

저작권 기술 트렌드



COPYRIGHT TECH TREND



한국저작권위원회
KOREA COPYRIGHT COMMISSION

영화 산업의 스토리 표절 논쟁과 AI 서사 분석 기술의 부상

뉴스 브리프

최근 천만 관객을 돌파한 영화 '왕과 사는 남자'가 표절 논란에 휩싸이면서, 창작과 차용의 경계를 둘러싼 논쟁이 재점화되었다. 이와 관련해 AI를 활용한 새로운 기술적 접근이 주목받고 있다. AI 서사 분석 기술은 시나리오의 구조, 인물, 주제 등을 비교·분석해 유사도를 측정한다. 또한 사용자가 직접 줄거리를 입력하면 유사한 작품을 검색해주는 챗봇 형태의 표절 탐지 도구도 연구되고 있다. 본 보고서는 이러한 AI 기반 기술의 작동 원리와 잠재력을 살펴보고, 신뢰성 강화를 위한 기술 표준화 방안과 산업 생태계에 미칠 영향을 분석한다. 나아가 데이터 기반 논의의 가능성을 탐색하고, 창작자의 권리 보호와 산업 발전에 기여할 시사점을 제시한다.

뉴스 플러스

I. 서론 : 반복되는 저작권 분쟁과 AI 기반 대안의 모색

- 천만 영화 '왕과 사는 남자'가 던진 질문

2026년 상반기, 영화 '왕과 사는 남자'는 개봉 67일 만에 누적 관객 1,633만 명을 돌파(2026년 4월 11일 기준)하며 한국 영화 역대 흥행 순위 최상위권에 이름을 올렸다.¹⁾ 그러나 성공 뒤편에서는 작품의 원작성에 관한 쟁점이 부상하였다. 2019년 작고한 연극배우 엄 씨의 유족이, 고인이 2000년대에 집필한 드라마 시나리오 초고 '엄홍도'와 영화 사이에 다수의 유사점이 존재한다고 제작사에 내용증명을 보낸 것이다.²⁾ 유족 측은 역사적 사실인 단종의 영월 유배를 배경으로, 영화 속 핵심 설정과 사건 전개 방식이 고인의 시나리오와 상당 부분 겹친다고 주장한다.

1) 조지영, "'왕과 사는 남자', 1600만 관객 돌파 '극한직업' 잡고 역대 2위 갈까?', 조선일보, 2026.04.12, <https://www.chosun.com/entertainments/movie/2026/04/12/MU2TIMRWME4TEMBQMFSTQMRSMU/>

2) 서지영, "'7가지 유사점 있다' 천만 돌풍 '왕사남', 출연 표절 의혹...제작사 '사실무근'", 아시아경제, 2026.03.10., <https://www.asiae.co.kr/article/2026031008523154431>

유족이 제기한 유사성은 구체적이다. 유배지에서 음식을 거부하던 단종이 특정 음식을 통해 마음을 여는 장면의 흐름, 역사 속 여러 인물을 '매화'라는 단일 캐릭터로 압축한 설정, 주인공 엄홍도의 자녀를 외아들로 각색한 점 등 총 7가지 유사점이 제기되었다.³⁾ 이러한 주장은 단순한 아이디어나 소재의 차용을 넘어, 인물 구성과 서사를 풀어나가는 구체적인 표현 방식에 대한 문제 제기라는 점에서 논란의 무게를 더한다. 결국 이 사건은 역사적 사실에 기반한 창작물에서 어디부터 어디까지를 아이디어와 독창적 표현으로 보아야 하는지에 대한 근본적인 질문을 던지고 있다.

반면 제작사는 표절 의혹을 전면 부인하며 강경한 입장을 고수하고 있다. 제작사 온다웍스는 공식 입장문을 통해 해당 영화가 역사적 사실을 기반으로 한 순수 창작물이며, 창작 과정에서 해당 작품을 접한 경로나 인과성이 없고 타 저작물을 표절한 사실이 전혀 없다고 밝혔다.⁴⁾ 또한 역사적 인물을 다루는 작품의 특성상 일부 유사한 설정이 존재할 수 있으나, 이는 장르적 특성에 기인한 것일 뿐 저작권 침해와는 무관하다는 점을 분명히 했다. 양측의 주장이 첨예하게 대립하는 이 상황은 장기적인 법적 공방을 예고하며, 영화 산업 전반에 적지 않은 과제를 남기고 있다.

• 저작권 침해 판단의 어려움

'왕과 사는 남자'를 둘러싼 논쟁은 온라인 공론장으로 빠르게 확산하며 다양한 여론을 형성하고 있다. 일부 대중은 "역사 소재 작품이면 비슷한 장면이 나올 수도 있는 것 아니냐"며 신중한 태도를 보이는 반면, "원작이 있다면 이름을 올려주는 게 맞다"는 의견도 적지 않다.⁵⁾ 이처럼 엇갈리는 반응은 저작권 침해 판단이 명확한 증거보다 평가자의 인상과 해석에 좌우되는 경우가 많음을 보여준다.

이러한 딜레마는 법적 판단 과정에서도 동일하게 나타난다. 전문가들은 저작권 침해가 법적으로 인정되기 위해서는 두 가지 요건이 충족되어야 한다고 설명한다. 하나는 두 저작물의 구체적인 표현 방식에서의 실질적 유사성이 인정되어야 한다는 것이고, 다른 하나는 원작을 접했을 가능성을 의미하는 의거성(依據性)이 입증되어야 한다는 점이다.⁶⁾ 그러나 구체적인 표현 방식의 유사도를 객관적으로 측정하기란 쉽지 않아 결국 재판부의 해석에 의존할 수밖에 없다.

이러한 간극은 법적 결과와 무관하게 창작자에게 실질적인 위협으로 이어질 수 있다. 영화평론가 정지욱은 "법적 판단과 대중의 인식 사이에는 항상 간극이 존재한다"고 지적하며, 핵심 설정이나 연출 방식이 지나치게 닮아 있을 경우 관객은 이를 표절로 받아들일 수 있다고 설명한다.⁷⁾ 법적 책임을 넘어서는 사회적·윤리적 비판과 상업적 실패라는 위협이 창작자에게 동시에 가해질 수 있는 것이다. 신뢰할 만한 판단 기준의 부재는 논쟁을 반복시키고, 창작 생태계 전반의 신뢰를 저해하는 요인으로 작용한다.

3) 서지영, "7가지 유사점 있다" 천만 돌풍 '왕사남', 들연 표절 의혹... 제작사 '사실무근', 아시아경제, 2026.03.10., <https://www.asiae.co.kr/article/2026031008523154431>

4) 김도연, "표절이면 이름 넣어달라" vs '사실무근'...1200만 앞둔 '왕과 사는 남자' 표절 논란, 서울경제, 2026.03.10., <https://www.sedaily.com/article/20017405>

5) 한현정, "1200만 눈앞 '왕사남' 표절 논란...유족 문제 제기, 제작사는 부인", 스타투데이, 2026.03.10., <https://v.daum.net/v/06W71m5TGJ>

6) 강세준, "천만관객 '왕사남' 표절 논란..무한반복 스크린 흑역사", 법치뉴스, 2026.03.10., <https://www.bubb.co.kr/news/articleView.html?idxno=4141>

7) 강세준, "천만관객 '왕사남' 표절 논란..무한반복 스크린 흑역사", 법치뉴스, 2026.03.10., <https://www.bubb.co.kr/news/articleView.html?idxno=4141>

- '광해'에서 '터미네이터'까지, 반복되는 창작과 차용의 경계

영화 산업에서 스토리의 유사성을 둘러싼 저작권 분쟁은 새로운 현상이 아니다. 과거 천만 관객을 동원한 영화 '광해, 왕이 된 남자' 역시 미국 영화 '데이브(Dave)'와의 유사성 논란을 겪었으며, '최종병기 활'은 개봉 당시 할리우드 영화 '아포칼립토(Apocalypto)'와 추격 서사 구조 및 연출 방식이 유사하다는 지적을 받았다. 영화 '암살'은 소설 '코리안 메모리즈'와의 표절 분쟁에서 소송까지 이어졌으나, 법원은 역사적 사실과 장르적 특성에서 비롯된 유사성은 창작적 표현의 침해로 볼 수 없다고 판단했다.⁸⁾ 이처럼 법적 결론이 엇갈리는 사례들은 창작과 모방, 영감과 도용의 경계가 얼마나 모호한지를 잘 보여준다.

이러한 현상은 한국에만 국한되지 않는다. 할리우드 역시 오랜 기간 아이디어 도용과 시나리오 표절 문제로 수많은 법적 분쟁을 겪어왔다. 작가 아트 부크월드(Art Buchwald)가 자신의 아이디어가 무단으로 사용되었다며 제작사를 상대로 소송을 제기한 영화 '커밍 투 아메리카(Coming to America)' 사건은 아이디어 제공자의 권리에 관한 중요한 판례를 남겼다. SF 영화의 고전인 '터미네이터(The Terminator)'와 '에이리언(Alien)' 역시 각각 TV 시리즈 및 소설과의 유사성 문제로 분쟁을 겪었으며, 제작사가 합의금을 지급하고 원작자에 대한 감사를 표기하는 방식으로 사건을 마무리한 바 있다.⁹⁾

이처럼 국내외를 막론하고 반복되는 표절 논쟁은 영화 산업의 구조적 특성과 연관되어 있다. 상업 영화는 막대한 제작비가 투입되는 만큼 흥행 실패의 리스크를 최소화해야 한다는 압박이 크고, 그 과정에서 이미 검증된 서사 구조나 익숙한 이야기 패턴에 기대는 경우가 생겨난다. SNS의 발달로 작품 간 유사성에 대한 대중의 비교와 검증이 빠르게 확산되면서, 논란이 단기간에 공론화되는 환경도 형성되었다.¹⁰⁾ 이러한 흐름 속에서 보다 체계적이고 데이터에 기반한 검증 방식에 대한 논의가 부상하고 있다.

- AI 기반 서사 분석, 새로운 해법의 가능성

창작자의 양심과 법원의 사후적 판단에만 의존하는 기존 방식은 디지털 시대의 복잡하고 다양한 창작 환경에 효과적으로 대응하기에 한계가 있다. 분쟁을 사전에 예방하고, 창작자가 안심하고 창작 활동에 몰입할 수 있는 환경을 조성하기 위한 새로운 접근이 필요한 시점이다.

이러한 배경 속에서 AI 기술이 문제 해결의 실마리를 제공할 가능성을 보여주고 있다. 대규모 언어 모델(Large Language Model, 이하 LLM)*과 자연어 처리(Natural Language Processing, 이하 NLP)**의 발전은 과거 인간의 고유 영역으로 여겨졌던 서사 구조 분석을 가능하게 하고, 작품 간의 의미적·구조적 유사성을 데이터 기반으로 측정하는 길을 열고 있다. 이는 실질적 유사성을 정량적 수치와 구체적 근거로 제시하려는 시도로, 정성적 해석에 머물던 저작권 논쟁에 데이터 기반 판단 기준을 도입할 가능성을 시사한다.

8) 강세준, "천만관객 '왕사남' 표절 논란..무한반복 스크린 흑역사", 법치뉴스, 2026.03.10., <https://www.bubb.co.kr/news/articleView.html?idxno=4141>

9) 강세준, "천만관객 '왕사남' 표절 논란..무한반복 스크린 흑역사", 법치뉴스, 2026.03.10., <https://www.bubb.co.kr/news/articleView.html?idxno=4141>

10) 강세준, "천만관객 '왕사남' 표절 논란..무한반복 스크린 흑역사", 법치뉴스, 2026.03.10., <https://www.bubb.co.kr/news/articleView.html?idxno=4141>

본 보고서는 이러한 맥락에서 AI 기반 서사 분석 기술이 영화 산업의 저작권 분쟁에 대한 기술적 대안이 될 수 있는지 탐색하고자 한다. 이를 위해 먼저 AI가 시나리오의 서사 요소를 분석하고 유사도를 측정하는 기술적 원리와 한계를 살펴보고, 창작자나 일반 사용자를 위한 대중적 표절 탐지 도구의 가능성까지 함께 검토한다. 나아가 해당 기술이 영화 산업 생태계 전반에 미칠 파급 효과와 향후 전망을 조망한다.

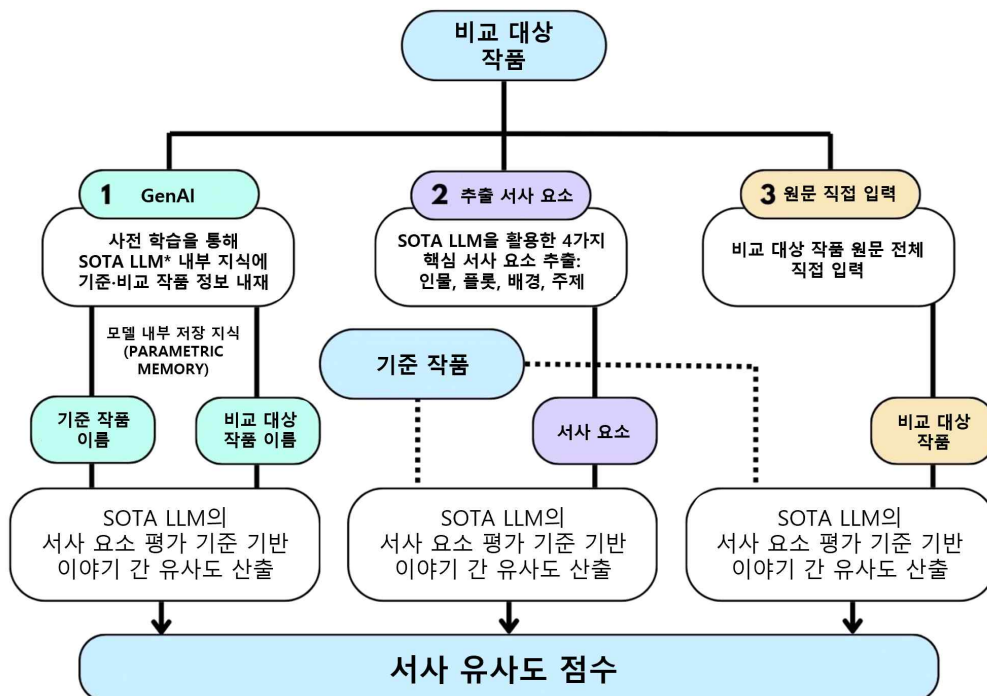
* 대규모 언어 모델(Large Language Model, LLM): 수십억 개 이상의 방대한 텍스트 데이터를 학습하여, 인간과 유사한 언어 이해 및 산출 능력을 갖춘 AI 모델
** 자연어 처리(Natural Language Processing, NLP): 인간이 사용하는 언어를 컴퓨터가 이해하고 분석·생성할 수 있도록 하는 인공지능 기술 분야

II. 본론 1: AI 서사 분석 기술의 원리와 한계

• AI 서사 유사도 분석의 개념과 접근 방식

AI 서사 유사도 분석은 인간의 직관과 정성적 해석에 의존하던 서사 분석을 데이터 기반의 정량적 평가로 전환하려는 시도이다. 이 기술의 핵심은 LLM이 방대한 텍스트 데이터를 학습하며 습득한 문맥 이해 능력을 활용하여, 두 시나리오 사이에 존재하는 구조적·의미적 유사성을 수치로 측정하는 데 있다. 키워드의 출현 빈도를 계산하는 기존의 텍스트 마이닝(text mining)과 달리, 이야기의 근간을 이루는 서사적 요소들을 종합적으로 비교·분석하는 접근 방식이다.

[그림 1] AI 서사 유사도 비교의 세 가지 방법론



출처: Jon Chun, "AIStorySimilarity: Quantifying Story Similarity Using Narrative for Search, IP Infringement, and Guided Creativity", Association for Computational Linguistics, 2024.11, <https://aclanthology.org/2024.conll-1.13.pdf>

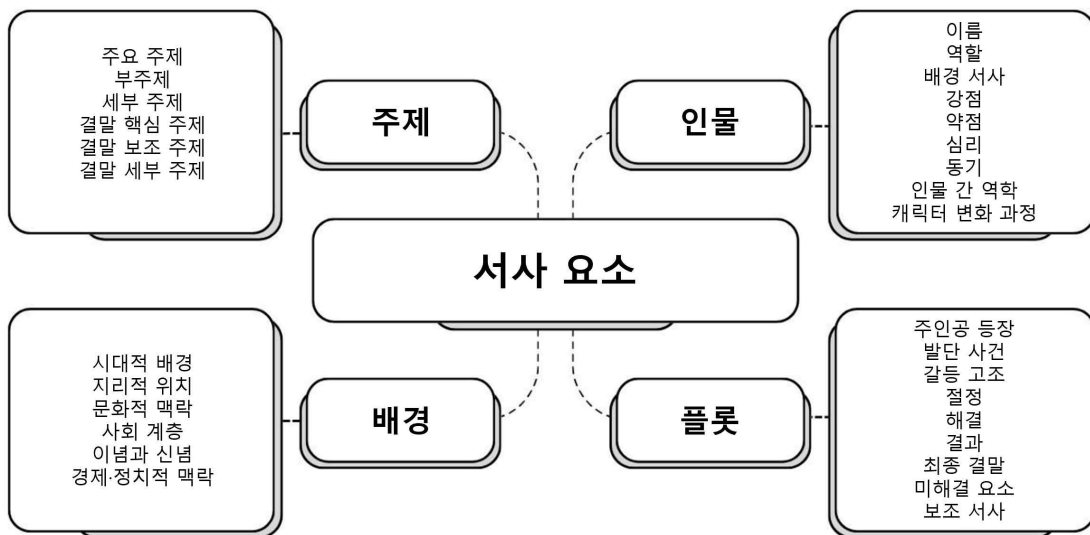
* SOTA LLM(State-Of-The-Art LLM): 특정 시점에서 가장 높은 성능을 보이는 대규모 언어 모델(LLM)을 통칭하는 표현. 본 논문에서는 OpenAI의 GPT-3.5-turbo가 사용됨

- 핵심 서사 요소 추출: 인물, 플롯, 배경, 주제

[그림 1]의 2번 경로에서 설명하듯 AI는 시나리오 텍스트를 분석하여 네 가지 핵심 서사 요소인 인물(characters), 플롯(plot), 배경(setting), 주제(themes)를 추출한다. 이 과정은 각 요소의 정보를 식별하도록 설계된 프롬프트로 수행된다. 예를 들어 인물 요소에 대해서는 "주요 인물들의 성격과 목표, 관계를 요약하라"와 같은 방식의 프롬프트가 설계될 수 있으며, "이야기의 핵심 사건들을 순서대로 나열하고 인과관계를 설명하라"는 식의 명령을 통해 플롯의 뼈대를 파악한다. 또한 "이야기의 핵심 메시지는 무엇인가?" 또는 "이야기의 시간적·공간적 배경은 어떠한가?"와 같은 질문 형태의 프롬프트를 통해 주제와 배경 정보를 데이터로 변환한다.

이처럼 각 서사 요소를 개별적으로 분석하는 방식은 중요한 의미를 가진다. 논문에 따르면 네 가지 요소는 분석의 복잡도 측면에서 차이가 있으며, 배경·플롯·인물·주제 순으로 요구되는 해석의 깊이가 깊어진다.¹¹⁾ 이를 개별적으로 측정함으로써 두 작품이 어느 서사 요소에서 얼마나 유사한지를 구체적으로 파악할 수 있고, 표면적인 소재의 일치 여부를 넘어 이야기의 본질적인 구조와 표현 방식의 유사성을 식별하는 것이 가능해진다. 각 서사 요소의 세부 분석 항목은 [그림 2]와 같다.

[그림 2] 서사 유사도 평가표(Narrative Similarity Rubric)



출처: Jon Chun, "AIStorySimilarity: Quantifying Story Similarity Using Narrative for Search, IP Infringement, and Guided Creativity", Association for Computational Linguistics, 2024.11, <https://aclanthology.org/2024.conll-1.13.pdf>

- 유사도 정량화: 프롬프트 기반 평가와 점수화

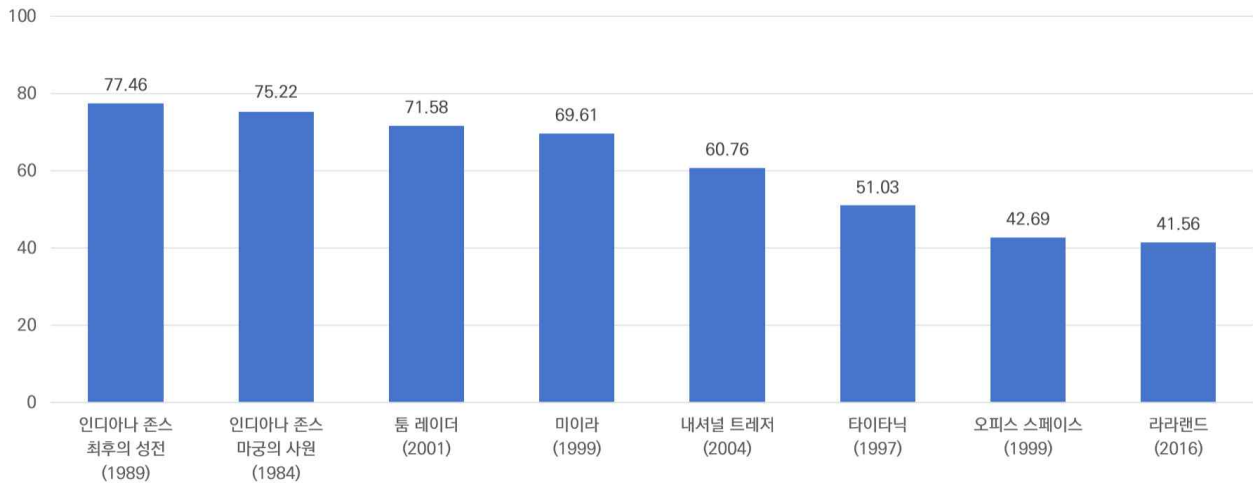
핵심 서사 요소가 추출된 후에는 두 이야기의 해당 요소들이 얼마나 유사한지를 평가하여 점수로 변환하는 정량화 단계가 이어진다. AI는 기준 작품(reference script)과 비교 대상 작품(test script)에서 각각 추출된

11) Jon Chun, "AIStorySimilarity: Quantifying Story Similarity Using Narrative for Search, IP Infringement, and Guided Creativity", Association for Computational Linguistics, 2024.11, <https://aclanthology.org/2024.conll-1.13.pdf>

서사 요소 정보를 입력받은 뒤, 특정 평가 기준에 따라 유사도를 판단하도록 설계된 평가 프롬프트를 수행한다. 이 과정의 신뢰도를 높이기 위해 논문에서는 방식에 관계없이 동일한 평가를 30회 반복하여 그 평균값을 최종 점수로 사용하는 방식을 채택했다.¹²⁾

인간 전문가 판단과의 일치율이 가장 높은 생성형 AI 기반 평가 방식(GenAI)의 실험 결과에 따르면 네 가지 서사 요소 중 플롯은 작품 간 유사도가 가장 높게 나타났는데, 이는 할리우드 영화 전반에 걸쳐 영웅의 여정(hero's journey)이라는 서사 구조가 광범위하게 공유되기 때문으로 분석된다. 반면 주제는 가장 낮은 유사도와 높은 다양성을 보였는데, 주제가 창작자의 개성이 가장 강하게 반영되는 추상적 영역이기 때문이다. 여러 영화의 서사 유사도를 분석한 결과는 아래 [그림 3]과 같다.

[그림 3] 레이더스(1981)와의 전체 유사도 비교(GenAI 방식)



출처: Jon Chun, "AIStorySimilarity: Quantifying Story Similarity Using Narrative for Search, IP Infringement, and Guided Creativity", Association for Computational Linguistics, 2024.11, <https://aclanthology.org/2024.conll-1.13.pdf>

위 그래프에서 볼 수 있듯 AI 분석은 기준 작품인 '레이더스(Raiders Of The Lost Ark)*'와 같은 인디애나 존스 시리즈에 속하는 후속편들에 대해 75점 이상의 높은 유사도 점수를 부여했다. 반면 장르와 내용이 전혀 다른 '라라랜드(La La Land)'나 '오피스 스페이스(Office Space)' 같은 작품들은 40점대의 낮은 점수를 기록하여 명확한 변별력을 보여준다. '툼 레이더(Laura Croft Tomb Raider)'나 '미이라(The Mummy)'처럼 주인공이 유물을 찾아 모험을 떠난다는 유사한 장르적 모티프를 가진 작품들에 대해서도 70점 내외의 높은 점수를 부여한다는 점도 의미있는 결과이다. 이는 AI가 장르적 관습이나 서사 원형의 유사성까지 포착할 수 있음을 보여준다. 이러한 정량적 데이터는 "두 작품이 얼마나 비슷한가?"라는 막연한 질문에 구체적인 수치와 근거를 제공하는 역할을 한다.

* 레이더스(Raiders of the Lost Ark, 1981): 스티븐 스피얼버그(Steven Spielberg) 감독의 인디애나 존스(Indiana Jones) 시리즈 1편. 고고학자 인디애나 존스가 나치로부터 성체를 지키기 위해 모험을 펼쳐는 내용의 영화로, 본 연구에서 비교 기준 작품으로 사용되었음. '인디애나 존스: 마궁의 사원 (Indiana Jones and the Temple of Doom, 1984)'과 '인디애나 존스: 최후의 성전(Indiana Jones and the Last Crusade, 1989)'은 각각 동 시리즈의 2편과 3편임

12) Jon Chun, "AIStorySimilarity: Quantifying Story Similarity Using Narrative for Search, IP Infringement, and Guided Creativity", Association for Computational Linguistics, 2024.11, <https://aclanthology.org/2024.conll-1.13.pdf>

- 기술적 난제와 신뢰성 문제: 약 70점의 미스터리

AI 서사 분석 기술은 정량적 데이터라는 장점을 지니지만, 동시에 명확한 기술적 한계와 신뢰성 문제를 안고 있다. 가장 큰 난제는 유사도 점수의 해석 문제이다. 앞선 예시에서 '미이라'가 약 70점(69.61점)의 유사도를 보였을 때, 이 점수가 표절을 의미하는지 아니면 장르적 유사성의 범위 내에 있는지를 판단할 명확한 기준이 확립되지 않았다. 즉 70점과 50점의 차이가 법적 또는 윤리적으로 어떤 의미를 갖는지에 대한 사회적 합의가 이루어지지 않은 상태이다. 이는 기술이 유사도를 측정할 수는 있지만, 표절이라는 가치 판단을 내릴 수는 없다는 근본적인 한계를 보여준다.

또 다른 문제는 AI 모델 자체의 편향성과 불투명성이다. AI는 학습한 데이터에 내재된 편향을 그대로 습득할 수 있으며, 특정 장르나 문화권의 서사에 더 익숙하게 반응할 가능성이 있다. 예를 들어 할리우드 영화 데이터로 주로 학습된 AI는 비서구권의 독특한 서사 구조를 제대로 이해하지 못하거나, 특정 클리셰(cliché)를 과도하게 유사하다고 판단할 수 있다. AI가 어떤 논리적 과정을 거쳐 특정 점수를 산출했는지 그 내부 작동 원리를 투명하게 설명하기 어려운 블랙박스(black box) 문제도 존재한다. 평가의 근거를 텍스트로 제시하긴 하지만, 그것이 AI의 모든 판단 과정을 대변하지는 못한다.

현재의 AI 서사 분석 기술은 저작권 침해 여부를 자동으로 판결하는 심판의 역할을 하기보다는, 인간 전문가의 판단을 돕는 보조 도구로서의 역할이 현실적으로 적합한 것으로 보인다. 다만 이 역할을 충실히 수행하기 위해서는 기술 자체의 신뢰성 확보가 먼저 이루어져야 의미를 갖는다. 기술의 신뢰성을 높이기 위해서는 지속적인 모델 개선과 함께 다양한 장르와 문화권의 데이터를 균형 있게 학습시키고, 그 결과를 해석할 수 있는 명확한 가이드라인을 수립하려는 논의가 병행될 필요가 있다.

II. 본론 2: 분석 신뢰성 확보를 위한 기술 조건과 활용 방식

- 신뢰성 강화를 위한 검증 절차의 표준화

본론 1에서 살펴본 AI 서사 분석 기술은 서로 다른 두 개의 이야기가 얼마나 유사한지를 수치화된 점수로 제시할 가능성을 열었다. 그러나 산출된 점수가 분석마다 달라지거나 평가자에 따라 결과가 흔들린다면 그 활용 가치는 떨어질 수밖에 없다. '왕과 사는 남자'의 사례처럼 분쟁 당사자들의 주장이 첨예하게 대립하는 상황에서, 어느 한쪽이 자신에게 유리한 AI 분석 결과를 제시한다면 상대방은 그 공정성을 신뢰하기 어렵다. "사용한 AI 모델이 편향되었다"거나 "의도적으로 특정 결과가 나오도록 프롬프트를 설계했다"는 식의 또 다른 논쟁으로 번질 여지도 있다. 기술의 중립성에 대한 의심이 쌓이면 그 유용성은 자연히 약해진다.

이러한 문제를 완화하는 데에는 분석에 사용되는 AI 모델과 절차의 일관성이 중요한 변수가 된다. 동일한 모델과 평가 프롬프트, 동일한 추론 설정에서 산출된 결과는 호출 시점이 달라져도 큰 편차 없이 재현될 수 있으며, 분석에 사용된 모델 버전과 프롬프트 구성이 함께 기록될 경우 결과의 검증 가능성도 높아진다. 반대로 사용 모델이나 프롬프트 설계가 분석마다 달라질 경우 결과 자체가 흔들리기 때문에, 이 영역은 AI 서사 분석 기술의 안정성을 좌우하는 핵심 기술 요소로 다뤄지고 있다.

아울러 일관성을 확보하기 위한 기술적 접근도 활발히 연구되고 있다. 첫째, 샘플링 기법을 통한 안정화 방식이다. 동일한 평가를 여러 차례 반복하여 평균값을 사용하면 한 번의 응답이 가진 우연성을 줄일 수 있다. 예를 들어 동일한 시나리오 쌍에 대한 유사도 측정을 30회 반복하여 평균을 구하면, 단일 측정에서 발생할 수 있는 편차를 보정할 수 있다. 둘째, 하이퍼파라미터(hyperparameter)* 설정을 통한 무작위성 조절이다. AI 모델은 추론 과정에서 일정한 무작위성을 가지는데, 이를 조절하는 설정값을 낮게 유지하면 같은 질문에 대해 매번 비슷한 응답을 얻을 수 있다. 이는 분석 결과가 호출 시점에 따라 크게 달라지는 것을 방지한다. 셋째, 구조화된 출력 형식의 활용이다. JSON(JavaScript Object Notation)**과 같은 구조화된 포맷으로 응답을 받으면, '유사도: 75점'처럼 매번 다른 문장 형식으로 출력되던 결과를 일정한 틀에 담을 수 있다.¹³⁾ 이는 결과를 자동으로 처리하고 데이터베이스에 저장하는 후속 작업의 안정성을 높인다. 이러한 기술 요소들이 결합될 때 AI 서사 분석은 일회성 참고 자료를 넘어 반복 검증이 가능한 도구로 기능할 수 있다.

* 하이퍼파라미터(hyperparameter): AI 모델이 응답을 생성하는 방식을 조절하는 설정값. 대표적으로 응답의 무작위성을 조절하는 변수(temperature)가 있으며, 값이 낮을수록 일관된 응답이, 높을수록 다양한 응답이 산출됨

** JSON(JavaScript Object Notation): 데이터를 '키(key)와 값(value)'의 쌍으로 구조화해 표현하는 형식. 자유 형식의 텍스트와 달리 항목별로 정보가 정리되어 있어 컴퓨터가 안정적으로 처리할 수 있음

• 창작자를 위한 사전 검증 시스템으로서의 활용

AI 서사 분석 기술의 가치는 사후적인 분쟁 해결에만 국한되지 않는다. 영화 산업 생태계에서 더욱 주목할 만한 활용 가능성은, 창작자들이 잠재적인 저작권 문제를 사전에 예방하고 관리할 수 있도록 돕는 사전 검증 시스템으로서의 역할이다. 실제로 영화계에서는 기획 단계에서부터 참고 작품을 검토하고 창작 이력을 관리하는 등 검증 체계를 강화해야 한다는 목소리가 꾸준히 제기되어 왔다. 표절 논란이 한 번 불거지면 법적 승패와 무관하게 창작자와 작품의 명성에 적지 않은 타격을 입히기 때문이다.

이러한 산업계의 요구에 부응하여 AI 기술은 창작자에게 자신의 아이디어가 기존 작품과 얼마나 유사한지 스스로 점검해 볼 기회를 제공한다. 예를 들어 창작자가 시놉시스나 초기 시나리오를 시스템에 입력하면, AI가 기존 영화 데이터베이스와 비교 분석하여 유사도가 높은 작품 목록과 구체적인 유사 지점을 리포트 형태로 보여주는 방식이 그 예이다.¹⁴⁾ 이러한 시스템을 통해 창작자는 의도치 않은

13) Jon Chun, "AIStorySimilarity: Quantifying Story Similarity Using Narrative for Search, IP Infringement, and Guided Creativity", Association for Computational Linguistics, 2024.11, <https://aclanthology.org/2024.conll-1.13.pdf>

14) Jothiragavan T 외 2인, "MOVIE PLAGIARISM DETECTION CHATBOT", International Journal of Research Publication and Reviews, 2024.04, <https://ijrpr.com/uploads/V5ISSUE4/IJRPR24766.pdf>

유사성을 발견하고 작품을 더욱 독창적인 방향으로 수정·보완할 수 있으며, 이는 사후 분쟁의 소지를 줄이는 효과로 이어진다. AI 기술이 창작 과정의 파트너로 기능할 때, 기술은 규제의 도구를 넘어 창의성을 촉진하고 창작자를 보호하는 지원 수단이 될 수 있다.

- 분석 결과의 활용 가능성과 기술적 한계

현재의 AI 기술은 두 텍스트 사이의 유사성을 측정할 뿐 후속 창작자가 선행 저작물을 접했는지, 즉 의거성을 직접 증명하지는 못한다는 한계를 가진다. 두 작품의 유사도가 매우 높게 측정되었다 하더라도 의거성이 인정되지 않는 경우에는 저작권 침해가 성립하지 않는다. 저작권법이 독자적인 창작의 가능성을 폭넓게 인정하기 때문이다. 결국 AI 분석 결과는 유사성 측정이라는 기술의 본래 영역에서 의미를 가지며, 그것을 어떻게 해석하고 활용할지는 인간의 판단 영역으로 남는다.

II. 본론 3: 챗봇 기반 표절 탐지 시스템과 산업적 활용

- 시나리오 피칭 및 투자 심사 과정에의 도입

이러한 한계에도 불구하고, AI 서사 분석 기술은 다양한 영역으로 그 활용 범위를 넓혀갈 수 있다. 앞서 살펴본 창작 단계의 사전 검증을 넘어, 영화 산업의 핵심 단계인 기획 및 투자 결정 과정에도 유의미한 변화를 가져올 수 있다. 영화 투자는 막대한 자본이 투입되지만 성공 여부가 불확실한 고위험 사업 영역으로, 투자사나 제작사는 잠재적 리스크를 최소화하기 위해 다각적인 검토를 수행한다. 그중에서도 개봉 이후 불거질 수 있는 저작권 분쟁은 흥행에 상당한 영향을 미칠 수 있는 위험 요인으로 꼽힌다.

이러한 맥락에서 AI 서사 분석은 투자 심사 과정을 지원하는 실사(due diligence) 도구로 자리할 수 있다. 투자 심사역이나 프로듀서가 검토 중인 시나리오를 AI 시스템에 입력하여 기존 작품들과의 유사도를 정량적으로 확인하는 방식이 그 예이다. 특정 작품과 예상보다 높은 유사도가 측정된다면 이는 해당 시나리오가 표절이라는 결론이 아니라, 사전에 법률적 검토나 수정·보완이 필요할 수 있다는 신호로 해석할 수 있다. 숙련된 전문가의 판단에 데이터 기반의 근거를 더함으로써 투자 결정의 합리성을 높이고, 잠재적인 법적 분쟁 리스크를 선제적으로 관리하는 데도 도움이 될 수 있다.

- 챗봇 형태의 표절 탐지 도구의 작동 원리

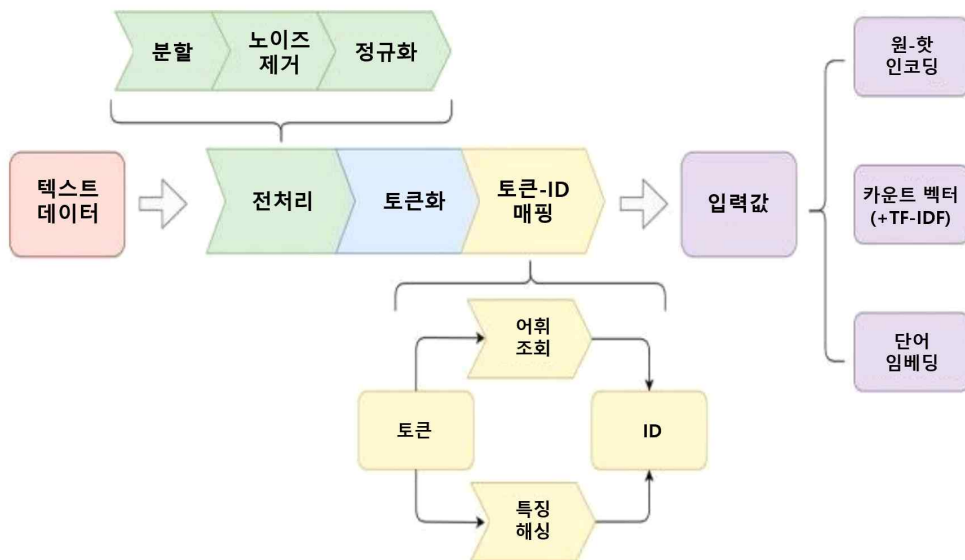
AI 서사 분석 기술의 산업적 파급력은 전문가나 기업 단위의 활용에만 머무르지 않고, 일반 창작자들도 손쉽게 접근할 수 있는 대중적 서비스로 확장될 때 더욱 커진다. 이러한 가능성을 보여주는 실제 사례로, 최근 연구에서는 사용자가 영화 시놉시스나 아이디어를 입력하면 유사한 작품을 찾아주는 챗봇(chatbot) 형태의 표절 탐지 도구가 제안되었다.¹⁵⁾ 이 도구는 사용자 친화적인 인터페이스 뒤에서 자연어 처리

15) Jothiragavan T 외 2인, "MOVIE PLAGIARISM DETECTION CHATBOT", International Journal of Research Publication and Reviews, 2024.04, <https://ijrpr.com/uploads/V5ISSUE4/IJRPR24766.pdf>

(Natural Language Processing, 이하 NLP) 과정을 수행하여, 입력된 줄거리와 데이터베이스 내 기존 작품들 사이의 의미적 거리를 정량적으로 계산한다.

이러한 도구의 작동은 크게 전처리(preprocessing), 수치 변환, 유사도 계산의 세 단계로 이루어진다. 첫 번째는 전처리 단계로, 입력된 텍스트를 분석 가능한 형태로 정제하는 작업이다. 분할(splitting)과 노이즈 제거(noise removal), 정규화(normalization)를 거쳐 텍스트를 표준화한 뒤, 토큰화(tokenization)를 통해 문장을 단어 단위로 분리한다. 이 과정에서 어간 추출(stemming)이나 표제어 추출(lemmatization) 같은 기법이 함께 적용되어, 같은 어근에서 파생된 단어들이 동일한 형태로 처리되도록 한다. 토큰화된 단어들은 어휘 조회(vocabulary lookup)나 특징 해싱(feature hashing) 같은 방식을 거쳐 고유 식별자(토큰-ID)로 매핑된다.¹⁶⁾

[그림 4] 자연어 처리(NLP)를 위한 전처리 단계



출처: Jothiragavan T 외 2인, "MOVIE PLAGIARISM DETECTION CHATBOT", International Journal of Research Publication and Reviews, 2024.04, <https://ijrpr.com/uploads/V5ISSUE4/IJRPR24766.pdf>

두 번째는 수치 변환 단계이다. 토큰-ID로 매핑된 단어들을 분석에 활용할 수 있는 수치 형태로 변환하는 작업이며, 분석 목적에 따라 여러 방식이 선택될 수 있다. 단순히 단어 출현 여부만을 표시하는 원-핫 인코딩(one-hot encoding), 단어의 출현 빈도를 반영하는 카운트 벡터(count vector), 빈도에 더해 단어의 중요도까지 가중치로 반영하는 TF-IDF*, 그리고 단어의 의미를 다차원 공간에 배치하는 단어 임베딩(word embedding)**이 대표적이다. 표면적인 어휘 일치 여부를 확인하는 데에는 TF-IDF가 효율적이지만, 줄거리의 의미적 유사성을 측정하는 작업에서는 단어 임베딩이 보다 적합한 방식이다.

16) Jothiragavan T 외 2인, "MOVIE PLAGIARISM DETECTION CHATBOT", International Journal of Research Publication and Reviews, 2024.04, <https://ijrpr.com/uploads/V5ISSUE4/IJRPR24766.pdf>

챗봇 도구는 이 중 단어 임베딩을 채택하여, 대표적인 사전 학습 임베딩 모델인 Word2Vec이나 GloVe을 활용해 각 단어를 수백 차원의 수치 벡터로 변환한다. 이 변환의 핵심은 의미가 비슷한 단어가 벡터 공간상에서 가까운 위치에 배치된다는 점이다. 예컨대 '왕'과 '군주'는 공간상 가까이 위치하고, '왕'과 '컴퓨터'는 멀리 위치한다. 입력된 시놉시스 전체는 이러한 단어 벡터들의 집합인 임베딩 행렬(embedding matrix)로 표현되어, 작품의 의미적 표상이 만들어진다.

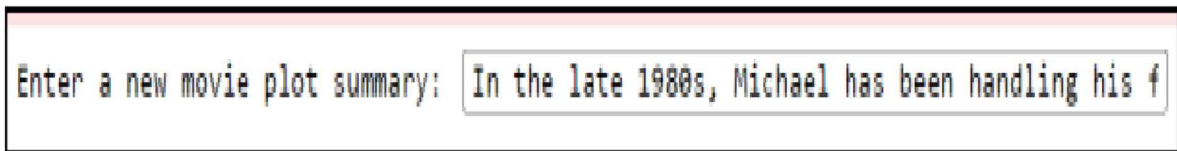
세 번째는 유사도 계산 단계이다. 두 작품의 임베딩 표상 사이의 코사인 유사도(cosine similarity)**를 계산하여 의미적 거리를 정량화한다. 값이 1에 가까울수록 두 작품의 의미가 유사하며, 0에 가까우면 무관하다는 의미이다.

- * TF-IDF(Term Frequency-Inverse Document Frequency): 단어가 특정 문서에 얼마나 자주 등장하는지(빈도)와 전체 문서들 중 얼마나 흔하지 않은 단어인지(희소성)를 결합해 단어의 중요도를 계산하는 방식. 특정 문서에 자주 등장하는 단어는 높은 가중치, 반대의 경우에는 낮은 가중치를 받게 됨
- ** 단어 임베딩(word embedding): 단어를 고정된 길이의 수치 벡터로 변환해 의미를 수학적으로 표현하는 자연어 처리 기법. 유사한 의미의 단어는 벡터 공간상 가까운 위치에 배치되어, 단어 간 의미 관계를 계산 가능하게 함
- *** 코사인 유사도(cosine similarity): 두 벡터가 이루는 각의 코사인 값을 통해 유사도를 측정하는 지표. 값이 1에 가까울수록 의미가 유사함을 뜻한다. 문장이나 문서 간의 의미적 거리를 정량화하는 데 널리 쓰임

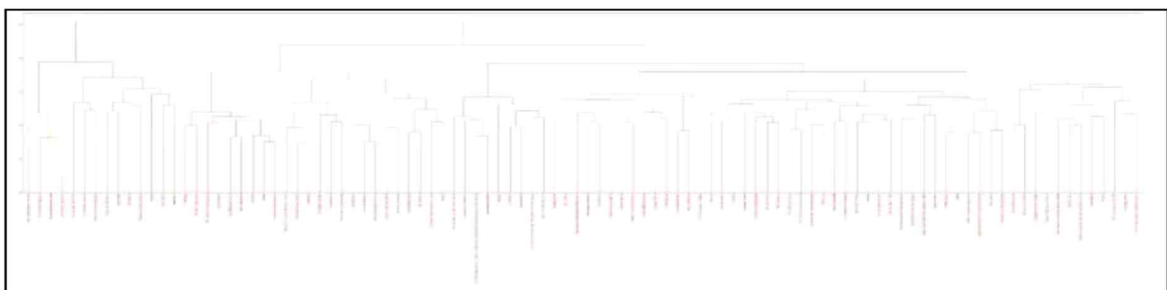
• 분석 결과의 시각화와 대중적 확산 가능성

이러한 의미 기반 분석은 단순한 키워드 일치 방식과 본질적으로 다르다. 단어의 출현 빈도만을 다루는 전통적 단어 가방(bag-of-words) 모델은 어휘를 다른 동의어로 바꾸기만 해도 우회될 수 있다. 반면 임베딩 기반 분석은 어휘를 바꾸어도 비슷한 의미를 지닌 줄거리는 유사한 벡터 표현으로 나타나기 때문에, "왕이 음식을 거부했다"와 "군주가 식사를 마다했다" 같은 표면적으로 다른 문장도 같은 의미로 인식된다. 임베딩 기반 도구는 이러한 의미적 증가성을 포착할 수 있기 때문에, 단순한 어휘 변경으로는 우회할 수 없는 표절 탐지가 가능하다.

[그림 5] 챗봇 기반 영화 표절 탐지 시스템의 입력 화면과 분석 결과



영화 플롯 입력 화면



영화 간 플롯 유사도 덴드로그램

출처: Jothiragavan T 외 2인, "MOVIE PLAGIARISM DETECTION CHATBOT", International Journal of Research Publication and Reviews, 2024.04, <https://ijrpr.com/uploads/V5ISSUE4/IJRPR24766.pdf>

분석 결과는 유사도가 높은 작품 목록과 함께 [그림 5]와 같이 시각화되어 제시된다. 결과 표현 방식으로는 덴드로그램(dendrogram)이 활용되는데, 이는 계층적 군집화(hierarchical clustering) 결과를 나무 모양으로 표현한 그래프이다. 가까운 의미를 가진 작품끼리 먼저 묶이고 점차 더 큰 군집으로 합쳐지는 구조를 보여주어, 사용자는 입력한 줄거리가 데이터베이스 내 어느 작품군에 가까운지 한눈에 확인할 수 있다. 해당 연구에서 보고된 도구의 성능은 정밀도(precision) 82%, 재현율(recall) 78% 수준으로 의미 기반 분석이 일정한 실용성에 도달했음을 보여준다.¹⁷⁾

챗봇과 같은 대중적 도구의 확산은 창작 생태계의 저변을 넓히는 데 기여할 수 있다. 소수의 대형 제작사뿐만 아니라 독립 영화감독, 시나리오 작가 지망생, 웹소설 작가 등 개인 창작자들이 창작 초기 단계에서부터 자신의 아이디어를 스스로 성찰하고 독창성을 강화할 기회를 얻을 수 있기 때문이다. 이는 의도치 않은 유사성으로 인한 분쟁을 예방하는 동시에, 창작자가 자신의 작품에 대한 저작권 관련 확신을 가지고 활동할 수 있는 환경을 조성하는 데도 기여한다. 향후 이러한 서비스는 구독형 모델이나 플랫폼 부가 기능 형태로 발전하며 새로운 시장을 열어갈 잠재력을 지닌다.

- 데이터베이스 구축과 데이터 처리 기술의 과제

AI 표절 탐지 시스템의 실효성은 분석 기반이 되는 데이터베이스의 규모와 질에 좌우된다. 앞서 살펴본 챗봇 도구는 TMDb(The Movie Database)와 같은 공개 영화 데이터베이스에서 등장인물, 주요 사건, 주제 등의 정보를 수집해 활용했다.¹⁸⁾ 그러나 공개 데이터를 그대로 활용하기에는 한계가 있다.

시나리오에는 작가나 제작사마다 작성 형식이 달라 장면 표제(scene heading), 지문, 대사 등이 일관되지 않게 표기된 경우가 많아, AI가 의미 있는 분석을 수행하려면 이러한 형식적 다양성을 표준화된 형태로 정제하는 전처리 과정이 필요하다. 또한 장르, 제작 연도, 등장인물, 주요 사건 같은 메타데이터를 구조화하여 함께 저장해야 분석 결과의 해석 가능성과 정확성도 높아진다.

데이터셋의 다양성도 중요한 기술적 변수이다. 영어권 작품 중심으로 구성된 데이터베이스는 영어가 아닌 언어로 작성된 작품이나, 비영어권 문화의 색채가 강한 작품을 분석할 때 정확도가 떨어지므로, 다국어 임베딩 모델과 문화권별로 미세 조정된 학습 데이터가 함께 갖춰져야 한다. 이러한 기술적 토대가 마련될수록 AI 서사 분석의 적용 범위도 확장될 수 있다.

- 영화 산업의 창작 윤리 강화와 신뢰 생태계 구축

궁극적으로 AI 서사 분석 기술이 나아갈 방향은 단순히 표절을 적발하는 감시 도구가 아니라, 산업 전반의 창작 윤리를 강화하고 구성원 간의 신뢰를 강화하는 생태계를 조성하는 데 있다. 반복되는 표절

17) Jothiragavan T 외 2인, "MOVIE PLAGIARISM DETECTION CHATBOT", International Journal of Research Publication and Reviews, 2024.04, <https://ijrpr.com/uploads/V5ISSUE4/IJRPR24766.pdf>

18) Jothiragavan T 외 2인, "MOVIE PLAGIARISM DETECTION CHATBOT", International Journal of Research Publication and Reviews, 2024.04, <https://ijrpr.com/uploads/V5ISSUE4/IJRPR24766.pdf>

논쟁은 창작자, 제작사, 대중 사이에 많은 논란을 불러왔다. AI 기반의 검증 시스템이 사회적 신뢰를 얻고 보편화된다면, 창작자는 독창성을 입증할 근거를 확보하고, 제작사는 불필요한 분쟁에 휘말릴 위험을 줄일 수 있으며, 대중은 작품의 독창성에 대한 신뢰를 높일 수 있게 된다.

기술의 존재 자체가 산업의 관행을 긍정적인 방향으로 유도하는 역할도 기대해 볼 수 있다. 누구나 손쉽게 유사성을 검증할 수 있는 환경에서는 안일한 모방이나 차용이 이루어지기 어려워지고, 창작자들은 자연스럽게 독창성을 확보하기 위한 노력을 기울이게 된다. 이는 흥행작의 성과가 불필요한 논란으로 퇴색되는 사례를 줄이고, 창작자가 창작 활동의 가치로 평가받는 문화가 자리 잡는 데 기여할 수 있다. AI 기술은 영화 산업 구성원 모두가 창작의 가치를 존중하고 보호하는 신뢰 생태계를 구축하는 인프라로 작동할 수 있다.

III. 결론 및 전망

- 요약: 주관적 논쟁에서 데이터 기반 논의로의 전환

본 보고서는 영화 '왕과 사는 남자'의 표절 논란을 계기로 반복되어 온 영화 산업의 저작권 분쟁 문제를 분석하고, 그 대안으로서 AI 기반 서사 분석 기술과 챗봇 형태의 표절 탐지 도구의 가능성을 조명했다. 지금까지 영화 시나리오를 둘러싼 저작권 분쟁은 유사성 판단 기준이 명확하지 않아 당사자 간 이견이 발생하는 경우가 많았으며, 이는 산업 생태계 전반에 걸쳐 정량적 근거에 기반한 판단 기준의 필요성을 높여왔다. 이러한 상황에서 AI 기반 서사 분석 기술과 챗봇 형태의 표절 탐지 도구는 유사성을 정량적 데이터로 제시함으로써 논의의 패러다임을 전환할 기회를 제공한다.

현재 기술은 유사도 점수의 해석 기준 부재, 모델의 편향성 등 명확한 한계를 안고 있다. 그러나 기술의 정확도와 신뢰성이 높아지고 활용 범위가 넓어진다면, 분쟁의 사후 해결뿐만 아니라 창작자를 위한 사전 검증, 챗봇 형태의 대중적 표절 탐지 도구, 나아가 투자 심사에 정량적 근거를 더하는 도구로까지 확장될 수 있다. AI 서사 분석 기술이 가진 진정한 가치는 표절을 판결하는 기계적 심판이 아니라 데이터에 기반한 합리적 토론의 장을 여는 데 있다. 기술의 등장이 영화 산업의 오랜 과제였던 저작권 분쟁을 보다 성숙하게 다루는 전환점으로 이어질 수 있을지, 앞으로의 기술 발전과 활용 사례의 축적이 그 가능성을 가능하게 될 것이다.

- 기술적 제언: 설명가능 AI(XAI) 도입과 모델 투명성 확보

AI 서사 분석 기술이 사회적 신뢰를 얻기 위해서는 기술 자체의 지속적인 발전이 뒤따를 때 가능하다. 특히 본론 1에서 지적된 블랙박스 문제는 해결이 쉽지 않은 과제로 남아 있다. AI가 어떤 논리적 과정을 거쳐 특정 유사도 점수를 산출했는지 투명하게 이해할 수 없다면, 그 결과는 설득력을 얻기 어렵다. 향후

기술 개발은 정확도를 높이는 것을 넘어, AI의 판단 과정을 인간이 이해할 수 있는 형태로 설명하는 설명가능 AI(XAI, eXplainable AI) 도입으로 확장될 전망이다.

이는 AI의 판단 근거에 대한 투명성을 높여 결과의 신뢰도를 뒷받침하고, 사용자가 결과를 비판적으로 검토하는 데도 도움이 된다. XAI 기반의 서사 분석 모델 개발이 진전되고 모델의 편향성이 지속적으로 검증·개선될 때, 기술의 사회적 수용 가능성도 높아질 것으로 기대된다.

- 미래 비전: AI 기술이 만드는 건강한 창작 생태계를 향하여

AI 서사 분석 기술의 도입이 가져올 궁극적인 미래는 모든 창작 활동이 데이터에 의해 감시받는 사회가 아니라, 창작의 가치가 더욱 존중받고 보호받는 생태계의 구현이다. AI가 제공하는 정량적 데이터는 창작자에게 자신의 독창성을 증명할 수 있는 근거가 되고, 표절에 대한 두려움 없이 창작 활동에 몰입할 수 있는 안전망으로 작동할 수 있다. 나아가 산업계 전반에서는 소송 대응 비용이 줄고, 그 자원이 창의적인 콘텐츠 개발로 향하는 선순환 구조도 형성될 것으로 전망된다.

AI 기술을 단순히 저작권 침해를 탐지하는 도구로만 볼 것이 아니라, 창작 과정을 지원하고 산업의 신뢰를 구축하며 문화의 다양성을 풍성하게 하는 인프라로 바라보는 시각도 점차 힘을 얻고 있다. 기술이 발전하고 창작자와 산업 구성원 모두가 그 혜택을 공정하게 누릴 수 있을 때, 창의성이 발휘되는 창작 환경에 한 걸음 가까워질 수 있다. AI 서사 분석 기술 및 표절 탐지 도구의 발전과 활용이 영화 산업의 창작 환경을 어떻게 바꾸어 나갈지, 앞으로의 기술 발전과 활용 사례의 축적 속에서 그 방향이 서서히 드러날 것이다.

참고문헌

- Jon Chun, "AIStorySimilarity: Quantifying Story Similarity Using Narrative for Search, IP Infringement, and Guided Creativity", Association for Computational Linguistics, 2024.11, <https://aclanthology.org/2024.conll-1.13.pdf>
- Jothiragavan T 외 2인, "MOVIE PLAGIARISM DETECTION CHATBOT", International Journal of Research Publication and Reviews, 2024.04, <https://ijrpr.com/uploads/V5ISSUE4/IJRPR24766.pdf>
- 조지영, "'왕과 사는 남자', 1600만 관객 돌파! '극한직업' 잡고 역대 2위 갈까?", 조선일보, 2026.04.12., <https://www.chosun.com/entertainments/movie/2026/04/12/MU2TIMRWME4TEMBQMFSTQMRSMU/>
- 서지영, "'7가지 유사점 있다' 천만 돌풍 '왕사남', 돌연 표절 의혹...제작사 '사실무근'", 아시아경제, 2026.03.10., <https://www.asiae.co.kr/article/2026031008523154431>
- 김도연, "'표절이면 이름 넣어달라' vs '사실무근'...1200만 앞둔 '왕과 사는 남자' 표절 논란", 서울경제, 2026.03.10., <https://www.sedaily.com/article/20017405>
- 한현정, "1200만 눈앞 '왕사남' 표절 논란...유족 문제 제기, 제작사는 부인", 스타투데이, 2026.03.10., <https://www.mk.co.kr/news/movies/11983516>
- 강세준, "천만관객 '왕사남' 표절 논란..무한반복 스크린 흑역사", 법치뉴스, 2026.03.10., <https://www.bubb.co.kr/news/articleView.html?idxno=4141>