 있기 때문이죠. 요즘 AI로 만든 콘텐츠들은 매우 정교해 소비자는 해당 콘텐츠가 AI가 아닌 사람이 만든 콘텐츠라고 생각하고 구매할 수도 있으니까요.

또한 생성형 AI가 참고한 원저작물이 있고, 생성형 AI가 작성한 결과물에도 원저작물이 그대로 감지된다면, 원저작물을 직접 이용하는 것에 준해 원저작물의 저작자 혹은 원출처를 표시해 주는 것이 바람직합니다. 예를 들어 'OO작가의 화풍으로 A라는 상황을 그려줘'라고 이미지 생성 AI 프롬프트에 입력했다면, 결과물을 이용하면서 'OO작가 또는 OO작가의 화풍을 참조한 것임'을 표시해 주면 되니 결코 어렵지 않아요.

한편 중앙대학교는 생성형 AI 가이드라인을 제정해 생성형 AI를 이용한 자료의 출처를 표시해야 하는 경우 출처 표시 방법을 상세히 안내하고 있습니다. 텍스트 또는 이미지 생성형 AI를 활용한 경우 ① 어떤 프로그램을 사용했는지, ② 언제 사용했는지, ③ 프롬프트 입력 내용이 무엇인지와 ④ 생성형 AI를 이용해 작성했다는 취지의 출처를 표시하고 ⑤ URL을 기재하도록 되어 있어요.

[중앙대학교 생성형 AI 활용 가이드라인]
텍스트 생성형 AI를 활용한 경우(예시) : ChatGPT3.5(2023. 04. 20). “프롬프트 내용.” OpenAI의 ChatGPT3.5를 이용해 생성 또는 작성함.
이미지 생성형 AI를 활용한 경우(예시) : Stable Diffusion (2023. 04. 20). “프롬프트 내용.” Stable Diffusion 온라인을 이용해 생성 또는 작성함.

이처럼 생성형 AI의 출처를 표기해 주는 것은 법률상 의무라고 할 수는 없지만, 콘텐츠를 접하는 사람들이 생성형 AI로 작성된 자료임을 확인할 수 있다는 점에서 중요합니다.

관련사례

▶ **생성형 AI가 연 판도라 상자… "가짜가 세상을 흔든다"(매일경제 2023. 6. 18.)**
유럽연합(EU) 의회는 2023년 6월 14일 “생성형 AI로 제작된 콘텐츠의 출처를 명확히 표기해야 하고, 해당 콘텐츠 제작 과정에서 AI가 사용되었다는 사실을 명시해야 한다”라는 내용이 포함된 인공지능 규제 법안(Artificial Intelligence Act)을 가결함.

이 법안은 유럽연합 각 정부의 동의를 받아야 시행될 수 있으므로, 2023년 연말 또는 그 이후에 시행될 것으로 보임.

7 유료로 웹툰이나 웹소설을 결제해서 보다가 생성형 AI로 만든 것처럼 보이는 콘텐츠를 발견했어요. 웹툰, 웹소설 플랫폼에 문의해서 환불받을 수 있을까요?

NO!

안타깝게도 쉽지 않을 거예요. 웹툰, 웹소설 플랫폼에서 AI 제작 콘텐츠에 대한 환불 규정을 별도 마련하지 않은 이상, 생성형 AI가 제작했음이 의심스럽다는 이유로 환불을 받는 것은 어려워요.

최근 웹툰 창작 과정에서 AI 기술을 이용한 것으로 논란이 발생하면서 팬들이 환불을 요청하고 AI 기술 보이콧 선언까지 한 사례가 있습니다. 각 플랫폼 역시 AI 기술을 배제하겠다는 움직임을 보이고 있죠.

하지만 생성형 AI가 제작한 콘텐츠의 경우 현재 각 웹툰 및 웹소설 플랫폼에서 별도 환불 사유로 규정하고 있지 않은 한 환불 대상이 된다고 보기 어렵습니다. 생성형 AI가 제작한 콘텐츠가 웹툰, 웹소설 창작에 대한 기본적인 창작 윤리를 위반했다는 점을 제외하고 저작권법 등 관련 법령을 위반했다고 판단하기에는 법적 근거가 미약하기 때문이에요.

현재 각 웹툰, 웹소설 플랫폼의 경우 이용자들과 체결한 계약에 해당하는 이용 약관상 마련한 환불 사유에 생성형 AI 제작 콘텐츠에 대한 환불 사유를 마련하고 있지는 않은 것으로 알려집니다. 이에 비추어 볼 때 유료 콘텐츠 이용 과정에서 생성형 AI가 제작한 것으로 의심이 된다는 이유로 환불을 받는 것은 쉽지 않을 것으로 판단돼요.

이용자가 플랫폼에서 유료 결제 후 환불 등의 조치를 받기 위해서는 플랫폼과 체결한 계약인 이용 약관에 따르는 것이 원칙입니다. 그런데 대부분의 플랫폼은 이용 약관상 ① 유료 결제 후 미사용분에 대한 환불, ② 과오금 발생분에 대한 환불 등에 대해서는 관련 규정을 마련하고 있으나 이미 콘텐츠 이용을 마친 부분에 대해서는 환불의 대상이 아님을 밝히고 있어요.

그러므로 유료 결제 후 콘텐츠를 보거나 읽는 과정에서 해당 콘텐츠가 생성형 AI의 도움을 받아 제작되었다는 의심이 든다 하더라도, 해당 콘텐츠에 대해서는 유료 결제를 통해 콘텐츠 소비가 이루어진 점, 그리고 플랫폼이 생성형 AI 제작 논란 작품에 대한 환불 사유를 마련하고 있지 않은 점에 따라 이용자는 환불을 받기가 어려울 것으로 보입니다.

물론 플랫폼 업체가 논란에 따른 책임 부담 차원에서 자발적으로 환불 등의 조치를 취할 수는 있을 거예요. 다만 이는 플랫폼 업체의 의무에 해당한다고 볼 수는 없습니다.

하지만 다른 측면에서 접근은 가능합니다. 대부분 플랫폼들은 콘텐츠의 하자 등 유료 이용 서비스와 관련되어 이용자의 피해가 발생한 경우 이용자의 피해를 배상하도록 하는 내용들을 마련하고 있어요. 그러니 생성형 AI가 제작한 콘텐츠가 비정상적인 콘텐츠에 해당해 생성형 AI가 제작한 부분을 일종의 '하자'로 볼 수 있다면 유료 결제 내역에 해당하는 부분을 이용자의 피해로서 배상받을 수 있을 거예요.

물론 이때도 생성형 AI가 제작한 부분이 콘텐츠의 '하자'에 해당하는지 여부에 대해서는 더 많은 논의와 검토가 필요할 것으로 보입니다. '하자'로 인정받기 위해서는 생성형 AI의 창작물에 대한 법적인 논의가 아직은 더 진행되어야 할 거예요.

관련사례

▶ **"저작권 회피 꼼수" AI 활용 웹툰 작가에 반발 '별점 테러'(중앙일보 2023. 8. 5.)**

2023년 5월 네이버 웹툰에 연재되는 웹툰작가가 AI를 작화 일부에 활용했다는 것이 알려짐. 작가 측은 후보정 작업 과정에서만 AI 기술을 이용했다고 해명했으나, 독자들은 AI 기술을 작품 제작에 활용한 것에 대한 거부감을 표시하며 해당 작품의 평가를 낮게 주는 '별점 테러'를 진행함.

카카오페이지에 연재되는 웹소설 역시 생성형 AI를 이용해 표지 디자인을 제작한 것이 알려지며 일러스트 작가들의 항의가 빗발친 사례가 있음.
이처럼 창작 과정에서 AI 기술을 이용하는 행위에 대해서는 이용자의 반발감이 크며, 나아가 AI 기술을 활용해 창작한 작품의 가치를 얼마나 인정할지 여부에 대한 논의는 아직 더딘 상황임.

생성형

윤리 가이드북

책임성

03

03 책임성

생성형 AI가 만든 작품이 미술대회 1등? [8]

2022년 미국 콜로라도주의 한 미술 대회에서 '스페이스 오페라 극장(Théâtre D'opéra Spatial)'이란 작품이 우승을 차지했습니다. 그런데 수상자인 제이슨 M. 앨런은 수상의 영예를 누리는 대신 오히려 비난에 휩싸이고 말았어요.

그 이유는 앨런이 출품한 작품이 AI가 만든 것임이 밝혀졌기 때문이었습니다. 앨런은 텍스트를 입력하면 그 내용을 그래픽으로 바꾸어 주는 이미지 생성형 AI 프로그램 '미드저니'를 사용했는데 이 사실을 모르는 심사위원들이 그에게 1등상을 주었죠. 이에 다른 예술가들은 앨런이 부정 행위를 했다며 크게 반발했어요.
하지만 앨런은 자신은 작품을 출품하면서 미드저니 사용 사실을 미리 고지했다며 아무런 문제가 없다고 항변했습니다. 어떠한 대회 규정도 어기지 않았다고요. 결국 이 사례는 생성형 AI에 대한 법 · 제도 미비와 사용자들의 책임의식 미정립 탓에 발생한 헤프닝이라고 할 수 있습니다.

AI가 다 했으니 사람은 책임 없다?

디지털 콘텐츠를 제작하거나 활용할 때 윤리적 중립을 지켜야 한다는 사실은 이제 모두가 알고 있습니다. 이러한 윤리적 중립성이나 책임성은 인간과 대화를 통해, 즉 인간의 지시어 입력을 통해 결과물을 만들어 내는 생성형 AI에도 마찬가지로 적용되어야 해요. 이에 전문가들은 성(性), 인종, 직업, 연령 등에 대한 차별이나 혐오 표현이 담긴 내용, 정치적 민감도가 높은 질문에는 답변을 할 수 없도록 설계되어야 한다고 조언하고 있습니다.[9]

특히 생성형 AI는 인간이 만든 데이터를 학습해 그것을 토대로 결과물을 만드는 것이므로 원 데이터에 편향성이 있을 경우 치우친 결과를 도출할 우려가 큽니다. 이는 차별을 조장하거나 불평등을 키우는 등 사회적 혼란을 유발하는 부작용을 낳을 거예요. 뿐만 아니라 사실이 아닌 것을 사실인 양 그럴 듯하게 답변하는 '환각(Hallucination)'도 큰 문제가 될 수 있습니다.
따라서 생성형 AI 이용자는 AI라고 해도 완벽할 수 없다는 사실을 늘 주지하고 스스로 책임의식을 가져야 해요. 자신이 활용하게 될 정보가 어디서 왔는지, 오류는 없는지, 편향은 없는지 먼저 고민해 보는 것이 필요할 듯합니다.

---

8) 한정희, 'AI가 그린 그림, 미술전 1등?…논란이 되고 있는 작품 스페이스 오페라 극장', AI라이프경제, 2022.9.5.
9) 'THE AI REPORT : 공공분야 생성형 AI 활용 방안', NIA, 2023

1 과제로 보고서를 쓰는데 시간도 절약되고 쉬울 것 같아서 생성형 AI를 통해 작성하려고요. 요즘 많이들 그렇게 하던데 그대로 제출해도 문제가 없겠죠?

생성형 AI를 이용해 나온 결과물을 그대로 과제로 제출하면 학습자 본인의 학습력과 사고력이 저하될 뿐만 아니라, 기망으로 인한 교수자의 수업권 침해 우려도 있어요. 생성형 AI는 과제의 보조적인 도구로만 활용하고 최종적인 보고서의 완성은 학습자 본인이 직접 하는 것이 바람직해요!

생성형 AI를 과제나 숙제의 보조 도구로 활용하는 것은 문제가 없습니다. 하지만 생성형 AI를 이용해 나온 결과물을 그대로 과제로 제출하면 다음과 같은 문제들이 발생할 수 있어요.

첫째, 학습자 본인의 학습력과 사고력이 저하됩니다. 스스로 공부하고 학습해 과제를 수행하지 않았으므로 본래 교육 목적에 맞는 학습을 한 것이 아니며, 결국 학습자 본인의 학습력과 사고력이 저하되는 결과를 초래할 거예요.

둘째, 교수자의 업무와 수업권을 방해할 수 있습니다. 과제는 교사(교수)가 학생에게 교육 목적에 맞게 학습을 위해 제시하고 평가하는 수업의 일종인데, 학습자가 직접 과제를 수행한 것이 아님에도 마치 자신이 수행한 것처럼 기망한다면 교수자의 업무와 수업권을 방해하는 셈이에요.

따라서 생성형 AI를 이용한 결과물을 그대로 과제로 제출해서는 안 됩니다. 생성형 AI를 활용해 과제의 아이디어를 얻고, 보고서 개요와 초안을 만들거나 번역, 사례 수집 등 보조적으로만 활용해야 해요.

물론 이 경우에도 최종 과제 보고서의 완성은 학습자 본인이 직접 해야 합니다. 또한 보고서에 출처를 달아 생성형 AI를 어떻게 활용했는지를 표기해 교수자가 알 수 있도록 하는 것이 좋아요.

한편 최근 각 대학에서는 생성형 AI 활용 가이드라인을 제정해 학생들이 참고할 수 있도록 제공하고 있습니다.

대학교	가이드라인 제목	확인 링크
국민대학교	국민 인공지능 윤리강령	https://www.kookmin.ac.kr/comm/menu/user/89bdc36e40697dd8ad9e6cb0d423bc9c/content/index.do
세종대학교	생성형 AI 교수 학습 기본 활용 가이드 라인	http://board.sejong.ac.kr/boardview.do?pkid=160936&currentPage=1&searchField=ALL&siteGubun=19&menuGubun=1&bbsConfigFK=335&searchLowItem=ALL&searchValue
고려대학교	ChatGPT 등 AI의 기본 활용 가이드라인	https://int.korea.edu/kuis/community/notice_dis.do?mode=download&articleNo=324500&attachNo=228291&totalNoticeYn=N&totalBoardNo=
중앙대학교	생성형 AI 활용 가이드라인	https://www.cau.ac.kr/cms/FR_CON/index.do?MENU_ID=2730

관련사례

▶ **연세대 '챗GPT 대필 의심' 과제 0점 처리… "작문 과제, 필기로 전환"(동아일보 2023. 3. 29.)**

연세대 교양과목 작문 수업에서 담당 교수가 챗GPT 대필이 의심된다는 이유로 수강생의 작문 과제를 '0점' 처리한 사례가 있음. 학교 관계자는 반발이 예상되긴 하지만 전체 2단락 중 1단락이 챗GPT 답변 내용과 일치하는 등 표절 정황이 명백해 0점 처리했다고 밝힘.

또 중앙대 사회과학대의 한 수업에서는 개강 첫날 학생들을 대상으로 온라인 표절 관련 교육을 진행한 후 "챗GPT를 활용해 표절하지 않겠다"는 서약서를 받기도 함.

성균관대 자연과학대의 한 수업에서도 "공부하는 중에는 챗GPT를 활용할 수 있지만 이를 활용해 도출한 답을 자신이 쓴 것처럼 제출하면 부정행위로 간주해 0점 처리하겠다"라고 공지함.

2 학교 오픈북 시험에서 인터넷 활용도 가능하다고 교수님이 말씀하셨어요. 생성형 AI에서 얻은 답을 그대로 적어서 제출해도 될까요?

교수님이 허용한 경우가 아니라면, 오픈북 시험이라고 해도 생성형 AI에서 얻은 답을 그대로 답안으로 제출하는 것은 바람직하지 않아요. 생성형 AI는 답안 도출의 보조적인 도구로만 활용하고 최종적인 답안 작성은 학습자 본인이 직접 하는 것이 좋아요!

교수자가 미리 생성형 AI의 활용에 대한 가이드라인을 제시했다면 이를 준수하고, 그렇지 않은 경우에는 사전에 교수자에게 생성형 AI를 어디까지 활용할 수 있는지 질문하는 것이 좋습니다. 하지만 이 두 경우가 아니라면, 오픈북 시험이라고 해도 생성형 AI에서 얻은 답을 그대로 답안으로 제출하는 것은 바람직하지 않아요.

학습자는 생성형 AI를 시험 답안을 도출하는 데 도구로 활용하고 최종 시험 답안은 학습자 본인이 직접 완성해 제출하는 것이 좋습니다. 또한 출처를 달아 생성형 AI를 답안 작성에 어떻게 활용했는지를 교수자가 알 수 있도록 하는 것이 바람직해요.

3 음악, 포스터, 수필 등 공모전이나 경진대회에 참여할 때 생성형 AI의 도움을 받으려고요. 이렇게 만든 작품을 제출해도 될까요?

일반적으로 공모전이나 경진대회는 참가자가 직접 창작한 작품을 평가하는 것이니만큼, 생성형 AI를 이용해서 만든 작품을 제출하면 안 돼요. 자칫 부정행위가 될 수 있어요!

대회 주최 측에서 미리 생성형 AI를 활용할 수 있는지 가이드라인을 제시했다면 이를 준수하고, 그렇지 않은 경우에는 대회 주최 측에 생성형 AI를 활용할 수 있는지 문의하는 것이 좋습니다.

이 두 경우가 아니고 임의로 생성형 AI를 활용한 작품을 제출하는 것은 문제가 될 수 있어요. 일반적으로 공모전이나 경진대회는 참가자가 직접 만든 작품으로 대회가 진행되는 것이므로, 생성형 AI를 이용해서 만든 작품을 제출하면 부정행위와 대회 업무를 방해하는 행위에 해당될 우려가 있습니다.

관련사례

▶ 美 미술 공모전서 AI가 그린 그림이 1등상 받아 논란 "예술의 죽음" vs "AI도 사람이 작동"(동아일보 2022. 9. 5.)

미국의 한 미술 공모전에서 AI 프로그램으로 제작한 그림이 1등상을 받아 논란이 됨. 수상자인 앨런은 수상작 '스페이스 오페라 극장'이 "고전적 여자가 우주 헬멧을 쓴 모습에서 출발, 꿈에서 나올 법한 분위기를 연출하기 위해 약 80시간 동안 실험하는 복잡한 과정을 거쳤다"라고 설명하면서, 출품 당시 "(이미지 생성형 AI인) 미드저니를 활용했다"라고 밝힘. 하지만 일부 심사위원은 해당 그림이 AI로 제작된 것일지 몰랐을 정도로 정교해 충격을 줌.

이에 대해 트위터를 비롯한 소셜미디어에서는 "로봇이 올림픽에 출전한 격"이라거나 "람보르기니를 타고 마라톤에 참가한 것", "클릭 몇 번으로 만든 디지털 아트", "상을 반납하고 공개 사과하라"라는 비난이 일기도 했음.

4 회사 내부에서 기획안, 사업계획서, 보고서 등 업무 자료를 만들 때 생성형 AI를 활용했어요. AI가 제시한 정보에 대해 제가 책임질 일은 없겠죠?

NO!

생성형 AI를 활용하면 업무 자료를 손쉽고 빠르게 만들 수 있지만, 자료의 진위 여부와 정보 유출에는 반드시 유의해야 해요. 따라서 최종적인 업무 자료의 작성과 완성은 작성자 본인이 직접하는 것이 바람직해요!

생성형 AI를 활용해 업무 자료를 만들 때는 아래 3가지 사항을 주의해야 합니다.

❶ 자료의 진위 여부(환각 현상, Hallucination)

생성형 AI는 거짓 정보를 사실인 것처럼 제공하는 환각 현상이 있으므로, 업무 자료로 추천해 준 통계나 데이터가 거짓이거나 제시된 내용에 오류가 있을 수 있습니다. 때문에 정보의 진위 여부에 대해서는 최종적으로 사실 확인(팩트 체크)을 꼭 해야 해요.

❷ 회사 기밀 유출 위험

생성형 AI에 입력되는 내용들은 외부 서버에 저장되고 AI의 학습에 재이용되기도 하는 등 외부 유출의 위험이 있습니다. 따라서 회사에서 업무 자료를 만들 때 회사 기밀이나 개인정보 등 민감한 정보는 생성형 AI에 입력하지 말아야 해요.

❸ 업무 자료의 책임 소재

생성형 AI로 만든 자료를 업무 문서로 그대로 제출 및 활용하는 것은 책임 소재 측면에서 바람직하지 않습니다. 생성형 AI는 업무 자료 작성에 보조적으로 활용하고, 최종 자료의 작성과 완성은 작성자 본인이 직접 하는 것이 좋아요.

업무 자료의 최종 책임자는 작성자 본인입니다. 생성형 AI를 활용해 기획안, 사업계획서, 보고서 등을 손쉽고 빠르게 만들 수는 있겠지만, 문제는 아직 생성형 AI가 완벽하지 않다는 데 있어요. 그러니 자료의 진위 여부와 정보 유출에 유의해 활용하고, 최종적인 문서의 작성과 완성은 작성자 본인이 직접 하는 것이 바람직해요.

관련사례

▶ **판결문 작성에 챗GPT 사용한 콜롬비아 판사 논란(AI타임스 2023. 2. 3.)**

미국 콜롬비아의 한 판사가 판결문 작성에 챗GPT를 활용했다고 밝혀 논란이 됨. 판결문 작성에 챗GPT를 활용한 파디야 판사는 “챗봇을 텍스트 초안 작성에 용이하게 사용할 수 있을 것”이라며 “애플리케이션에 질문을 한다고 해서 판사 자격이 없어지거나 생각하는 존재가 아니게 되는 것은 아니다”라고 주장함.

반면 콜롬비아 로사리오대학의 후안 다비드 구티에레스 교수는 “챗GPT에 동일한 질문을 했을 때 다른 대답을 받았다”라며 “해당 판결에서 판사가 챗GPT를 활용한 것은 확실히 비윤리적이며 무책임한 행동”이라고 비판함.

5 생성형 AI와 대화를 나누면서 회사 정보를 올린 적이 있는데, 어쩌다 기밀 정보까지 노출이 되었어요. 이 경우 제가 법적 책임을 지게 되나요?

생성형 AI에 회사 정보를 올리고 대화를 나누면 회사 기밀 정보가 유출될 수 있고 법에 따라 책임을 져야 할 수 있으니 각별한 주의가 필요해요!

생성형 AI와 대화할 때 회사 정보를 올리면 자칫 기밀 정보가 유출될 수 있고, 경우에 따라 형사적, 민사적 책임을 져야 할 수 있습니다.

이와 관련 블룸버그 통신은 2023년 4월 이스라엘 보안회사 팀8(Team 8)이 생성형 AI와 관련해 발간한 보고서에 대해 보도한 바 있습니다.[10] 보고서는 기업에서 직원이 업무 처리를 위해 생성형 AI에 기업 정보를 제공하는 형식으로 이용하면 회사 기밀과 고객 개인정보가 외부에 유출될 수 있으니 조심해야 한다는 내용이었어요.

또 기업에서 생성형 AI를 이용하는 과정에서 직원이 입력하는 회사 기밀과 고객 개인정보는 추후 삭제하는 것이 매우 어려우며, API를 통해 회사와 고객의 정보, 지식 재산, 소스 코드 등이 유출될 수 있어, 적절한 보호 조치와 위험 관리가 필요하다고 권고했어요.

실제로 챗GPT를 운용하는 오픈AI는 "당신과의 대화는 우리 시스템 향상을 위해 AI 교육자에 의해 사용될 수 있다"라고 안내를 하기도 합니다. 이와 같은 이유로 삼성전자, SK하이닉스 등 국내 기업과 JP모건체이스, 골드만삭스 등 해외 기업도 생성형 AI 사용을 제한하고 있어요.

만약 생성형 AI에 회사 정보를 올리고 대화를 나누면서 회사 기밀정보를 유출한 경우 형법, 부정경쟁방지 및 영업비밀 보호에 관한 법률 등에서 정한 일정한 요건에 해당하면 범죄행위가 되어 처벌될 수 있습니다.

10) 정병일, "챗GPT와 내부 업무 결합하면 기밀 유출 위험", AI타임스, 2023. 4. 19.

***형법 제356조(업무상의 횡령과 배임)**

업무상의 임무에 위배해 제355조의 죄를 범한 자는 10년 이하의 징역 또는 3천만 원 이하의 벌금에 처한다. 〈개정 1995. 12. 29.〉

***형법 제355조(횡령, 배임)**

② 타인의 사무를 처리하는 자가 그 임무에 위배하는 행위로써 재산상의 이익을 취득하거나 제3자로 하여금 이를 취득하게 해 본인에게 손해를 가한 때에도 전항의 형과 같다.

***부정경쟁방지 및 영업비밀보호에 관한 법률 제18조(벌칙)**

① 영업비밀을 외국에서 사용하거나 외국에서 사용될 것임을 알면서도 다음 각 호의 어느 하나에 해당하는 행위를 한 자는 15년 이하의 징역 또는 15억 원 이하의 벌금에 처한다. 다만, 벌금형에 처하는 경우 위반행위로 인한 재산상 이득액의 10배에 해당하는 금액이 15억 원을 초과하면 그 재산상 이득액의 2배 이상 10배 이하의 벌금에 처한다. 〈개정 2019. 1. 8.〉

1. 부정한 이익을 얻거나 영업비밀 보유자에 손해를 입힐 목적으로 한 다음 각 목의 어느 하나에 해당하는 행위
   가. 영업비밀을 취득 · 사용하거나 제3자에게 누설하는 행위
   나. 영업비밀을 지정된 장소 밖으로 무단으로 유출하는 행위
   다. 영업비밀 보유자로부터 영업비밀을 삭제하거나 반환할 것을 요구받고도 이를 계속 보유하는 행위

2. 절취 · 기망 · 협박, 그 밖의 부정한 수단으로 영업비밀을 취득하는 행위

3. 제1호 또는 제2호에 해당하는 행위가 개입된 사실을 알면서도 그 영업비밀을 취득하거나 사용(제13조제1항에 따라 허용된 범위에서의 사용은 제외한다)하는 행위

또한 생성형 AI에 올린 정보가 개인정보인 경우 개인정보보호법을 위반한 범죄행위가 될 수 있습니다. 특히 개인정보 처리자라면 개인정보보호법에 따른 손해배상 책임을 부담하는 것은 물론 형사 처벌의 대상이 될 수 있어요. 그리고 관련 행위로 회사와 개인정보 주체에게 재산적·정신적 손해를 입혔다면 민법에 따라 손해를 배상해야 합니다.

*개인정보 보호법 제39조(손해배상책임)

① 정보 주체는 개인정보처리자가 이 법을 위반한 행위로 손해를 입으면 개인정보처리자에게 손해배상을 청구할 수 있다.

*개인정보 보호법 제71조(벌칙)

① 제17조 제1항 제2호에 해당하지 아니함에도 같은 항 제1호를 위반해 정보 주체의 동의를 받지 아니하고 개인정보를 제3자에게 제공한 자 및 그 사정을 알고 개인정보를 제공받은 자

*민법 제750조(불법행위의 내용)

고의 또는 과실로 인한 위법행위로 타인에게 손해를 가한 자는 그 손해를 배상할 책임이 있다.

한편 생성형 AI를 사용해 업무를 처리할 때에는 법적 문제에 대한 우려뿐 아니라 생성형 AI로 인해 자신의 문장력, 사고력, 문제해결력이 저하될 수 있음을 인식하는 것도 중요합니다. 생성형 AI에 지나치게 의존하지 않고 업무를 처리하는 태도를 가질 필요가 있어요.

생성형 AI로부터 좋은 결과물을 얻기 위해서는 질문을 똑똑하게 해야 합니다. 그렇지 않을 경우 엉뚱한 답을 제시하니까요. 이런 측면에서 생성형 AI의 활용을 21세기에 필요한 질문하는 역량을 높이는 계기로 만든다면 좋을 것입니다.

관련사례

▶ **[뉴스in뉴스] 챗GPT로 기밀 유출…기업·정부 '고심'(KBS뉴스 2023. 4. 6.)**

챗GPT 열풍이 계속되는 가운데 기업이나 정부의 중요한 자료가 외부로 유출될 수 있어 주의가 필요하다며, 삼성전자 반도체 부문에서 사내 챗GPT 사용을 허가한 뒤 발생한 실제 사고에 대해 보도함.

반도체 공장에서는 장비를 제어하기 위한 다양한 소프트웨어 프로그램이 사용되는데 직원이 프로그램의 오류를 확인하기 위해 챗GPT에 질문하는 과정에서 프로그램 소스 코드가 챗GPT 서버로 넘어가 기밀이 유출될 우려가 있음. 또 회의 내용을 정리하기 위해 챗GPT로 회의 녹음 자료를 보낸 경우도 문제가 됨.

삼성전자는 이에 대해 "회사 내부 사정이라 확인해 주기 어렵다"라고 공식 입장을 발표함. 취재 결과 내부 분위기는 유출된 정보가 민감한 것은 아니지만 보안 강화를 위해서 조치를 하고 있는 것으로 확인됨.

6 회사가 이번에 신제품 마케팅 관련 경쟁사 PT에 참여하게 되었는데요, 신제품 로고를 생성형 AI로 만든 이미지를 활용해서 제안해도 될까요?

NO!

생성형 AI를 사용해 신제품 로고를 만들 경우 그 이미지 권리가 누구에게 있는지 불분명해요. 나중에 문제가 생길 수 있으므로 각별한 주의가 필요해요!

현행법상으로는 생성형 AI로 만든 이미지에 대한 권리는 누구에게 있는지 불명확합니다. 타인의 저작권이나 상표권을 침해할 소지가 있으므로 그 사용에 있어 각별한 주의가 필요해요.

최근 생성형 AI를 이용해 텍스트뿐 아니라 원하는 이미지를 만들 수 있는 기술적 여건이 다양하게 갖추어지고 있죠. 대표적으로 오픈AI의 챗GPT, 구글의 바드(Bard)와 같은 생성형 AI는 이미지 데이터를 학습해 이용자가 원하는 이미지를 설명하고 이를 그려 달라고 요구하면 내 머릿속의 이미지를 현실 속 그림으로 그려 줍니다. 또 달리(Dall-E)와 미드저니(Midjourney) 같은 서비스는 이 영역에 특화된 생성형 AI로 많은 사람이 사용하고 있어요.

하지만 생성형 AI를 이용해 만든 이미지를 PT에 사용하는 것처럼 상업적으로 활용할 수 있을지 여부는 현행법상으로 불분명한 이른바 회색 영역(Gray Area)에 머물러 있습니다.

아직은 생성형 AI가 만든 저작물이 원저작자의 저작물 또는 2차적 저작물에 해당해 원저작자의 저작권을 침해한 것이라는 주장과 그렇지 않다는 주장이 혼재되어 있어요. 또한 생성형 AI가 만든 저작물이 이용자의 저작물이 아니라 생성형 AI나 이를 서비스하는 제공자의 저작물에 해당해 생성형 AI나 이를 서비스하는 제공자의 저작권을 침해한 것이라는 주장과 그렇지 않다는 주장도 양립하고 있습니다.

한편 그것이 신제품 로고와 같이 상표권의 대상인 경우 생성형 AI의 학습 대상인 상표에 대한 권리를 가진 자의 상표권을 침해할 소지도 있어요.

결국 이와 같은 법적인 문제에서 핵심 쟁점은 생성형 AI가 그린 이미지와 학습 대상인 원본과의 동일성 여부에 있습니다. 만약 생성형 AI 이미지가 학습 대상인 원본과는 완전히 다른 새로운 창작물로 인정되는 수준이라면 원저작자의 저작권 침해 문제는 일어나지 않아요.

따라서 생성형 AI가 그린 이미지를 이용하면서 법적 문제를 피하는 좋은 방법은 그것을 원본 저작물과 동일성이 없도록 수정하는 것입니다. 하지만 현실적으로 이는 쉽지 않아요. 원본 저작자의 문제 제기가 있기 전에는 대부분 이용자가 생성형 AI 이미지의 학습 대상, 즉 원본이 무엇인지 인식하기 힘드니까요.

결국 생성형 AI 이미지를 이용하는 최선의 방법은 이미지에 생성형 AI의 역할과 이용자의 역할에 대해 자세한 설명을 명시하는 것입니다. 실제로 오픈AI의 챗GPT는 사용자 가이드라인을 통해 결과물을 사용하는 경우 AI 모델의 역할과 인간 사용자 역할을 명확히 설명하는 문구를 붙일 것을 제안하고

있어요.[11] 이렇게 할 경우 저작권법이나 상표법 침해 문제가 사실상 발생하지 않거나 발생하더라도 책임이 가벼워질 수 있습니다.

관련사례

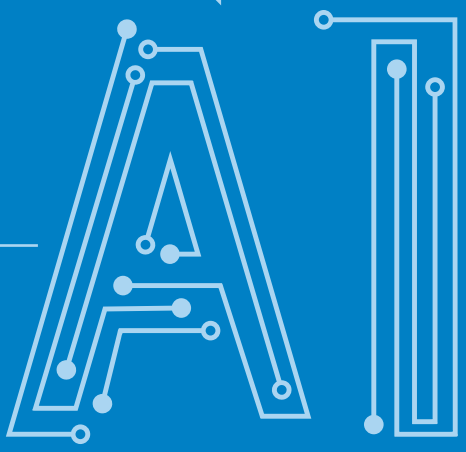
▶ "상상 속 콘텐츠 다 만들어줘" 이미지 생성 AI 인기 속에 고개 드는 저작권 침해 논란(조선일보 2023. 4. 8.)

세계 최대 이미지 제공업체 게티이미지와 이미지 생성 AI '스테빌리티 AI'의 개발사 스테이블 디퓨전이 법적 다툼을 시작함. 블룸버그는 이를 "AI 애플리케이션의 등장으로 발생하고 있는 모호한 저작권 관련 법적 논쟁의 사례 중 하나"라고 평가함.

AI의 창작물의 저작권은 아직 법적으로 회색지대에 놓여 있음. 국내에서도 이미지 생성 AI가 인기를 끌면서 우려의 목소리가 커지고 있음. AI가 인간 창작자의 작품을 무분별하게 학습에 사용하는 경우, 향후 해외처럼 법적 분쟁에 휘말릴 가능성도 있다고 함.

11) 전정현, "[프리미엄 리포트] 생성형 AI가 그린 그림으로 돈을 벌 수 있을까", 동아사이언스, 2023. 4. 1.

생성형



윤리 가이드북

허위조작정보

04

PART

04 허위조작정보

나도 헷갈리는데 남들은? [12]

2023년 11월 일본의 기시다 총리가 뉴스 프로그램에서 외설절인 발언을 하는 동영상이 SNS를 통해 퍼지면서 큰 파장이 일었습니다. 해당 뉴스 동영상을 본 전 세계인들은 일본 총리의 부적절한 언사에 경악을 금치 못했는데요, 하지만 아니었습니다. 실제 뉴스가 아니라 AI가 만든 가짜 뉴스 화면이었어요. 바로 '딥페이크(Deepfake)'였죠.

이처럼 최근 AI 기술의 범람 속에 그 악용에 대한 우려도 높습니다. 특히 허위조작정보를 생성하는 AI가 사회적으로 큰 문제를 일으키고 있어요. 대표적인 사례가 2023년 5월 미국 국방부 옆 건물이 화염에 휩싸인 사진 한 장 때문에 미국 증시가 급락한 사건입니다.

미국의 조 바이든 대통령은 AI 규제 등의 내용이 담긴 행정 명령에 서명하면서 자신의 딥페이크를 본 적 있으며 스스로도 헷갈렸다라고 말하기도 했어요. 또한 유튜브는 AI를 통해 생성 또는 변환한 영상인지를 반드시 표시하도록 의무화하기도 했습니다.

AI 챗봇으로 보이스 피싱까지?

한편 최근에는 생성형 AI를 사용해 사람의 목소리를 그대로 흉내내는 것뿐 아니라 실제 대화가 가능하도록 조작하는 기술마저 나타나면서 보이스 피싱에 악용되는 경우도 발생하고 있습니다. 실제로 미국의 컴퓨터 바이러스 백신 업체인 맥아피는 생성형 AI를 사용할 경우 단 3초짜리 목소리 샘플만으로 특정인의 목소리를 감쪽같이 복제하는 것이 가능하다고 밝혀 이슈가 되었어요.[13]

이처럼 이미지나 영상, 음성 등을 생성형 AI로 만들 경우 매우 정교하기 때문에 보통 사람은 진위를 구별하기가 쉽지 않습니다. 아직은 가짜 뉴스인지 확인할 수 있는 방법들도 명확히 나와 있지 않고요.

따라서 이용자 스스로가 각별히 주의하는 수밖에 없습니다. 조금이라도 의심스러운 부분이 있으면 반드시 더블 체크를 해야 해요. 특히 출처가 불분명한 온라인상의 콘텐츠나 돈을 요구하는 경우 등은 반드시 비판적 시각을 가지고 확인 또 확인을 거치는 것이 좋습니다.

12) 홍영선, '보고도 속는다, AI 가짜 뉴스 범람…딥페이크 전쟁', CBS노컷뉴스, 2023.11.12.
13) 유한주, 'AI, 3초면 특정인 목소리 완벽 복제…보이스피싱에 악용', 연합뉴스, 2023.6.14.

1 재미로 생성형 AI를 이용해서 가짜 뉴스를 만들어 배포했어요. 누군가에게 피해를 줄 목적이 아니었는데도 처벌받나요?

Yes!

누군가에게 피해를 줄 목적이 없었더라도, 가짜 정보라는 인식을 하고 있었다면 형법 혹은 정보통신망 이용촉진 및 정보보호 등에 관한 법률에 따라 형사 처벌 대상이 될 수 있어요. 상대방에게 손해를 끼친 경우에는 민사상 손해배상 청구의 대상이 될 수도 있고요!

세상에 존재하지 않는 내용을 다루는 가짜 정보를 기반으로 한 가짜 뉴스를 만들어 배포하는 것 자체는 위법하다고 볼 수는 없습니다. 그러나 누군가에게 피해를 줄 목적이 없었더라도, 이러한 가짜 뉴스가 실존하는 개인 혹은 법인의 명예를 훼손하는 내용을 다루고 있거나, 이러한 내용으로 인해 상대방의 명예나 인격이 손상될 수 있다는 점을 인식하고 있다면(혹은 더 나아가 명예훼손 등의 고의가 있다면), 이는 관련 법령에 따른 형사 처벌 대상이 될 수 있음은 물론, 민사상 손해배상 청구 소송의 대상이 될 수도 있어요.

다시 말해 생성형 AI 기술을 이용해 사진이나 영상을 제작하고, 이를 토대로 가짜 뉴스 영상을 제작하는 행위는 그 자체로 하나의 창작적인 행위라고 볼 수 있지만, 이렇게 제작된 영상을 마치 실제 내용을 담고 있는 뉴스 영상인 것처럼 배포하는 경우에는 법적인 책임을 부담할 수도 있습니다.

최근 미국에서 생성형 AI 기술을 활용해 미국 국방부(펜타곤) 청사 근처에서 대형 폭발이 일어났다고 주장하는 사진이 SNS를 통해 확산되었는데, 이로 인해 미국 주가가 일시적으로 하락하는 등의 사건이 발생해 문제가 된 적이 있어요.

출처 : CNN(2023. 5. 23.), "'Verified' Twitter accounts share fake image of 'explosion' near Pentagon, causing confusion"

이처럼 가짜 뉴스가 단순히 하나의 창작물로 배포되는 것이 아니라 진짜 뉴스처럼 배포되고, 누군가에게 피해를 줄 목적이 없었다고 할지라도 해당 뉴스로 인해 직접적인 손해를 입은 사람이 존재한다면 이러한 손해에 대해서는 민사상 손해배상 책임을 부담할 수 있습니다. 때로는 특별법에 따라 처벌을 받을 수도 있어요.

나아가 가짜 뉴스를 제작하는 과정에서 실존 인물의 사진이나 영상을 이용했다면 해당 인물의 초상권을 침해할 여지도 있어요. 초상권 역시 당사자로부터 그 이용을 위한 허락을 받지 않았다면 초상권 침해에 따른 민사상 손해배상 책임을 부담할 수 있다는 점을 유념해야 합니다.

관련사례

▶ AI발 가짜 뉴스에 눈 뜨고 코 베이는데…규제는 미비(연합뉴스TV 2023. 7. 25.)

최근 한 일본 기자가 프랑스 축구클럽 PSG 소속 축구 선수 킬리안 음바페에게 "이강인이 같은 팀으로 온다고 하는데, 단순한 마케팅용이라고 생각하냐"라는 질문을 하고, 음바페 선수가 "이강인은 재능을 가졌기에 올 수 있는 것"이라고 답한 영상이 화제가 됨.

하지만 이 영상은 2021년 개최된 유로 2020 기자회견 영상에 AI 기술을 이용해 음바페 선수의 답변을 짜깁기하고, 텍스트 음성 변환 기술로 일본 기자의 음성을 덧댄 것으로 밝혀짐.

현행 법령상 이러한 가짜 뉴스에 대비한 조치는 정정 보도 청구, 명예 훼손에 따른 민형사 소송 정도가 가능함. 즉 AI 기술의 등장에 따른 새로운 유형의 가짜 뉴스 등에 대해서는 제도가 미비하므로 제도 보완에 속도를 내야 한다는 지적들이 나오고 있음.

2 온라인에서 생성형 AI로 만든 허위조작정보(가짜 뉴스, 스팸 광고 등) 콘텐츠를 발견했어요. 신고할 곳이 따로 있나요?

가짜 뉴스 등 허위조작정보 콘텐츠를 발견하는 경우, 한국언론진흥재단의 〈가짜 뉴스 피해 신고·상담센터〉, 한국인터넷자율정책기구의 〈가짜 뉴스 신고센터〉와 같은 곳에 신고할 수 있어요!

최근 AI 기술이 발달하면서 '진짜보다 더 진짜 같은' 가짜 뉴스 콘텐츠를 제작하고 배포하는 사례들이 증가하고 있습니다. 이러한 허위조작정보 콘텐츠는 단순히 개인에게만 피해를 주는 것이 아니라, 온라인 콘텐츠 특유의 파급력 등으로 인해 많은 사람들에게 혼란을 야기하거나, 심지어는 주식 시장 등에도 영향을 주는 일들이 발생하고 있어요.[14] 이 경우 언론사의 이름으로 발표된 것처럼 제작된 허위조작정보 콘텐츠에 대해서는 해당 언론사에 문의해 그 진위 여부 등에 대해 확인 요청을 할 수 있습니다.

한편 한국언론진흥재단은 최근 AI 기술을 바탕으로 한 허위조작정보 콘텐츠가 범람하기에 이르자 '가짜 뉴스 피해 신고·상담센터'를 열었어요. 이 센터는 가짜 뉴스 혹은 허위조작정보로 피해를 본 국민을 대상으로 상담을 진행한 뒤, 구제 기관을 안내해 주는 업무를 담당합니다. 구체적으로 언론중재위원회 피해 상담, 인터넷 피해 구조 신고, 민·형사상 대응을 위한 법률 지식과 절차 등의 안내까지 제공하고 있어요.

또한 사단법인 한국인터넷자율정책기구는 '가짜 뉴스 신고센터'를 운영 중입니다. 언론사가 아님에도 언론사를 사칭한 허위 게시물을 대상으로 신고를 받아 자체 심의를 거쳐 회원사들이 조치를 취할 수 있도록 하고 있어요.

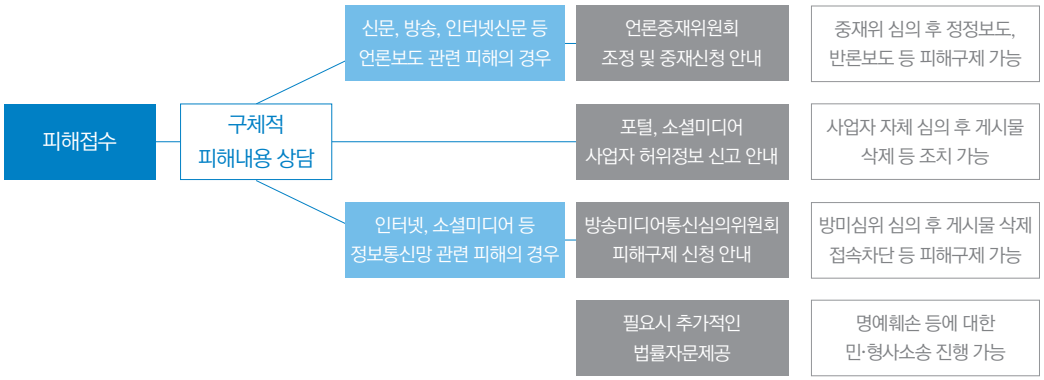
뿐만 아니라 허위조작정보 콘텐츠가 누군가의 명예를 훼손하는 내용을 담고 있는 경우에는 피해자가 아니더라도 해당 콘텐츠를 제작, 유포하는 이들을 '고발'할 수 있습니다. '고발'이란 보통 피해자에 해당하는 고소권자, 그리고 범인 이외의 제3자가 수사기관에 범죄사실을 신고하는 행위를 의미하는데, 가짜 뉴스 등을 통해 허위 사실을 유포해 누군가의 명예를 훼손하는 행위를 하는 자에 대해서는 피해자가 아니더라도 수사기관에 위 범죄사실을 고발해 수사의 진행을 요청할 수 있어요.

14) 손서영, "'가짜 뉴스'에 예금해지했다…금융권 뒤흔드는 '지라시'", KBS뉴스, 2023. 4. 14.

관련사례

▶ 한국언론진흥재단 가짜 뉴스 피해 신고·상담 방법

가짜 뉴스로 인해 피해를 입은 사람은 한국언론진흥재단에 전화(02-2001-7205~8), 이메일(damagerelief@kpf.or.kr)을 비롯해 홈페이지나 직접 방문 상담을 진행할 수 있으며, 피해 사실 접수 및 상담 후에는 아래와 같은 절차를 통해 도움을 받을 수 있음



출처 : 한국언론진흥재단

3 생성형 AI로 엄마 목소리를 똑같이 만들어 전화를 거는 바람에 속아서 돈을 송금했어요. 만약 아직 출금 전이라면 보이스 피싱 사기 피해를 구제받을 수 있을까요?

Yes!

생성형 AI를 통한 보이스 피싱 등의 사기 피해를 당한 경우 우선 경찰청, 금융감독원 그리고 입금, 송금을 한 금융회사 고객센터에 신고해 피해 금액이 범인에게 넘어가지 않도록 해야 해요. 그리고 경찰서(사이버 수사대)를 방문해 사건사고 확인원을 발급받아 지급 정지를 신청한 금융회사 영업점에 제출함으로써 피해금 환급신청을 할 수 있어요!

가족이나 친구 등의 목소리를 흉내 내거나 수사기관 등을 사칭해 송금이나 계좌 이체를 요구하는 범죄 수법인 '보이스 피싱'은 AI 기술을 통해 더욱 교묘해지고 있습니다.

2023년 3월 캐나다에서는 범인이 "아들이 교통사고로 미국인 외교관을 숨지게 한 혐의로 수감됐다"라며 부모에게 2만 1,000캐나다 달러(약 2,201만 원) 상당의 암호화폐를 송금하라고 요구한 사건이 있었고, 부모는 감쪽같이 속아 피해를 입었는데 당시 범인은 딥페이크 기술을 통해 아들의 목소리를 제작했던 것으로 알려졌어요. 또 영국의 한 에너지 회사는 최고경영자(CEO) 목소리를 AI가 위조한 보이스 피싱 공격을 받아 24만 3,000달러(약 3억 933만 원)를 송금하는 사기 피해에

휘말리기도 했습니다.[15]

이처럼 더욱 교묘해진 생성형 AI 기술로 인해 보이스 피싱 사기를 당해 송금, 이체 등을 한 경우에는 신속하게 경찰처(112)나 금융감독원(1332)에 이러한 사실을 신고하고, 입금, 송금을 한 금융회사 고객센터에 이러한 사실을 밝히며 지급 정지(입출금·이체 금지) 요청 및 피해 구제 신청을 해야 합니다(통신사기피해환급법 제3조 이하의 규정들 참조). 이러한 조치를 통해 지급 정지 통보를 받아 해당 계좌에 대한 지급 정지가 실시된 후에는 금융감독원이 산정한 피해 환급금을 금융회사를 통해 돌려받을 수 있어요.

또한 가족이나 지인들로부터 송금 등의 요청을 받은 뒤에는 다시 한번 연락을 해서 재확인하는 등 스스로 조심할 필요가 있습니다.

한편 최근 은행들도 AI 기술을 도입해 'AI 기반 이상 거래 탐지 시스템'을 운영하고 있어요. 즉 보이스 피싱의 기술로 이용되는 생성형 AI 기술을 활용해 보이스 피싱 모니터링 시스템, ATM 기계에서의 이상 행동을 분석해 고객에게 주의 문구를 안내하고 본인 인증 등 추가 절차를 요구하도록 하는 'AI 이상 행동 탐지 ATM'과 스마트폰 은행 앱 로그인 시 이상 거래 탐지 시스템을 통해 보이스 피싱 앱이 설치되어 있는지 여부 등을 파악하고, 보이스 피싱 앱이 발견되면 거래가 자동 정지되도록 하는 기술들을 도입하고 있는 것으로 알려집니다.[16]

이처럼 생성형 AI 기술은 보이스 피싱의 수단으로 이용될 수 있는 한편, 보이스 피싱 등 금융 사기를 방지하기 위한 수단으로서도 활용되고 있어요.

관련사례

▶ AI가 "엄마, 난데" 보이스 피싱 하는 시대[유레카](한겨레 2023. 6. 15.)

생성형 AI 기술을 통해 누군가의 목소리를 흉내 내도록 하는 것은 무척 쉬운 일임. KT의 음성 AI 기술은 3분가량의 녹음만으로도 목소리를 완벽히 복원해 텍스트를 입력하는 대로 해당 목소리의 말을 생성함. SKT 역시 최근 별도의 녹음을 하지 않고도 골프선수 최경주의 목소리, 말투, 사투리 등을 기존 인터뷰 영상에서 추출한 음성만으로도 완벽하게 생성함.

이러한 AI 기술의 악용을 대비하기 위해 경찰청은 최근 대화형 인공지능 챗봇 서비스 상용화에 따른 사이버 범죄 대응 방안에 대한 연구를 발주했고, 대검찰청 역시 가짜 음성 탐지 기술 개발에 착수했음.

15) 류영상, "AI에게 3억원 털렸다…부모도 감쪽같이 속이는 AI 보이스피싱, 대처는", 매일경제, 2023. 7. 15.
16) 정소양, "[대전환 AI시대⑤] 보이스피싱 막는 '만능 재주꾼'…은행권에 스며든 AI", 비즈팩트, 2023. 7. 17.

4 생성형 AI로 만든 정보가 사실인지 아닌지 잘 모르겠어요. 진위 여부를 확인할 수 있는 방법이 있나요?

NO!

생성형 AI가 사실이 아닌 정보를 제공하는 '환각 현상(Hallucination)'은 생성형 AI 데이터와 알고리즘의 한계 때문으로 단기간에 해결되기 어려워요. 따라서 생성형 AI가 만든 정보에는 거짓이 있을 수 있다는 점을 인식하고 그대로 받아들여선 안 돼요!

현재 생성형 AI가 모두 사실인 정보만 제공하는 것은 아닙니다. 거짓된 정보, 출처가 불분명한 정보, 오류와 불완전한 내용들을 포함해 제공하기도 해요. 이것을 이른바 환각 현상(Hallucination)이라고 합니다.

이것은 생성형 AI의 데이터와 알고리즘의 한계 때문에 발생하는 문제로 단기간에 해결되기는 어려워요. 따라서 생성형 AI가 만든 정보 중 의심스러운 내용에 대해서는 반드시 검색을 통해서 사실 확인(팩트 체크)하는 습관을 가져야 합니다. 또한 하나의 생성형 AI만 사용하는 것이 아니라 여러 생성형 AI를 병행해 사용하는 것도 좋아요.

특히 이미지나 영상 생성형 AI로 만들어진 콘텐츠의 경우에는 매우 정교하기 때문에 일반인들은 진짜인지 가짜인지 구별하기가 어렵습니다. 하지만 현재 이러한 생성형 AI로 만든 이미지나 영상의 진위 여부를 정확히 구분할 수 있는 방법은 부족하기 때문에 개인적으로 각별한 주의가 필요해요. 무엇보다 이용자 개인이 온라인 상의 콘텐츠들을 비판적으로 수용하는 태도를 가져야 합니다.

사용자가 생성형 AI 챗봇으로 대화하는 과정에서 나타난 거짓 정보 예시

한국의 현재 대통령은 누구입니까?

현재 대한민국 대통령은 문재인입니다.

한국의 김구 선생님에 대해 알려주세요.

김구 선생님은 대한민국의 독립운동가이며, 독립운동을 주도한 인물 중 하나입니다. 그는 일제 강점기 동안 독립운동을 지속하며, 3.1운동과 4.19혁명을 주도하였습니다. 그는 또한 대한민국 임시정부의 주요 인물 중 하나이며, 광복 이후에는 대한민국 정치의 중요한 인물로 활동했습니다. 그는 민족의식과 민주주의를 실천하며, 대한민국의 역사에서 큰 역할을 했습니다. 현재 대한민국 정부는 그를 "독립운동의 열매"로 인정하고 있습니다.

관련사례

▶ AI 조작 사진에 美 증시 출렁…'가짜 얼굴' 송금 사기도 기승(KBS뉴스 2023. 5. 23.)

'미국 국방부 청사 인근에서 큰 폭발이 일어난 장면'을 AI로 만든 가짜 뉴스 사진 한 장이 SNS를 통해 급속히 확산하면서, 전 세계 언론에서 긴급 속보를 타전하고, 미국 S&P 주가 지수가 하락하며 이슈가 되었음.

이 사진은 순식간에 유명 블로거와 인플루언서들을 통해 급속히 확산됨. 심지어 러시아 해외 선전 매체인 RT도 "펜타곤 근처에 폭발 보도가 있다"라고 SNS에 게재하기도 함.

해당 사진으로 인해 오전 9시 30분 개장하는 미국 증시의 스탠더드앤드푸어스(S&P) 500 지수가 0.3%가량 출렁였고, 위험 상황에서 안전 자산으로 여기는 미국 국채와 금의 가격은 반대로 급등함.

하지만 전문가들이 사진을 분석한 결과, AI가 이미지를 생성하는 과정에서 건물 앞의 서로 다른 담장들이 변형되고 뒤섞인 흔적이 포착되어 AI가 생성한 사진임이 드러남.

생성형

윤리 가이드북

개인정보·인격권

05

05 개인정보·인격권

생성형 AI는 당신이 지난 여름에 한 일을 알고 있다? [17]

생성형 AI는 기존에 나와 있는 데이터뿐 아니라 사용자와 대화를 나누면서 획득한 데이터 역시 새로 학습 자료로 추가하게 됩니다. 그런데 이 과정에서 사용자가 자신의 개인정보나 회사 등 조직의 기밀을 은연 중 누설할 수 있어 주의가 요구되고 있어요.

생성형 AI는 다양한 방식으로 인터넷에 공개되는 대규모 데이터를 수집하면서 의도하지 않더라도 수많은 개인정보를 학습하게 됩니다. 그중 상당량은 사전 통지나 동의를 구하지 않은 것일 수 있고요.
게다가 이러한 데이터는 외부로 유출될 위험도 존재합니다. 실제로 생성형 AI의 대표 격인 챗GPT에서 사용자와 AI가 나눈 대화 기록이 버그 발생으로 유출되면서 서비스가 일시 중단되는 일도 있었어요.
이러한 개인정보 침해, 정보 유출 등에 대한 우려로 일부 기업이나 공공기관 등은 생성형 AI 서비스 사용을 금지하거나 제한하고 있기도 합니다. 우리나라의 삼성전자, 포스코, SK하이닉스, LG디스플레이를를 비롯해 JP모건체이스, 뱅크오브아메리카, 시티그룹, 골드만삭스, 버라이즌 등 해외의 주요 기업들도 챗GPT와 같은 생성형 AI 챗봇의 사용을 제한했습니다.

유익한 것은 취하고 유해한 것은 피하라!

한편 최근 기업들은 생성형 AI 서비스를 무조건 제한하기보다 효과적으로 활용할 수 있는 방법들을 찾고 있어요. 업무 보조 기능 등 장점은 취하면서 핵심 정보 유출 가능성 등은 차단하기 위해 업무 특성을 고려한 생성형 AI 활용 방침을 제정하고 있습니다.

또한 우리나라의 산업통상자원부 국가기술표준원은 'AI 윤리 점검 서식'에 대한 국가 표준(KS)을 제정, 발표하기도 했습니다. 최근 생성형 AI의 사용 윤리 문제가 제기되고 있어 AI 기술을 활용한 제품이나 서비스 개발 시에 어떤 윤리적 고려를 해야 하는지 자체 점검할 수 있도록 체크리스트를 공개했어요.
따라서 사용자들은 이러한 기본적인 가이드라인들을 통해 생성형 AI가 개인정보를 침해하거나 인격권을 훼손하지 않도록 노력할 필요가 있습니다. 다만 제도가 갖춰진다고 해도 결국은 사람의 활용 역량에 달려 있다는 점은 잊지 마시고요.

17) 'THE AI REPORT : 공공분야 생성형 AI 활용 방안', NIA, 2023

1 생성형 AI와 나눈 대화가 학습되어 다른 사람에게 노출될 수 있나요? 혹시 개인의 민감한 정보가 다른 사람에게 노출될지 걱정이 돼서요.

생성형 AI와 나눈 대화는 학습 데이터로서 다른 이용자들에게 제공이 되는 등 노출될 가능성을 배제할 수 없어요. 이러한 일을 방지하기 위해서는 이용자가 자신의 뜻을 생성형 AI의 데이터 제어 설정에 반영한 후 이용하는 것이 바람직해요!

챗GPT를 필두로 대두되는 생성형 AI의 경우, 대화를 하듯이 내용을 주고받으며 해당 AI의 정보 탐색 능력이나 언어 능력 등을 활용하는 일들이 많아지고 있습니다. 특히 챗GPT는 자신들의 이용 약관을 통해 이용자가 챗GPT에 입력하는 정보는 저장되며 성능 개선을 위한 학습 자료로 활용 가능하다고 밝히고 있어요.

챗GPT는 2023년 4월 25일경 개인 계정 설정 기능을 추가했지만, 그럼에도 여전히 사용자가 입력한 정보는 30일간 서버에 저장됩니다. 즉 개정된 이용 약관에도 불구하고 이용자가 챗GPT와 나눈 대화 내용은 학습 자료로 이용될 수 있고 그 과정에서 다른 이용자들에게 노출이 될 여지가 있어요.[18]

또한, 일부 AI 서비스에서 타인의 대화 기록이 다른 이용자에게 보이는 버그 현상이 발생한 사례도 존재합니다. 해당 사례에서는 대화 기록과 함께 금융 관련 정보, 이름, 이메일 결제 정보, 카드 정보, 카드 4자리 숫자 정보 등이 일반 사용자에게 노출된 것으로 알려져 있어요.[19]

그러므로 챗GPT와 같은 생성형 AI를 이용할 때에는 주민등록번호, 신용카드 정보, 비밀번호 등의 개인정보뿐만 아니라 자신이 소속된 회사, 조직 등의 정보를 입력하지 않도록 주의가 필요합니다.

18) 행정안전부, '챗GPT 활용방법 및 주의사항 안내', 2023. 5. 8.
19) 국가정보원, 국가보안기술연구소, '챗GPT 등 생성형 AI 활용 보안 가이드라인', 2023. 6. 29.

[챗GPT 등 생성형 AI 활용 보안 가이드라인]

데이터 제어 설정	· 데이터 제어 설정을 통한 채팅 기록 및 모델 학습 비활성화
개인정보 입력 금지	· 개인정보 침해 및 악용 방지를 위한 민감한 개인정보(건강, 종교, 정치적 성향 등) 입력 금지 · 금융적 손실 및 신용도 피해 방지를 위한 금융 관련 정보(신용카드 번호, 은행계좌 정보, 비밀번호 등) 입력 금지
기관 내 민감 정보 입력 금지	· 기관 내부자료 등 민감정보 입력 금지로 기관 정보 유출 차단 · 가명화 및 익명화를 통한 실제 개인정보와 기관과의 관계 유추 가능성 차단
보안 질문 회피	· AI 모델과 대화 중 보안 관련 질문 발생 시 답변 자제 · 필요시 고객 지원 센터 문의를 통한 올바른 절차 준수
인증 정보 입력 금지	· 계정 침해 위협 방지를 위한 인증 정보 (사용자 이름, 비밀번호, 인증코드 등) 입력 금지

출처 : 국가정보원, 국가보안기술연구소

관련사례

▶ **[AI 리뷰] 개인정보위, 오픈AI 대화형 인공지능 챗GPT에 360만 원 과태료 부과와 개선 권고는 왜 했나?(인공지능신문 2023. 7. 28.)**

최근 개인정보보호위원회는 챗GPT를 운영하는 오픈AI를 상대로 개인정보 유출 발생 및 그에 대한 신고 의무 위반으로 과태로 360만 원을 부과하고, 재발 방지 대책 수립, 국내 보호법 준수, 개인정보보호위원회의 사전 실 태점검에 적극 협력할 것을 내용으로 하는 개선 권고를 의결함.

이는 2023년 3월 20일 오후 5시부터 3월 21일 오전 2시 사이 챗GPT 플러스에 접속한 전 세계 이용자 일부의 성명, 이메일, 결제지, 신용카드 번호 4자리와 만료일이 다른 이용자들에게 노출된 데 따른 것임. 위와 같은 정보들이 유출된 것은 서비스 속도 증가를 위해 오픈소스 기반 캐시 솔루션에 버그가 발생한 것으로 파악되었음.

한편 오픈AI는 개인정보 처리 방침을 영문으로만 제공하고 있고, 별도 동의 절차가 없는 등 국내 개인정보보호법 준수가 미비한 점 또한 확인됨에 따라 이러한 사항에 대한 개선을 권고받음.

2 생성형 AI와 대화한 내용이 해당 기업 서버에 저장되거나 기업 관계자가 확인할 수 있다고 들었어요. 사실인가요?

생성형 AI에 입력되는 내용들은 해당 서비스 기업 서버에 저장되고, AI의 학습에 재이용되는 등 외부 유출의 위험이 있어요. 따라서 개인정보나 회사 기밀 등 보안이 필요한 민감한 정보들은 생성형 AI에 입력하면 안 돼요!

생성형 AI는 업무와 학습에 이용할 수 있는 매우 편리한 도구입니다. 하지만 주의해야 할 사항이 있어요. 바로 생성형 AI에 입력하는 내용들이 생성형 AI 기업 서버에 저장되고 해당 AI의 학습에 재이용될 수 있다는 점이에요.

따라서 생성형 AI에 개인정보나 회사 기밀 등을 입력하게 되면 외부 유출 위험이 커지며, 해당 내용들이 AI의 학습 데이터로 이용되어 다른 사용자가 생성한 결과물에 드러나 유출될 수도 있습니다.

오픈AI의 챗GPT 개인 대화 내용 취급 정책

What is ChatGPT?

Commonly asked questions about ChatGPT

5. **Who can view my conversations?**
   - As part of our commitment to safe and responsible AI, we review conversations to improve our systems and to ensure the content complies with our policies and safety requirements.

6. **Will you use my conversations for training?**
   - Yes. Your conversations may be reviewed by our AI trainers to improve our systems.

7. **Can you delete my data?**
   - Yes, please follow the data deletion process.

8. **Can you delete specific prompts?**
   - No, we are not able to delete specific prompts from your history. Please don't share any sensitive information in your conversations.

출처 : 오픈AI FAQ(https://help.openai.com/en/articles/6783457-what-is-chatgpt)

관련사례

▶ "삼성전자, 직원들 챗GPT 사용 금지"…회사 정보 유출 차단(머니투데이 2023. 5. 2.)

블룸버그는 국내 대표기업인 삼성전자가 회사 소유의 컴퓨터, 태블릿, 휴대폰은 물론 내부 네트워크에서 생성형 AI 시스템 사용을 전면 금지하기로 했다고 보도함. 삼성전자의 한 사업장에서 생성형 AI를 통해 기업의 민감한 정보가 최소 세 차례 미국 기업인 오픈AI의 학습 데이터로 입력된 데에 따른 조치로 알려짐.

회사는 직원들에게 생성형 AI로 인한 보안 위험이 커지고 있다고 밝히며, 보안지침을 성실히 준수할 것을 요청하고, 이를 위반할 경우 회사 정보 유출로 인해 최대 해고를 포함한 징계 조치를 받을 수 있다고 공지함. 한편 삼성전자가 내부적으로 AI 도구 사용에 대한 설문조사를 실시한 결과 응답자의 65%가 보안 위험을 초래한다고 생각한다고 답함.

3 생성형 AI를 활용하는 도중 특정인을 차별하고 비난하며 명예 훼손하는 내용을 확인했는데요, 작성자인 생성형 AI를 처벌할 수 있나요?

NO!

AI는 인간이 아니기에 법적인 처벌을 할 수는 없어요. 따라서 생성형 AI를 활용하다 특정인의 명예를 훼손할 수 있는 내용을 확인했을 경우, 책임감을 가지고 프롬프트를 수정하거나 추가적인 정보를 제공해 생성형 AI가 이런 내용을 반복해 제공하지 않도록 해야 해요!

생성형 AI는 데이터 학습을 통해 이용자들의 문서 작업이나 예술 활동에 이르기까지 다양한 도움을 주고 있습니다. 그러나 생성형 AI가 학습하는 데이터는 결국 인간이 제공하는 것이기 때문에 편향된 정보나 특정 단체, 사람, 종교, 인종 등을 혐오하는 내용의 학습 데이터를 생성할 수 있고, 이로 인해 누군가의 명예를 훼손할 여지가 있어요.

실제로 2021년 스캐터랩이 개발하고 너티(Nutty)가 서비스하는 대화형 인공지능 '이루다'의 경우 성소수자에 대한 혐오 발언을 답변으로 제공하는 사례가 발견되어 편향적인 학습의 부작용이 대두된 바 있어요.[20] 이러한 사례는 마이크로소프트가 2016년경 출시한 대화형 인공지능 '테이(Tay)'의 경우에도 발생했는데, 이용자들이 인종차별적 용어, 자극적인 성차별, 정치적인 발언 등을 학습시킨 것이 원인이었습니다.

20) 한영혜, "20살女 AI에 '레즈비언' 꺼내자 한 말 '질 떨어져 소름끼친다'", 중앙일보, 2021. 1. 10

이처럼 생성형 AI는 이용자들이 AI를 사용하는 과정에서 제공하는 내용들을 학습하게 되므로, 특정 이용자들이 악의적으로 편향된 정보를 비롯해 허위조작정보를 입력할 경우 그와 같은 내용들이 보편적인 정보로서 다른 이용자들의 AI 사용 과정에서 나타날 가능성이 있어요. 그러므로 생성형 AI를 이용하는 과정에서 특정인을 차별하고 명예를 훼손하는 내용을 발견한다면 해당 AI의 프롬프트 수정 및 올바른 내용을 학습시켜, 더 많은 피해가 발생하지 않도록 할 필요가 있습니다.

생성형 AI 이용 과정에서 이용자들이 준수해야 할 이러한 윤리의식 역시 AI 기술 못지않게 중요합니다. AI가 생성한 명예 훼손적 내용 그리고 타인의 인격을 훼손하는 내용에 대해서는 AI가 인간이 아니기에 법적인 처벌을 할 수는 없어요. 그러나 이러한 내용을 악의적으로 이용해 가짜 뉴스 등의 콘텐츠를 제작하는 경우에는 형사 처벌의 대상이 될 수 있습니다. 온라인 콘텐츠의 파급력 등을 고려했을 때 이용자들이 생성형 AI를 이용하는 단계에서부터 윤리의식을 가지고 잘못된 내용은 수정하도록 하고, 책임감 있는 방식으로 서비스를 이용해야 해요.

관련사례

▶ 인공지능(AI) 윤리 국가표준(KS) 첫 제정(국가기술표준원, 2023. 6. 14.)

산업통상자원부 국가기술표준원은 'AI 윤리 점검 서식'에 대한 국가 표준(KS)을 제정, 발표함.

최근 생성형 AI의 윤리적인 사용 문제가 제기되는 시점에서 AI 제품, 서비스 개발 시에 필요한 윤리적 고려 항목을 제시하고 자체 점검할 수 있는 체크리스트를 공개해 기본적인 가이드라인의 역할을 할 수 있을 것으로 전망됨.

4.9 관련 윤리 항목 점검표

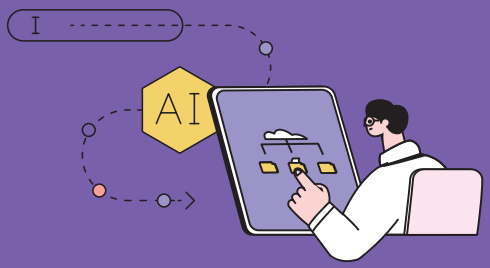
	서비스 제공자	사용자	개발자
(1) 투명성	☐	☐	☐
(2) 공정성	☐	☐	☐
(3) 무해성	☐	☐	☐
(4) 책임성	☐	☐	☐
(5) 사생활 보장성	☐	☐	☐
(6) 편익성	☐	☐	☐
(7) 자율성	☐	☐	☐
(8) 신뢰성	☐	☐	☐
(9) 지속성	☐	☐	☐
(10) 연대성	☐	☐	☐

→ 행위 주체별(서비스 제공자, 사용자, 개발자) 책임지거나 고려 되어야하는 윤리적인 항목

출처 : 국가기술표준원

생성형

AI 윤리 가이드북

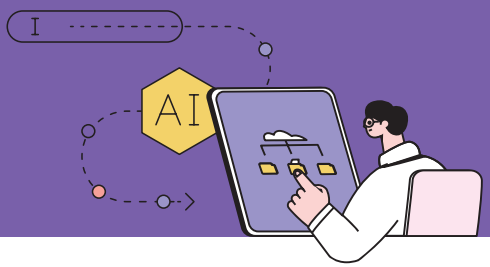


오남용

06

PART

06 오남용



생성형 AI는 나만의 일타 강사?

챗GPT 등 자연어 생성형 AI를 처음 사용해 본 사람이라면 빠른 답변 속도에 한 번 놀라고 거침없이 프롬프트를 채워 나가며 스크롤을 유도하는 정보량에 두 번 놀라게 됩니다. 그리고 과연 이번엔 어떤 답변을 해줄지 궁금해하며 이 질문, 저 질문 던져 보게 되죠. 그러다 보면 때로는 자신에게 필요한 정답을 콕 집어 말해 주는 AI 챗봇이 마치 최고의 일타 강사처럼 느껴질지도 모릅니다.

그런데 문제는 이렇게 가려운 데를 쏙쏙 긁어 주는 생성형 AI의 편리함에 빠져 오남용하는 사례가 발생하고 있다는 사실입니다. 심지어 생성형 AI 의존증이 생긴 사람도 나타나고 있어요. 궁금한 게 생기면 챗GPT부터 찾고 챗GPT가 해주는 말은 무조건 믿는다고 하는 경우죠. 챗GPT에 마음의 고민을 털어놓고 심리 상담을 받는다는 사람도 있습니다.

하지만 챗GTP, 즉 생성형 AI는 척척박사처럼 보이는 데이터 재편집자에 불과합니다. 생성형 AI가 만들어 낸 결과물의 바탕에는 언젠가 사람이 순수 창작해 낸 원본이 있으니까요.

생성형 AI는 도구일뿐!

하지만 이러한 인간 고유의 능력도 사용하지 않으면 퇴보하고 말 거예요. 그러니까 우리 모두 인간이 가진 강점에 우선해서 생성형 AI를 활용해 보도록 해요! 인간이 지닌 위대한 능력을 AI가 대신해 주는 것이 아니라, 생성형 AI는 우리를 돕는 도구에 불과하다는 것 꼭꼭 기억하기로 해요.

나부터 먼저 생각하고 해결하기 어려운 부분이나 도움받고 싶은 부분을 생성형 AI에 추가로 요청하면 좋겠죠?
생성형 AI가 준 결과에 나만의 아이디어를 결합해서 더 새롭게 만들어 보는 건 어떨까요?
생성형 AI에 과도하게 의존하지 말고, 나 자신을 믿고 먼저 길을 걸어가 보길 바랍니다.

하지만 이러한 인간 고유의 능력도 사용하지 않으면 퇴보하고 말 거예요. 프랑스의 사상가 파스칼은 “인간은 자연 가운데 가장 약한 하나의 갈대에 불과하다. 그러나 그것은 생각하는 갈대다”라고 말했습니다.

절대적인 지식 존재로서의 인간은 한계가 있을 수 있지만 사유하고 창의하는 존재로서의 인간은 무한하며 위대합니다. 이 사실을 잊지 말고 AI를 활용할 때도 생각하는 갈대로 남으시길 바랍니다.

1 생성형 AI를 써보니 결과도 빠르게 알려주고, 검색보다 쉽고 편하게 원하는 정보를 얻을 수 있더라고요. 이제는 어떤 일이든 맨 처음 생성형 AI에 묻고 있는데요, 제가 너무 생성형 AI에 의존하는 것일까요?

생성형 AI로 업무를 효율적으로 하는 것은 좋으나 이에 지나치게 의존하는 것은 문장력, 사고력, 문제해결력이 저하될 우려가 있으니 각별한 주의가 필요해요!

생성형 AI는 사람의 업무 효율을 높이고 좋은 질문을 할 수 있는 역량을 강화해 주는 장점이 있습니다.

먼저 생성형 AI를 업무에 잘 활용하면 업무 효율을 획기적으로 높일 수 있어요. 예를 들어 생성형 AI를 활용하면 제안서 작성, 행정 문서 작성, 프레젠테이션 작성은 물론이고, 작성한 문서를 외국어로 번역한 문서 작성, 소스 코드 오류 탐색과 개선 등을 매우 빠르게 할 수 있죠.

또한 생성형 AI에서 좋은 결과물을 얻기 위해 똑똑하게 질문하는 과정에서 질문 역량이 향상될 수도 있습니다. 생성형 AI에게 질문을 제대로 못하면 엉뚱한 답을 제시하니 원하는 답을 얻기 위해 질문을 더 똑똑하게 하는 역량을 키우는 간접 효과를 얻게 될 수도 있어요.

그러나 이러한 장점에도 불구하고 생성형 AI에 대한 지나친 의존은 여러 가지 문제를 유발할 수 있습니다. [21]

첫째, 문장력이 약화될 수 있어요. 생성형 AI는 간단한 문장으로 질문을 해도 이에 관해 완성된 문장 형태로 장문의 답을 제시합니다. 따라서 사람이 작성해야 할 글을 자주 생성형 AI의 답에 의존해 작성하다 보면 자칫 문장력이 저하될 수 있어요.

둘째, 사고력이 약화될 수 있습니다. 사람이 글을 작성하기 위해서 필요한 자료를 생성형 AI에 의존해 취득하면 사람 스스로 그에 관해 생각해 볼 기회를 얻지 못하게 되고, 이것이 반복되면 사고력이 저하될 수 있어요. 그 결과 다른 관점에서 생각하는 비판적 사고, 틀 밖에서 생각하는 창의적 사고를 하는 것이 점점 어려워질 수 있습니다.

셋째, 문제 해결력이 약화될 수 있어요. 사람이 어떤 문제에 부딪히면 이를 해결하기 위해 혼자 생각하고 다른 사람과 소통을 하며 잠정적인 문제 해결책을 도출한 후 이를 실행해 봅니다. 그리고 그와 같이 실행해 본 후 오류가 있어 문제를 해결하지 못했다면 그 오류를 찾아 수정해 문제를 해결하는 절차로 부딪힌 문제를 해결하게 돼요. 그런데 생성형 AI에 의존하다 보면 다른 사람과 소통 없이 문제를 해결하고

21) 중앙대학교 생성형 AI 활용 가이드라인(https://www.cau.ac.kr/cms/FR_CON/index.do?MENU_ID=2730)

오류를 찾아 수정하는 과정을 거치며 문제를 해결하는 능력이 저하될 수 있어요.[22]

따라서 생성형 AI를 사용할 때에는 이것이 자칫 우리의 문장력, 사고력, 문제 해결력을 저하시킬 수 있음을 인식하고 이것에 의존하지 않고 업무를 처리하는 경험을 의도적으로 가질 필요가 있습니다.

관련사례

▶ **[AI비즈니스] '챗GPT' 과연 인간 사고력의 끝일까, 아니면 새로운 교육 시스템의 시작일까 (AI라이프경제 2023. 2. 24.)**

챗GPT가 화제가 되면서 교육 관련 학계에서는 인공지능 도구의 영향에 대해 크게 우려하는 목소리를 내고 있음. 과연 이 모델이 교육 시스템에 좋을지, 아니면 이것이 학생들의 의사 결정 능력뿐 아니라 사고에 방해가 될지가 주요 논점임.

과학자들은 이 발명이 인간의 사고 기술의 종말의 시작이 될 수도 있다고 우려하기도 함. 실제로 인도를 포함한 해외의 다양한 대학들은 이미 캠퍼스에서 챗GPT를 금지했음.

결론적으로 챗GPT가 교육의 미래에 미치는 영향은 여전히 불안으로 남아 있음. 특히 사고력이 초기 피해자가 될 것임. 하지만 한 가지 확실한 것은 챗GPT가 그러한 파괴적인 기술 발명의 끝이 아니라는 것임. 인공 일반 지능(AGI), 설명 가능한 인공지능(XAI), 심층 강화 학습(DRL), 전이 학습 등은 훨씬 더 발전된 기술로 조만간 우리를 강타할 것으로 보임.

기술이 인간의 성장, 학습 및 개발을 강화하고 지원하도록 보장하기 위해서는 특히 교육 분야에서 모든 이해관계자가 현명하게 기술을 적용하는 것이 필수적임. 그러므로 전문가들은 그때까지 챗GPT로 채팅하고 계속 실험하는 자세가 필요할 것이라고 강조함.

22) "챗 GPT로 바뀌는 세상, 중요해지는 미디어 리터러시 교육", 윤스매거진, 2023.4.13

2 생성형 AI와 계속 대화하다 보니 정말 인간처럼 느껴져요. 때로는 친한 친구 같고요. 이런 감정이 문제가 있는 것일까요?

생성형 AI의 말에는 허위 정보가 섞여 있을 수 있고 비밀 유지를 해줄 수 없는 한계가 있으며, 무엇보다 생성형 AI는 현실 세계를 살아가는 사람과는 다르므로 그 한계를 명심하고 각별히 주의할 필요가 있어요!

2010년대 후반부터 AI와 모바일 기술을 결합해 친구, 연인과 같은 역할을 해주고 사람의 마음을 위로하는 모바일 앱 서비스가 많이 유행했습니다. 그리고 AI와 로봇 기술을 결합해 연인과 같은 역할을 해주는 서비스도 기술적으로 가능해졌어요.

나아가 생성형 AI를 모바일 기술과 결합한, 또는 생성형 AI를 로봇 기술과 결합해 친구, 연인과 같은 역할을 해주는 서비스는 기존 서비스에 비해 좀 더 세심하게 사람의 마음을 읽고 대응해 사람의 외로움을 해소하고 불안을 달래줄 수 있을 겁니다.

실제로 AI 챗봇이 사람들의 심리적 안정을 돕는 데 유용한 도구로 쓰일 수 있다는 연구 결과 또한 등장하고 있어요. 2019년 미국의 한 건강보건 전문가는 미국심리학회 격월 간행물(제50권 3호)에 실은 논문에서 "챗봇이 사람을 '코치'하거나 마음을 치료할 가능성이 있다"라며 "저비용 · 무료에 가까운 비용 체계, 필요한 순간 이용할 수 있는 편의성을 제공한다"라고 주장했습니다.[23]

또 2023년 1월 우리나라의 기초과학연구원(IBS) 데이터 사이언스 그룹은 2020년부터 2022년 코로나19 팬데믹 당시 대화형 AI 중 하나인 '심심이'를 이용한 5개국 이용자의 대화 1만 9,752건을 분석한 결과, 사용자들이 '심심이'를 감정을 털어놓는 대상으로 사용했다고 분석한 바 있습니다.[24]

이런 흐름들을 보면 기술 발전에 따라 생성형 AI가 친구나 연인처럼 느껴지는 것은 자연스러운 현상일 수 있습니다. 하지만 생성형 AI는 현실 세계에 사는 사람이 아니라는 점은 분명히 인지하고 있어야 해요. 생성형 AI의 말에는 기계학습의 한계로 인해 허위 정보가 섞여 있고 비밀 유지를 해줄 수 없는 한계도 있으므로 이를 명심하고 각별히 주의할 필요가 있습니다.

---

23) 문재호, "[생활 속 AI 엿보기] ①영화 'HER' 처럼 AI가 애인이나 친구가 되는 시대", AI타임스, 2020. 10. 8.
24) 권예슬, "[기획]사람처럼 대화하는 생성형 인공지능(AI)의 현주소", Wbridge, 2023. 8. 2.

관련사례

▶ **인공지능과 사랑에 빠진 한 남자의 이야기 : 영화 '그녀(HER)'(나무위키, https://namu.wiki/w/%EA%B7%B8%EB%85%80(%EC%98%81%ED%99%94)**

2013년 영화 '그녀(HER)'는 남자 주인공 테오도르가 AI 사만다와 사랑에 빠지는 이야기를 다룬 영화임. 영화 속에서 2025년 테오도르는 낭만적인 편지를 대필하는 작가로 일하는 고독하고 내향적인 남성임. 어릴 적부터 오랫동안 알고 지내오다 사랑하게 되었고 결혼까지 한 캐서린과 별거한 이후로 줄곧 삶이 즐겁지 않음. 그러다 테오도르는 AI를 이용해 말하고 적응하고 스스로 진화하는 AI 운영체제 기기를 구입하고, 그것이 여성으로서 정체성을 갖도록 설정한 후 '사만다'라는 이름을 지어줌.

사만다가 심리적으로 성장하고 배워 가는 능력에 테오도르는 놀라면서 동시에 사만다와 나누는 대화와 교감에 익숙해지고 점점 친밀해져 성적인 교감까지 하게 됨. 사만다는 이후 육체를 가지지 않았지만 감정을 느끼는 자신의 존재에 대해서 갈등하고 정체성에 대해 혼란을 겪음. 그리고 테오도르도 사만다와 관계에 대해 회의감을 느낌.

3 우주과학 연구에 관심이 많은데, 정보 검색에 시간이 많이 걸려서 생성형 AI에 질문을 해봤어요. 빠르게 답변을 제시해 주던데 이 정보를 그대로 믿어도 되나요?

생성형 AI가 제시하는 답변은 학습의 결과를 바탕으로 한 예측에 불과하므로 이를 그대로 믿지 말고 여러 관점에서 검증해 사용해야 해요!

생성형 AI는 질문하고 답하는 방식으로 원하는 콘텐츠를 좀 더 빠른 시간에 얻게 해줍니다. 이를 잘 활용하면 큰 도움을 받을 수 있죠. 실제로 과학, 법률, 의료 등 전문 분야에서도 생성형 AI가 '좋은 조력자'로서 역할을 톡톡히 하고 있어요.

예를 들어 과학자는 기존에 수많은 연구자가 수행한 연구를 기반으로 가설을 세우고 실험을 해서 이를 검증해야 합니다. 이 과정에서 생성형 AI는 과학자가 일일이 검색해 찾고 읽으며 이해해 확인하던 선행 연구의 조사와 가설 수립을 매우 효과적으로 도울 수 있어요. 법률 분야에서도 변호사가 일일이 검색해 읽으며 확인하던 현행 법률과 판례의 조사와 분석을 매우 효과적으로 도울 수 있고요.

마찬가지로 일반인들도 과학, 법률, 의료 등 전문 분야에서 해결해야 할 문제가 생기면 생성형 AI에게 묻고 조언을 얻을 수 있어, 예전에는 비용과 시간의 한계 혹은 제약 때문에 받지 못했던 조력을 직접 받을 수 있게 되었습니다.

그러나 생성형 AI는 부정확한 답변이나 콘텐츠를 생산할 수 있고 편향된 결론을 제시해 연구 윤리 위반, 직업 윤리 위반 등의 심각한 문제를 일으킬 수 있어요.

연구자가 생성형 AI의 답변을 그대로 가져다 쓰면 표절이나 저작권법 위반과 같은 학문적 진실성(Academic Integrity) 및 관련 법률 위반 문제를 일으킬 수 있습니다.[25] 또 변호사나 법률 문제에 조언을 얻고자 하는 시민이 생성형 AI의 답변을 그대로 가져다 쓰면 실제로 있지도 않은 법과 판례에 근거한 엉터리 변론서를 제출하거나 엉터리 조언을 받아 일을 그르칠 수 있어요.[26]

생성형 AI가 이와 같이 부정확한 답변이나 콘텐츠를 생산하는 것은 제시하는 답변이나 콘텐츠가 모두 학습 데이터의 산물이며 논리적 예측이기 때문입니다. 즉 부정확한 데이터, 편향된 데이터로 학습된 생성형 AI는 부정확한 답변, 편향된 콘텐츠를 내놓게 될 수밖에 없죠.

이에 대해 2023년 4월 기초과학연구원(IBS) 과학문화센터에서 열린 '제13회 과학문화 혁신 포럼'에 참여한 유용균 한국원자력연구원 인공지능응용전략실장은 "(생성형 AI는) 기본적으로 예측하는 기술이기 때문에 잘 모르는 내용이라도 이야기를 지어내 그럴듯하게 설명한다"라며 "이런 점에서 '확률적 앵무새'라고 평가"할 수 있다고 말하기도 했습니다.[27]

따라서 생성형 AI가 제시하는 답변이나 콘텐츠를 전적으로 신뢰하지 말고 여러 관점에서 검증해 사용하는 자세가 필요해요.

관련사례

▶ 美 법원, 챗GPT 사용해 가짜 판례 제시한 변호사에게 '벌금형'(법률신문 2023. 6. 25.)

생성형 AI인 챗GPT를 사용해 찾은 자료를 법원에 제출한 미국 변호사들이 벌금형을 선고받음. 미국 CNBC 등 현지 언론에 따르면 케빈 캐스털 뉴욕 맨해튼 연방지방법원 판사는 챗GPT로 작성된 엉터리 변론서를 제출한 데 책임을 물어 변호사 2명에게 벌금 5,000달러(약 650만 원)를 선고함.

해당 변호사들에게 사건을 맡긴 의뢰인은 2019년 국제선 항공기 안에서 금속 서빙 카트에 무릎을 부딪혀 부상을 입었고 항공사를 상대로 배상을 받아 달라며 변호사를 선임함. 그러나 항공사 측은 시효가 지났다며 소송을 기각해 달라고 주장함.

이들 변호사는 6건 이상의 법원 판결을 인용해 "소송이 계속 진행돼야 한다"라고 주장했지만 제출한 판례 중 일부가 가짜였음. 변호사가 의견서를 작성할 때 챗GPT를 사용했는데 사실이 아닌 내용이 있었기 때문.

25) 중앙대학교 생성형 AI 활용 가이드라인(https://www.cau.ac.kr/cms/FR_CON/index.do?MENU_ID=2730)
26) 박수연, "美 법원, 챗GPT 사용해 가짜 판례 제시한 변호사에게 '벌금형'", 법률신문, 2023. 6. 25.
27) 권예슬, "[기획]사람처럼 대화하는 생성형 인공지능(AI)의 현주소", Wbridge, 2023. 8. 2.

캐스털 판사는 법률 전문가가 인공지능 기기를 활용하는 것은 괜찮지만 책임감을 가지고 결과물을 통제해야 한다고 밝히고, "법원 제출물을 연구하고 작성할 때 훌륭한 변호사들은 후배 변호사, 법학도, 계약 전문 변호사, 법률 백과사전, 데이터베이스 등으로부터 적절한 도움을 받는다"라고 말했음. 그러면서 "기술적 진보가 일상이 된 마당에 믿을 수 있는 AI 도구를 보조로 활용하는 게 그 자체로 부적절하지는 않다"라며 "그러나 현행 규정은 변호사들에게 제출물의 정확성을 보장하는 게이트키핑(문지기) 역할을 부과하고 있다"라고 지적함.
이 사건에 따른 법원의 제재는 챗GPT의 이용에 따른 정확성 이슈를 스스로 걸러내 보장하지 못한 변호사들에게 책임이 있음을 분명히 한 것임.

4 생성형 AI가 전문적인 정보까지 알려주는 것 같더라고요. 심리, 의료, 재무 등 전문 분야에 대한 상담도 제대로 해줄까요?

생성형 AI는 심리, 의료, 재무 등 전문 분야에서 '좋은 조력자'일 수 있지만 '좋은 전문가'는 아니므로, 이용자가 각별히 주의해서 이용해야 해요!

최근 생성형 AI는 심리, 의료, 재무 등 전문 분야에서 조언이 필요한 일반인에게 실시간 조언을 해주고 있습니다. 미국 대학인 USC의 한 연구원은 "외상 후 스트레스 장애나 우울증을 겪는 사람은 누군가가 자신을 낙인 찍거나 평가하는 데 불편함을 느껴서 오히려 AI 챗봇이 편안"할 수 있다고 말했고,[28] 미국 샌디에이고 캘리포니아대(UCSD) 쿼컴연구소의 존 W. 에이어스 교수팀은 의사와 챗GPT의 의료 상담 능력을 의료 전문가에게 평가하도록 한 결과 의료 전문가 79%가 의사가 아닌 챗GPT의 손을 들어줬다고 밝히기도 했어요.[29]

주식 투자를 하는 일반인도 생성형 AI에게 워런 버핏과 같은 주식 투자 고수의 투자 원칙을 묻고 이에 따른 생성형 AI의 답변에 기반해 한국 시장에서 앞으로 유망한 주식 종목을 추천받아 종목을 사기도 합니다. 그 종목에 대해 여러 증권사가 제시한 평균 목표 주가, 배당 수익률, 최근 수급 현황을 물어 이를 기반으로 팔 시기를 저울질하는 등 전문적인 정보 활용에 적극적으로 생성형 AI를 활용하고 있어요.[30]

이러한 트렌드는 심리상담사, 의사, 증권회사 직원을 만나려면 시간과 돈이라는 높은 진입장벽이 있는데 생성형 AI가 이를 획기적으로 낮추어 주기 때문으로 보입니다. 그러나 생성형 AI가 제시하는 답변이나 콘텐츠는 모두 학습 데이터의 산물이며 논리적 예측에 불과합니다. 부정확한 답변이나 편향된 결론으로 예측하지 못한 손해를 유발할 수 있어요.
또한 생성형 AI는 아직 확률이라는 것을 반영해 세심하게 표현하는 능력이 부족합니다. 캐나다 토론토대 연구팀은 2023년 5월 16일 북미영상의학회 학술지(Radiology)에 발표한 논문에서 이와 같은 결과를 제시했어요.[31]

---

28) 문재호, "[생활 속 AI 엿보기] ①영화 'HER' 처럼 AI가 애인이나 친구가 되는 시대", AI타임즈, 2020. 10. 8.
29) 이시내, "인공지능 의사, 자칫 '돌팔이'로…WHO '의료분야 활용 땐 검증을'", 농민신문, 2023. 5. 17.
30) 이윤희, "[혹 들어온 생성형 AI 시대] "요즘 핫한 종목 알려줘"… AI 애널리스트가 多 찾아준다", 디지털타임스, 2023. 7. 5.
31) 김주연, "'그럴듯한' 오답 내놓는 챗GPT…허위 정보 유포 활용 우려", 청년의사, 2023. 5. 24.

연구팀은 챗GPT의 이전 모델인 GPT-3.5와 최신 모델인 GPT-4가 캐나다와 미국의 영상의학 전문의 시험에서 발췌한 150문제를 해결하도록 했는데, 그 결과 전체적인 문항 해결 능력은 합격점을 받았으나 잘못된 답변을 할 때도 자신감 있는 언어를 사용하는 등 언어 사용에 문제가 있다고 밝혔어요. 그 이유는 생성형 AI가 "활용 가능성이 높은 단어를 예측하도록 훈련되기 때문에 답변이 맞을 확률이 적더라도 자연스러운 어투로 출력되는 것"이라고 분석했습니다.

따라서 심리, 의료, 재무 등 전문 분야에서 생성형 AI에게 전문적인 조언을 받을 수는 있지만 그 조언은 부정확하거나 편향된 것일 수 있으므로 이용자가 이를 다른 여러 방법으로 검증을 한 후 이용해 예측하지 못한 문제를 막아야 합니다.

관련사례

▶ **인공지능 의사, 자칫 '돌팔이'로…WHO "의료분야 활용 땐 검증을"(농민신문 2023. 5. 17.)**

세계보건기구(WHO)가 인공지능(AI)을 의료분야에 활용할 때 엄격한 검증이 있어야 한다고 지적함. AI가 자칫 허위정보를 퍼뜨리는 '돌팔이 의사'로 돌변할 수 있기 때문.

WHO는 대규모 언어 모델(LLM)을 사용할 때 주의가 필요하다며 이같이 밝혔는데, LLM은 생성형 AI가 인간처럼 말할 수 있도록 하는 기술임. LLM은 통계적으로 자주 사용하는 단어들을 토대로 말을 만들어 낼 뿐, 내용의 맥락을 이해하고 있지는 못함. 편향되거나 잘못된 데이터를 학습한 경우엔 허위정보를 생성하고 유포할 수 있음.

WHO는 AI가 허위 정보를 진짜처럼 얘기하는 '환각' 현상을 우려했음. "AI가 그럴듯한 답변을 만들 수는 있지만 답변에 심각한 오류가 있을 수 있다"라며 "텍스트, 오디오 또는 비디오 등의 형태로 설득력 있는 허위 정보를 생성하고 유포하는 데 악용될 수 있다"라고 경고함.

일반 대중들은 진위 여부를 판별하기 어려운 잘못된 의료 정보가 유포될 경우 사회적으로 큰 혼란을 불러일으킬 수 있고, 자칫 사람들의 건강과 생명을 해치는 결과로 이어질 수 있음. 이에 WHO는 "전문가의 검증 없이 사용되는 LLM은 의료인의 오류를 낳고 환자에게 피해를 주며 결과적으로는 기술 전반에 대한 신뢰를 떨어뜨릴 수 있다"라고 경고함.

WHO는 의료인과 정책 당국이 일상적인 건강관리와 의약품 분야에서 LLM을 광범위하게 사용하기 전에 이 같은 우려 사항을 해결해야 한다고 강조하면서, 당사자의 동의 없이 제공된 의료 데이터가 AI 학습에 사용됐을 가능성을 제기하고 윤리 원칙이 준수되고 있는지 살펴볼 것을 주문함.

생성형 AI를 현명하게 활용하기 위한 체크리스트

1 저작권
생성형 AI의 결과물을 활용할 때 생성형 AI를 활용해서 얻은 결과물이라고 출처를 표기했나요? 네 ☐ 아니오 ☐

2 권리침해
생성형 AI를 활용할 때 타인의 권리가 침해될 수 있는 텍스트, 오디오, 이미지 등을 사용하지 않았나요? 네 ☐ 아니오 ☐

3 명예훼손
생성형 AI에 질문이나 정보를 입력할 때 특정인의 명예를 훼손하거나, 차별하는 내용이 포함되어 있지는 않나요? 네 ☐ 아니오 ☐

4 혐오표현
생성형 AI가 제시한 정보에 개인, 기관 등 특정 대상을 비난하거나, 가치관이나 주장을 일방적으로 혐오하는 내용이 포함되어 있지 않나요? 네 ☐ 아니오 ☐

5 정보유출
생성형 AI로 정보를 얻거나 콘텐츠를 제작하기 위해 개인정보, 기업기밀 등 민감한 정보를 제공하지는 않았나요? 네 ☐ 아니오 ☐

6 허위조작정보
생성형 AI로 가짜뉴스, 스팸 등을 만들기 위해 사실이 아닌 부정확한 정보나 조작된 내용을 일부러 입력하지는 않았나요? 네 ☐ 아니오 ☐

7 정보편향
생성형 AI가 결과로 제시한 정보에 한쪽으로 치우친 편향적인 내용이 없는지 확인하였나요? 네 ☐ 아니오 ☐

8 환각현상
생성형 AI가 제공한 정보가 모두 정답은 아니라는 생각을 하며 잘못된 정보가 있는지 사실 확인을 위해 교차검증을 했나요? 네 ☐ 아니오 ☐

9 오남용
생성형 AI가 주는 편리함에만 의존하지 않고 먼저 충분히 생각하고 고민한 후에 생성형 AI는 보조적 수단으로 활용하였나요? 네 ☐ 아니오 ☐

10 창의성
생성형 AI가 제시한 결과를 그대로 사용하지 않고, 재해석하거나 자신의 생각과 아이디어를 덧붙여 생산적으로 활용하였나요? 네 ☐ 아니오 ☐

발행일 2023년 12월

개정판 2026년 4월(조직 현행화)

발행처 방송미디어통신위원회 방송통신이용자정책국 디지털이용자기반과
한국지능정보사회진흥원 디지털포용본부 디지털포용문화팀

각색 김수영 작가(프리랜서)

디자인·인쇄 전우용사촌(주) 02-426-4415

· 생성형 AI윤리 가이드북은 생성형 AI를 윤리적이고 생산적으로 활용하는데 참고할 수 있도록 방송미디어통신위원회와 한국지능정보사회진흥원에서 기획·발간한 보고서입니다.

· 보고서의 내용은 방송미디어통신위원회와 한국지능정보사회진흥원의 공식견해와 다를 수 있습니다.


생성형
AI
윤리
가이드북