

논문 2024-4-18 <http://dx.doi.org/10.29056/jsav.2024.12.18>

게임 콘텐츠 이용허락 계약을 위한 생성형 AI 기반 맞춤형 계약 조항 추천 시스템

김현수*, 최창준*, 조용준*, 신동명*†

Generative AI based Customized Contract Clause Recommendation System for Game Content License Agreement

Hyun-Soo Kim*, Chang-Jun Choi*, YongJoon Joe*, Dong-Myung Shin*†

요 약

본 논문은 게임 장르별 특성을 반영한 이용허락 계약 조항 생성을 위한 인공지능 기반 시스템을 제안한다. 게임 콘텐츠는 각 장르가 고유한 특성을 지니고 있으며, 이러한 특성의 차이는 계약 조건에도 중요한 영향을 미쳐 법적 분쟁의 가능성을 증대시킬 수 있다. 본 연구에서는 생성형 인공지능 모델인 GPT-4o를 파인튜닝하여, 게임 장르와 계약 목적에 최적화된 맞춤형 계약 조항을 생성하는 시스템을 제안한다. 이를 위해 공개된 표준 계약서를 분석하고, 학습 데이터의 부족을 해결하기 위해 TextGAN을 활용하여 데이터셋을 확장하였다. 실험 결과, 하이퍼파라미터 수정을 통한 파인튜닝된 모델은 Training Loss는 0.4635까지 감소하였으며, 기존의 GPT-4o 모델에 비해 게임 콘텐츠 이용허락 계약서 작성에 더욱 적합한 문구를 생성하는 성능을 보였다. 이러한 시스템은 계약 작성의 효율성을 향상시키고, 법적 분쟁의 가능성을 감소시키는 데 기여할 것으로 기대된다.

Abstract

This paper proposes an AI-based system for generating license agreement clauses customized to the characteristics of different game genres. Game content possesses unique traits depending on its genre, and these differences significantly influence contract terms, potentially increasing the probability of legal disputes. In this study, we fine-tuned a generative AI model, GPT-4o, to create personalized contract clauses optimized for specific game genres and contractual purposes. To achieve this, we analyzed publicly available standard contracts and addressed the lack of training data by expanding the dataset using TextGAN. Experimental results showed that the fine-tuned model, optimized through hyperparameter adjustments, achieved a decrease in Training Loss to 0.4635 and demonstrated improved performance in generating clauses suitable for game content license agreements compared to the base GPT-4o model. This system is expected to enhance the efficiency of contract drafting and reduce the potential for legal disputes.

한글키워드 : 디지털 콘텐츠, 생성형 AI, 이용허락, 추천 시스템

keywords : Digital Contents, Generative AI(Artificial Intelligence), License, Recommendation System

* 엘에스웨어(주) 소프트웨어연구소 연구개발본부

접수일자: 2024.11.17. 심사완료: 2024.12.12.

† 교신저자: 신동명(email: roland@lsware.com)

게재확정: 2024.12.20.

1. 서론

게임 콘텐츠는 장르에 따라 각기 다른 특성을 지니고 있으며, 이러한 특성에 따라 이용허락 계약 시 요구되는 계약 조항 역시 장르별로 상이하게 설정되는 경우가 많다[1]. 예를 들어, 어드벤처 게임과 롤플레이 게임(RPG), 공포 게임은 각각의 스토리라인, 캐릭터 설정, 연출 방식 등이 다르기 때문에 계약 조건에도 차이가 발생할 수밖에 없다. 이러한 다양성은 계약 작성 시 비효율성을 초래하고, 계약 당사자 간의 이해 불일치로 인해 법적 분쟁이 발생할 가능성을 높인다[2]. 특히, 각기 다른 계약 조항들이 불완전하게 포함되거나 법적 검토가 부족한 상태로 계약이 체결될 경우, 계약 이후 발생할 수 있는 책임 소재와 권리 문제는 심각한 갈등을 초래할 수 있다. 또한, 이용허락 계약서의 불완전함으로 인해 스토리의 주요 내용이나 중요한 게임 요소가 무단으로 사용되거나, 스포일링(spoiling)되는 문제가 발생할 수 있다. 이러한 저작권 침해는 게임의 몰입감을 저해하고, 콘텐츠의 가치를 손상시켜 개발자와 퍼블리셔에게 경제적 손실을 초래할 수 있다. 특히, 스토리형 게임의 경우 스포일링은 사용자의 게임 경험을 크게 훼손하기 때문에, 이러한 침해를 방지할 수 있는 명확한 계약 조항이 필수적이다[3].

이러한 문제를 체계적으로 해결하기 위한 대응책은 여전히 미흡한 상태이며, 게임 개발자나 퍼블리셔는 계약의 법적 전문성이 부족한 경우가 많아, 계약서 작성 시 법률 전문가의 도움을 받지 않는 한 적절한 조항을 포함하기 어려운 실정이다. 이에 따라, 본 논문에서는 생성형 AI를 이용한 게임 장르별 이용허락 계약 조항 추천 시스템을 제안한다. 이를 통해 계약 작성의 효율성을 높이고, 계약 당사자 간의 이해 충돌을 줄이며, 법적 분쟁의 가능성을 낮추는 것을 목표로 한다.

2. 관련 연구

2.1 게임 장르별 이용허락 계약 조항

게임 콘텐츠 이용허락 계약의 조항은 게임 장르와 그 특성에 따라 구체적으로 구성되어야 한다. MMORPG, FPS, 퍼즐, 스포츠 등 다양한 게임 장르는 각각 고유한 게임플레이 메커니즘과 사용자 상호작용 방식을 가지며, 이러한 차별화된 특성은 계약의 구조와 내용에 중요한 영향을 미친다[4]. 특정 장르의 특성을 고려하지 않고 획일적인 계약 조항을 적용할 경우, 무단 이용, 스포일링, 지적 재산권 침해 등의 법적 문제가 발생할 가능성이 높아진다. 따라서 각 장르에 최적화된 계약 조항을 마련하는 것은 필수적이다.

게임 장르별 이용허락 계약에서 일반적으로 포함되어야 하는 공통적인 계약 조항으로는 저작권 및 지적 재산권 보호, 사용권 허가 범위, 변경 및 수정 권한, 수익 배분, 품질 보증 및 유지보수, 법적 책임 및 면책 조항, 기밀 유지 등이 있다[5]. 이러한 공통 조항들은 모든 게임 콘텐츠 계약에서 필수적이며, 계약 당사자 간의 권리와 의무를 명확히 규정하는 데 중점을 둔다. 그러나 이러한 계약 조항이 명확하고 구체적이어야 함에도 불구하고, 법률 전문가의 도움 없이 계약 당사자들이 이를 직접 작성하기에는 상당한 어려움이 있다. 특히, 계약서 작성 시 게임 장르별 특성과 계약 목적을 충분히 반영한 맞춤형 계약 조항이 필수적이지만, 전문적인 법적 지식과 경험이 부족한 경우 이러한 중요한 요소들이 간과되기 쉽다.

게임 이용허락 계약의 효율성을 높이고 법적 리스크를 최소화하기 위해서는 법률 전문가의 지원이나 자동화된 계약 조항 생성 도구의 활용이 매우 중요하다[6]. 이를 통해 게임 장르별 특성과 계약 목적을 충분히 고려한 맞춤형 계약 조항을 적용할 수 있어야 한다.

2.2 TextGAN

TextGAN은 텍스트 데이터를 생성하는 데 특화된 생성적 적대 신경망(Generative Adversarial Network) 기반 모델로, 자연어 처리 분야에서 복잡한 언어의 문법적 일관성과 의미적 유사성을 학습하여 고품질의 텍스트 데이터를 생성하는 데 중점을 두고 있다[7]. TextGAN의 핵심 구조는 생성자(Generator)와 판별자(Discriminator)라는 두 개의 신경망으로 구성되며, 이들은 서로 경쟁하며 학습하는 적대적 학습 방식을 통해 텍스트의 품질을 개선시킨다. 생성자는 계약 조항과 같은 텍스트 데이터를 생성하는 역할을 하며, 판별자는 생성된 텍스트가 실제 데이터와 얼마나 유사한지 여부를 판단한다. 이러한 경쟁적 학습 구조를 통해 생성자가 더욱 현실적이고 자연스러운 텍스트 데이터를 생성할 수 있도록 한다.

TextGAN을 활용한 텍스트 데이터 생성의 가장 큰 장점은 기존 데이터셋의 제한성을 극복할 수 있다는 점이다[8]. 게임 콘텐츠 이용허락 계약과 같은 특수한 법적 텍스트는 데이터 수집 자체가 어렵다. 법적 계약 조항은 매우 구체적이고 특정 상황에 따라 다르게 작성되어야 하므로, 이를 위한 충분한 데이터셋을 구축하는 데는 많은 시간과 비용이 필요하다. TextGAN은 기존 데이터에 의존하지 않고 새로운 데이터 패턴을 학습하고 생성할 수 있어, 특히 계약과 같은 제한된 데이터셋 상황에서도 효과적으로 활용될 수 있다. 이를 통해 게임 장르별 이용허락 계약의 다양한 상황에 적합한 데이터를 생성하여, 모델의 학습 효율성을 크게 향상시킬 수 있다.

2.3 GPT-4o

GPT-4o는 OpenAI에서 개발된 최신 자연어 처리 모델로, 대규모 텍스트 데이터 학습을 통해 자연어 생성과 이해에 있어 탁월한 성능을 발휘하는 언어 모델이다[9]. 이는 Transformer 아키

텍처를 기반으로 하여 텍스트의 문맥적 이해와 고품질의 응답 생성을 가능하게 한다. GPT-4o는 문법적 일관성, 문맥적 적절성, 그리고 의미적 깊이를 이해하는 데 매우 뛰어나며, 이를 바탕으로 자연스럽게 일관된 텍스트 생성이 가능하다.

GPT-4o는 대규모의 텍스트 데이터를 학습한 후, 주어진 입력에 맞춰 적절한 응답을 생성하는 방식으로 동작한다. 이를 통해 입력된 텍스트의 문맥을 정밀하게 분석하고, 그에 적합한 형태의 언어 데이터를 생성한다. 특히, Transformer 아키텍처는 병렬화된 계산을 통해 대량의 데이터를 빠르게 처리하며, 각 단어 간의 관계를 정확히 파악하여 문맥적으로 자연스러운 문장을 생성한다[10]. 이 과정에서 모델은 다층 신경망을 통해 텍스트의 패턴과 규칙을 학습하고, 이를 바탕으로 문맥에 맞는 단어와 구문을 선택한다. 또한, 학습한 방대한 양의 데이터로부터 각 상황에 적합한 어휘와 표현을 선택하여 응답을 생성함으로써, 인간이 작성한 것과 유사한 수준의 자연스러운 텍스트를 생성한다.

GPT-4o는 특히 복잡한 문서 작성이나 특정 도메인 지식이 필요한 텍스트 생성에서 강점을 발휘한다. 게임 장르별로 요구되는 계약 조항은 매우 다양하고 구체적이며, 이러한 특성을 반영하기 위해서는 복잡한 언어적 이해와 정밀한 문맥 조정이 필수적이다. GPT-4o는 게임 장르와 관련된 방대한 데이터와 법적 문서를 학습하여 각 장르의 특성에 맞는 계약 조항을 자동으로 생성할 수 있다. 이는 Transformer의 셀프 어텐션(self-attention) 메커니즘을 통해 입력된 문서 내의 중요한 부분을 강조하고, 전체 문맥을 고려한 응답을 생성할 수 있다[11]. 결과적으로, 복잡하고 구조화된 정보에 대한 깊이 있는 이해를 바탕으로 고품질의 텍스트를 생성하며, 이를 통해 계약 작성자들은 게임의 특성과 요구사항에 맞는 맞춤형 계약서를 효율적으로 작성할 수 있다.

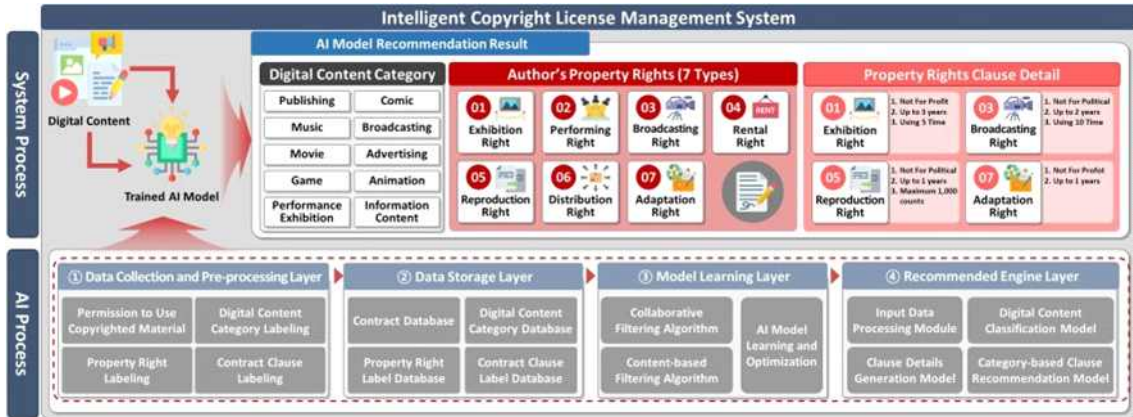


그림 1. 게임 콘텐츠 이용허락 계약 조항 추천 시스템 개요도
 Fig 1. Overview of the Game Content License Agreement Clause Recommendation System

3. 이용허락 계약 조항 추천 시스템

본 논문에서 대상으로 하는 게임 장르는 스토리 중심의 게임으로, 어드벤처, 롤플레이, 공포 게임에 대한 장르를 중점적으로 다룬다.

3.1 데이터 수집 및 전처리

데이터 수집 및 전처리 단계에서는 게임 장르별 이용허락 계약을 위해 관련 산업에서 공개된 표준 계약서(e.g., 저작권산권 비독점적 이용허락 표준 계약서, 게임 퍼블리싱 표준 계약서) 및 게임 이용허락과 관련된 저작권 가이드 등을 확보하여 의미 있는 계약 조항을 추출하고 이를 분석한다. 수집된 데이터는 가능한 한 다양한 계약 상황을 포함하도록 하여 계약 조항 추천 모델의 학습 범위를 확장하는 데 중점을 둔다.

수집된 데이터는 텍스트 정규화 과정을 거쳐 일관성 있는 형태로 변환된다. 이 과정에는 특수 문자 제거, 불필요한 숫자 제거 등이 포함되며, 각 문장에서 불용어(stopwords)를 제거하여 데이터의 품질을 향상시킨다. 또한, 계약서의 각 조항이 어떤 게임 장르에 해당하는지, 어떤 목적(스포일러 방지, 연령 제한, 무단복제 방지, 수익 창

출 등)을 갖는 조항인지를 추가한다. 이를 통해 계약 조항의 주요 요소와 각 장르에 특화된 법적 요구사항을 반영하여 스토리 보호, 캐릭터 IP 보호, 연출 요소 보호와 같은 특화된 계약 조항을 생성한다. 또한, 기존의 .csv 데이터 형식이 아닌 생성형 AI 모델인 GPT-4o 학습데이터 형식인 JSONL(JSON Line) 형식으로 학습 데이터를 변형한다. 이 데이터는 표 1과 같은 형태로 정보를 저장한다. 이 학습 데이터는 시스템의 목적, 사용자의 질문, 어시스턴트의 응답에 대한 정보를 포함하고 있다. 본 논문에서는 사용자는 게임 장르와 계약서 작성 목적을 입력하면 어시스턴트는 이에 부합하는 계약 조항을 생성한다.

표 1. JSONL 학습 데이터 구조
 Table 1. JSONL Training Data Structure

구분	번호	Key	Value
Text Data	Info #1	Role	System
		Content	Explanation
	Info #2	Role	User
		Content	Question
	Info #3	Role	Assistant
		Content	Answer

3.2 계약 데이터셋 확장 모델

본 논문에서는 이용허락 계약 조항 추천 모델 학습을 위한 학습 데이터 확보를 위해 TextGAN을 활용하여 기존 계약 조항 데이터셋을 확장한다. 게임 콘텐츠 이용허락 계약과 같은 특수한 법적 텍스트는 그 자체로 데이터의 확보가 어려운 경우가 많다. 따라서, TextGAN을 통해 기존에 제한된 형태로 존재하던 계약 조항 데이터셋을 다양하게 확장함으로써, 계약 조항 추천 시스템의 학습에 필요한 충분한 데이터를 확보한다.

TextGAN 학습의 첫 단계는 데이터 전처리로, 본 연구에서는 이용허락 계약 조항 추천 모델 학습 과정에 사용될 데이터셋을 사용한다. 이 데이터에서 'User Content'와 'Assistant Content' 필드 정보를 추출하고 이를 학습 가능한 형식으로 변환한다. 구체적으로, 'User Content' 필드에서 게임 장르와 계약 목적을 추출한다. 게임 장르는 데이터셋 내에 명시된 키워드를 기반으로 정의되었으며, 계약 목적은 계약의 배경과 의도를 나타내는 정보로, 자연어 처리 기반의 필터링 및 규칙 기반 추출 방법을 통해 식별된다. 계약 목적의 식별 과정에서는 텍스트 내에서 특정 패턴이나 의미를 추출하기 위해 자연어 처리 기술인 토큰화, 품사 태깅, 의존 구문 분석 등을 활용한다.

이후, 'Assistant Content' 필드에서는 계약 조항을 추출한다. 계약 조항은 실제로 계약서에 포함되는 주요 문구로, 계약 당사자 간의 권리와 의무를 명확히 규정하는 내용을 담고 있다. 추출된 정보는 TextGAN 학습에 적합한 형식으로 변환되었으며, 변환된 데이터는 '게임 장르: <게임 장르>; 목적: <계약 목적>; 조항: <계약 조항>'의 구조로 구성되어 새로운 학습 데이터셋 D' 에 저장된다. 이와 같은 데이터 구조를 통해 학습 모델이 게임 장르와 계약 목적을 명확히 이해하고, 적절한 계약 조항을 생성할 수 있다.

TextGAN 학습 단계에서는 생성자와 판별자의 두 신경망이 상호 경쟁적으로 학습하여 텍스트 데이터를 생성할 수 있는 모델을 구축한다. 생성자는 랜덤한 노이즈를 입력받아 가짜 텍스트 시퀀스를 생성하며, 이 텍스트는 학습 데이터와 유사한 구조를 가지도록 설계된다. 반면 판별자는 입력된 텍스트가 학습 데이터셋 D' 에서 온 진짜 텍스트인지, 생성자가 생성한 가짜 텍스트인지 구분하는 역할을 한다.

학습 과정은 두 단계로 나뉜다. 첫째, 판별자 학습 단계에서는 학습 데이터셋에서 진짜 텍스트 샘플 $R \sim D'$ 과 생성자가 생성한 가짜 텍스트 $F = G(Noise)$ 를 입력받아 판별자가 두 텍스트를 구분할 수 있도록 학습한다. 이때 손실 함수 L_D 는 진짜 텍스트와 가짜 텍스트에 대해 각각 계산되며, 이를 기반으로 판별자의 파라미터가 업데이트된다. 둘째, 생성자 학습 단계에서는 판별자를 속이는 방향으로 생성자를 학습시킨다. 랜덤 노이즈를 입력받아 생성된 새로운 텍스트 샘플 $F' = G(Noise)$ 를 판별자에 입력하고, 판별자가 이를 진짜로 판단하도록 생성자의 손실 함수 L_G 를 계산한다. 이 손실 값을 최소화하는 방향으로 생성자의 파라미터가 업데이트된다. 이러한 과정은 여러 에폭(epoch) 동안 반복되며, 생성자와 판별자가 경쟁적으로 학습하여 더 정교한 텍스트를 생성할 수 있도록 발전한다. 학습이 완료되면 최종적으로 학습된 모델이 저장된다.

학습된 TextGAN 모델을 활용하여 새로운 텍스트 데이터를 생성한다. 생성 과정은 랜덤한 노이즈를 입력하여 시작되며, 생성자는 이를 기반으로 학습 데이터셋과 유사한 형식의 텍스트를 생성한다. 생성된 텍스트 샘플 S' 는 게임 장르, 계약 목적, 계약 조항의 세 가지 요소를 포함한다. 이렇게 생성된 데이터셋은 이용허락 계약 조항 추천 모델의 학습데이터로 사용된다.

표 2. 계약 텍스트 데이터 생성 알고리즘
Table 2. Contract text data generation algorithm

Algorithm 1: TextGAN-based Generation of Game Genres, Contract Purposes, and Terms

Require: Training dataset D (each entry contains "User Content" and "Assistant Content")
Ensure: Trained TextGAN model, generated game genres, contract purposes, and terms

- 1: **Data Preprocessing:**
- 2: **for** each data entry $d \in D$ **do**
- 3: $U \leftarrow d$ ["User Content"], $A \leftarrow d$ ["Assistant Content"]
- 4: $G \leftarrow U$ ["Game Genre"], $P \leftarrow U$ ["Contract Purpose"],
- 5: $T \leftarrow A$ ["Contract Terms"]
- 6: Format into training sequence:
- 7: $S \leftarrow$ "Genre: G ; Purpose: P ; Terms: T "
- 8: **end for**
- 9: Save preprocessed data: $D' \leftarrow \{S_1, S_2, \dots, S_n\}$
- 10: **TextGAN Training:**
- 11: Initialize Generator G and Discriminator D
- 12: **for** epoch = 1 to N **do**
- 13: // Train Discriminator
- 14: Sample real text $R \sim D'$
- 15: Generate fake text $F \leftarrow G(\text{Noise})$
- 16: Compute Discriminator loss: $L_D \leftarrow \text{Loss}(R, F)$
- 17: Update Discriminator
- 18: // Train Generator
- 19: Generate new fake text $F' \leftarrow G(\text{Noise})$
- 20: Compute Generator loss: $L_G \leftarrow \text{Loss}(F', \text{Real_Label})$
- 21: Update Generator
- 22: **end for**
- 23: Save trained model
- 24: **Generation and Output:**
- 25: Generate new text sample: $S' \leftarrow G(\text{Noise})$
- 26: $G' \leftarrow S'$ ["Game Genre"], $P' \leftarrow S'$ ["Contract Purpose"],
 $T' \leftarrow S'$ ["Contract Terms"]
- 27: Output results: $G', P', T' = 0$

3.3 이용허락 계약 조항 추천 모델

본 논문에서는 최근 공개된 GPT-4o 모델을 활용하여 사용자가 게임 장르와 이용허락 계약서 작성 목적을 입력하면, 해당 계약서에 사용될 계약 조항을 생성하는 모델을 제안한다. 기존의 공개된 모델은 계약서 작성에 필요한 문구에 최적화되어 있지 않고, 일반적인 일상 용어 생성에

특화되어 있기 때문에, 본 논문에서는 이 모델을 파인튜닝(Fine-Tuning)하여 특정 도메인에 맞게 활용할 필요가 있다. 이를 위해 전처리된 데이터와 계약 데이터셋 확장 모델에서 생성한 확장 데이터를 사용하고 하이퍼파라미터와 학습률, 배치 크기, 에폭을 정의하여 GPT-4o 모델을 파인튜닝한다. 이 과정에서 GPT-4o 모델은 게임 장르와 계약서 작성 목적이 포함된 프롬프트와 이에 대한 응답 쌍을 반복적으로 학습함으로써, 최적화된 응답을 생성할 수 있도록 학습되었다.

표 3. 계약 조항 생성 알고리즘
Table 3. Contract clause generation algorithm

Algorithm 2: Generating Contract Terms using Fine-Tuned GPT-4o

Require: Dataset (JSONL format with "game genre", "contract purpose", "contract terms")
Require: User input: "game genre", "contract purpose"
Ensure: Generated contract terms

- 1: **Data Preprocessing:**
- 2: **for** each entry in the dataset **do**
- 3: game genre \leftarrow entry["game genre"]
- 4: contract purpose \leftarrow entry["contract purpose"]
- 5: contract terms \leftarrow entry["contract terms"]
- 6: prompt \leftarrow "Genre: " + genre + ", Purpose: " + contract purpose
- 7: response \leftarrow contract terms
- 8: Append (prompt, response) to fine-tuning dataset
- 9: **end for**
- 10: Save the fine-tuning dataset
- 11: **Fine-Tuning GPT-4o:**
- 12: Train GPT-4o model using the fine-tuning dataset
- 13: Set training parameters (e.g., learning rate, batch size)
- 14: Save the fine-tuned model after training
- 15: **Contract Terms Generation:**
- 16: Accept user-provided genre and contract purpose
- 17: prompt \leftarrow "Genre:" + user-provided genre + ", Purpose:" + user-provided contract purpose
- 18: Pass the prompt to the fine-tuned model
- 19: generated terms \leftarrow model output
- 20: **Return Results:**
- 21: Output the generated terms = 0

4. 제안 시스템 실험 및 평가

본 연구에서는 게임 장르에서 사용되는 이용허락 계약 조항 300개를 기본 데이터셋으로 사용하였다. 각 이용허락 계약 조항은 계약서 작성 목적에 따라 스포일러 방지, 연령 제한, 콘텐츠 제한으로 구분되었으며, 이러한 구분은 데이터셋 확장 및 학습에 활용되었다. 기본 데이터셋은 TextGAN을 통해 1,000개로 확장되었으며, 이는 이용허락 계약 조항 추천 모델의 학습 데이터셋으로 사용되었다. 데이터 전처리 및 학습 과정은 Colab T4 GPU 환경에서 수행되었으며, 파인튜닝은 OpenAI API를 통해 진행되었다[12].

이용허락 계약 조항 추천 모델의 기본 모델은 GPT-4o-mini-2024-07-18을 기반으로 하였으며, 학습은 에폭 3, 배치 크기 1, 토큰 수 13,650, 학습률 1.8의 하이퍼파라미터로 설정하여 진행되었다. 결과적으로, 그림 2에 나타난 바와 같이 평균 Training Loss는 0.4635로 다소 높은 수준을 보였다. 이는 OpenAI API의 비용적인 제약으로 인해 에폭, 배치 크기, 학습률 변경에 한계가 있었기 때문으로, 이러한 제한이 모델 성능의 향상을 저해한 것으로 판단된다.

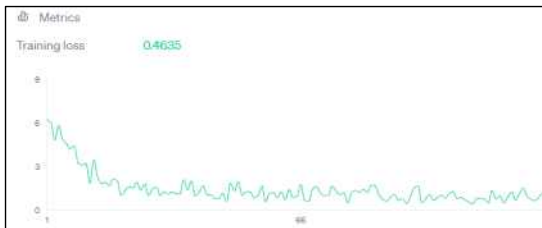


그림 2. 이용허락 계약 조항 추천 모델 성능 평가
Fig 2. License Contract Term Recommendation model evaluation

기존 GPT-4o 모델과 학습된 모델의 성능을 비교한 결과, 그림 3에서 확인할 수 있듯이 기존

GPT-4o 모델의 응답은 게임 콘텐츠 이용허락 계약서에 특화된 문구가 아닌 일반적인 계약서에 사용되는 문구와 유사하게 생성되었다. 이는 기존 모델이 특정 도메인에 대한 전문적인 이해 없이 일반적인 계약 조항을 생성하는 경향이 있음을 보여준다. 이러한 일반적인 응답은 게임 콘텐츠 이용허락 계약서와 같은 특화된 법적 문서 작성에 있어 충분히 구체적이지 않으며, 게임 장르와 계약 목적에 맞춘 맞춤형 조항을 반영하는 데 한계가 있다.

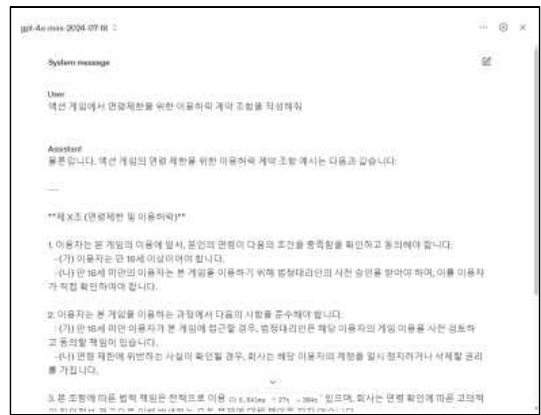


그림 3. 기존 GPT-4o 모델 응답
Fig 3. GPT-4o Response (Original)

반면, 그림 4에서 볼 수 있듯이 학습된 모델은 게임 이용허락 계약서에 특화된 계약 조항을 보다 정확하게 생성하는 성능을 보였다. 학습된 모델은 게임 장르별로 요구되는 특정 요구사항을 반영하여 계약서를 작성하도록 파인튜닝된 결과, 계약 목적과 게임의 특성에 부합하는 보다 구체적이고 타당한 계약 조항을 생성할 수 있었다. 예를 들어, 학습된 모델은 게임의 특정 장르에 따른 연령 제한, 스포일러 방지, 콘텐츠 사용 범위 등의 요소를 반영한 문구를 생성하며, 이는 게임 이용허락 계약의 특수성을 반영한 결과로 해석될 수 있다[13].

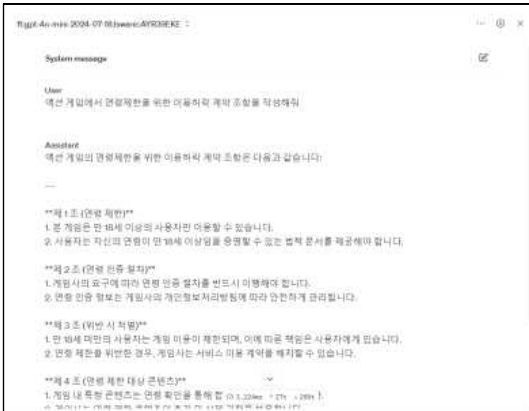


그림 4. 파인튜닝 GPT-4o 모델 응답
Fig 4. GPT-4o Response (Fine-Tuned)

이러한 결과는 학습된 모델이 게임 콘텐츠 이용허락 계약서와 같은 특수한 분야에서 더 높은 수준의 맞춤형 응답을 제공할 수 있는 가능성을 보여준다. 특히, 특정한 법적 요구사항을 반영하여 게임 장르에 맞춘 계약 조항을 자동으로 생성하는 능력은 계약서 작성의 효율성을 높이고, 게임 콘텐츠의 무단 이용 방지와 같은 법적 리스크를 최소화하는 데 중요한 역할을 할 수 있다.

5. 결론

본 논문에서는 게임 콘텐츠 이용허락 계약서 작성 시, 게임 장르와 계약 목적에 부합하는 맞춤형 계약 조항을 생성하는 시스템을 제안하였다. 이를 위해, 게임 산업에서 공개된 표준 계약서를 분석하여 계약 조항을 도출하고, 각 조항에 적합한 게임 장르와 계약 목적을 라벨링함으로써 학습 데이터셋을 구축하였다. 또한, 학습 데이터의 제한적인 규모를 보완하기 위해 TextGAN을 활용하여 이용허락 계약 조항 추천 모델의 학습에 사용할 추가 데이터를 생성하였다.

제안된 이용허락 계약 조항 추천 모델은 기존 GPT-4o 모델을 기반으로 파인튜닝을 수행하여, 게임 콘텐츠 이용허락 계약서 작성에 최적화된 계약 조항을 생성하도록 설계되었다. 동일한 프롬프트를 활용한 실험 결과, 파인튜닝된 모델이 기존 GPT-4o 모델보다 게임 콘텐츠 이용허락 계약서에 더 적합한 문구를 생성하는 성능을 보였음을 확인하였다. 그러나, 객관적인 성능 지표에서 최적의 성능을 달성하지 못한 이유로는 OpenAI API 비용 제약이 있었으며, 추가적인 비용을 투입하여 하이퍼파라미터를 조정한다면 성능 향상을 기대할 수 있을 것으로 보인다.

향후 연구에서는 다양한 게임 장르와 계약 목적에 맞는 계약 조항을 추가적으로 확보하고, 이를 바탕으로 데이터셋을 확장할 계획이다. 또한, GPT-4o 모델 이외의 생성형 AI 모델을 파인튜닝하여, 맞춤형 계약 조항 생성 및 추천에 최적화된 모델을 개발하는 연구도 진행할 예정이다. 이러한 연구를 통해 게임 콘텐츠 이용허락 계약서 작성의 효율성을 극대화하고, 법적 리스크를 최소화할 수 있는 방안을 모색할 예정이다.

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 신기술 융합 저작권 기술 개발 사업으로 수행되었음(과제명 : Web3.0 탈중앙화 환경에서 창작자간의 저작권 이용허락 거래 자동화 기술 개발, 과제번호 : RS-2024-00441360, 기여율 : 100%)

참고 문헌

[1] Samuelson, P. (1993). Fair use for computer programs and other copyrightable works in digital form: the implications of Sony, Galoob and Sega. J.

- Intell. Prop. L., 1, 49, <https://digitalcommons.law.uga.edu/jipl/vol1/iss1/6/>
- [2] Wallace, R. (2014). Modding: Amateur authorship and how the video game industry is actually getting it right. *BYu L. REv.*, 219, <https://digitalcommons.law.byu.edu/lawreview/vol2014/iss1/7/>
- [3] Black, E. G., & Page, M. H. (1992). Add-On Infringements: When Computer Add-Ons and Peripherals Should (and Should Not) Be Considered Infringing Derivative Works Under *Lewis Galoob Toys, Inc. v. Nintendo of America, Inc.*, and Other Recent Decisions. *Hastings Comm. & Ent. LJ*, 15, 615, https://repository.uclawsf.edu/hastings_comm_ent_law_journal/vol15/iss3/4/
- [4] Ackerson, T. (2020). It's Time for the Copyright Act to Patch-in a Statutory License for Video Game Streaming. *AIPLA QJ*, 48, 325, <https://tind.wipo.int/record/45133>
- [5] Bonadio, E., & Trapova, A. (2024). Intellectual property law in gaming and artificial intelligence. In *Research Handbook on Property, Law and Theory* (pp. 448-462). Edward Elgar Publishing, doi:10.4337/9781802202069.00037
- [6] Abrams, M. S. (2013). The legal status of video games: comparative analysis in national approaches. *WIPO Magazine*, 7-96, <https://epap.by/storage/files/documents/insights/527b43b56a4e6.pdf>
- [7] Zhang, Y., Gan, Z., Fan, K., Chen, Z., Heno, R., Shen, D., & Carin, L. (2017, July). Adversarial feature matching for text generation. In *International conference on machine learning* (pp. 4006-4015). PMLR, doi:10.48550/arXiv.1706.03850
- [8] De Rosa, G. H., & Papa, J. P. (2021). A survey on text generation using generative adversarial networks. *Pattern Recognition*, 119, 108098, doi:10.1016/j.patcog.2021.108098
- [9] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, doi:10.48550/arXiv.2303.08774
- [10] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (pp. 38-45), doi:10.18653/v1/2020.emnlp-demos.6
- [11] Shaw, P., Uszkoreit, J., & Vaswani, A. (2018). Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*, doi:10.48550/arXiv.1803.02155
- [12] Peng, B., Li, C., He, P., Galley, M., & Gao, J. (2023). Instruction tuning with GPT-4. *arXiv preprint arXiv:2304.03277*, doi:10.48550/arXiv.2304.03277
- [13] Mao, R., Chen, G., Zhang, X., Guerin, F., & Cambria, E. (2023). GPTEval: A survey on assessments of ChatGPT and GPT-4. *arXiv preprint arXiv:2308.12488*, doi:10.48550/arXiv.2308.12488

저 자 소 개



김현수(Hyun-Soo Kim)

2019.02 : 단국대학교 소프트웨어학과 학사
2023.08 : 숭실대학교 AI·SW융합학과 석사
2019.01-현재 : 엘에스웨어(주) 소프트웨어연구소
연구개발본부 팀장
<주관심분야> 소프트웨어 공학, 딥러닝,
컴퓨터 비전, 분산신원증명, 빅데이터



최창준(Chang-Jun Choi)

2019.02 : 상명대학교 컴퓨터공학과 학사
2021.08 : 세종대학교 정보보호학과 석사
2021.09-현재 : 엘에스웨어(주) 소프트웨어연구소
연구개발본부 선임연구원
<주관심분야> 정보보호, 블록체인, 네트워크
보안, 분산신원증명, 저작권 기술



조용준(YongJoon Joe)

2011.03 : 큐슈대학교 전기정보공학과 학사
2013.03 : 큐슈대학교 정보학부 석사
2016.03 : 큐슈대학교 정보학부 박사 수료
2013.04-2016.03 : 일본 학술진흥원 특별연구원
2016.04-현재 : 엘에스웨어(주) 소프트웨어연구소
연구개발본부 기술이사
<주관심분야> 오픈소스, 저작권, 병렬·분산
컴퓨팅, 게임이론, 분산 제약 최적화 문제



신동명(Dong-Myung Shin)

2003.08 : 대전대학교 컴퓨터공학과 박사
2001-2006 : 한국정보보호진흥원(KISA)
응용기술팀 선임연구원
2006-2014 : 한국저작권위원회
저작권기술팀 팀장
2014-2016 : 한국스마트그리드사업단
보안인증팀 팀장
2016-현재 : 엘에스웨어(주) 소프트웨어연구소
연구개발본부 연구소장/상무이사
<주관심분야> 오픈소스 라이선스, 저작권 기술,
시스템/네트워크 보안, SW 취약점 분석·감정, 블
록체인 기술