

# 자연언어처리 기반 프로젝트 제안서

2376250 장수연

## 1. 프로젝트 개요

### 가. 목표

본 프로젝트는 일반 사용자가 법률 전문가의 도움 없이도 계약서의 불공정한 조항을 쉽게 식별할 수 있도록 돕는 AI 기반의 불공정 조항 분석 시스템을 개발하는 것을 목표로 합니다. 사용자가 특정 조항을 텍스트로 입력하면, 시스템은 AI 분석을 통해 해당 조항이 사용자에게 불리하다면 왜 불리한지, 또 어떤 법률에 근거하여 문제가 되는지를 종합적으로 설명하는 결과물을 생성하여 제공합니다.

### 나. 예상 결과물 (예시)

구분	내용
사용자 입력	<p>제 1 조(손해배상): 서비스 유료화 이후는 회사의 귀책사유로 인해 전체 서비스 장애시간이 24 시간 이상 되었을 경우 사용자는 해당 기간의 이용요금에 대하여 환불을 요구할 수 있다.</p>
시스템 출력	<p><b>분야:</b> 정보통신</p> <p><b>유불리 판단:</b> 사용자에게 불리한 조항입니다.</p> <p><b>불리한 이유:</b> 상당한 이유 없이 사업자의 손해배상 범위를 제한하거나 사업자가 부담하여야 할 위험을 고객에게 떠넘기는 조항이기 때문입니다.</p> <p><b>판단 근거 법령:</b> "제 6 조(일반원칙) ② 약관의 내용 중 다음 각 호의 어느 하나에 해당하는 내용을 정하고 있는 조항은 공정성을 잃은 것으로 추정된다. 1. 고객에게 부당하게 불리한 조항", "제 7 조(면책조항의 금지) ... 2. 상당한 이유 없이 사업자의 손해배상 범위를 제한하거나 사업자가 부담하여야 할 위험을 고객에게 떠넘기는 조항"</p>

## 2. 개발 계획 및 방법론

### 2-1 데이터셋 구축 및 법령 DB 준비

#### 가. 데이터셋

AI-Hub 의 '법률/규정 텍스트 분석 데이터' 중 약관 데이터를 사용합니다.

<https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=580>

주요 필드: clauseField(분야), clauseArticle(약관 조항), dvAntageous(유불리 판단),

illdcssBasiss(위법성 판단 근거), relateLaword(관련 법령).

#### 나. 법령 DB 준비

시스템이 사용자에게 제시하는 법령을 생성할 때는 환각 문제를 방지하기 위해 DB에 저장되어 있는 법령만을 응답하도록 할 예정입니다.

relateLaword 필드의 법 조항들을 따로 추출하여 DB에 저장합니다. 이 때, nlpai-lab의 KURE-v1 모델을 사용하여 각 법령 조항의 의미 벡터를 미리 계산하여 DB에 같이 저장합니다.

<https://huggingface.co/nlpai-lab/KURE-v1>

### 2-2 LLM 파인튜닝

참조 논문에서 높은 신뢰도 점수를 보인 Llama-3.1 모델을 위 데이터셋으로 파인튜닝하여 다음 세 가지 핵심 정보를 순차적으로 추출하도록 학습시킵니다.

1. 약관 분야 분류,
2. 불공정 여부 판단,
3. 판단 근거 추출

학습 프롬프트 구조:

- Instruction: 다음 약관 조항의 문맥을 이해하여 분야 분류, 불공정 여부 판단, 판단 근거를 요약하시오.
- Input: (약관 내용)
- Output: "분야: [분야] / 불공정여부: [판단] / 근거: [요약된 근거]

이 때 Output에 들어가는 데이터는 다음과 같이 수집할 예정입니다.

분야: clauseField 필드

불공정 여부: dvAntageous 필드 (1: 유리, 2: 불리)

근거: illdcssBasiss 필드

### 2-3 RAG 방식의 법령 검색 기능 구현

LLM에게 직접 법률 원문을 생성하도록 훈련시킬 수도 있으나, 법 조항을 직접 생성하도록 할 경우, 존재하지 않는 조항을 만들거나 법률 번호 등을 틀릴 위험이 크다고 판단하여 RAG 방식을 채택했습니다. 법률 분야의 특성상 환각을 사용자에게 제공하는 것을 원천적으로 방지해야 한다고 생각했기 때문입니다.

nlpai-lab의 KURE-v1 모델을 검색 임베딩 모델로 사용합니다. 사용자가 입력한 원본 약관 조항과 LLM 파인튜닝 결과로 생성된 '판단 근거'문장을 합쳐 검색 Query로 활용합니다. Query 벡터와 법령 DB에 저장된 모든 법령 벡터들 사이의 유사도를 계산하여, 가장 의미적으로 유사한 법령 조항 원문을 찾아냅니다.

### 2-4. 최종 답변 생성

2-2와 2-3의 과정을 거친 후, LLM에게 2-2의 output과 2-3에서 찾은 법령을 전달하고, LLM은 검색된 법령 원문을 Context로 활용하는 동시에, 다음의 Instruction을 따라 최종 보고서를 생성합니다.

Instruction: 다음은 최종 보고서를 생성하기 위한 입력 데이터이다. 이를 참조하여 아래에 제시된 JSON 형식과 키 이름에 맞추어 통합 분석 결과를 생성하시오.

```
{
    "분야": "[약관 분야]",
    "유불리_판단": "[유불리 결과]",
    "불리한_이유": "[불리한 이유 결과]",
    "판단_근거_법령": [
        {
            "조항_번호": "[법령 번호]",
            "내용": "[법령 원문 내용]"
        }
    ]
}
```

### 3. 성능 검증

LLM 성능 검증: AI-Hub에서 제공하는 Validation 데이터를 사용해 LLM의 분류 및 판단 태스크 성능을 검증합니다.

RAG 검색 성능 검증: LLM의 쿼리 생성 품질 및 검색 파이프라인의 유효성을 정량적으로 평가하기 위해 MAP (Mean Average Precision) 및 NDCG (Normalized Discounted Cumulative Gain) 지표를 사용하여 검색 정확도를 측정합니다.

### 4. 참고 사이트 및 논문

AI-Hub 데이터셋:

<https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=580>

임베딩 모델:

<https://huggingface.co/nlpai-lab/KURE-v1>

법률 도메인 LLM/BERT 연구:

<https://www.dbpedia.co.kr/pdf/pdfView.do?nodeId=NODE12267821&width=854>

<https://www.dbpedia.co.kr/pdf/pdfView.do?nodeId=NODE12318461>

<https://medium.com/lbox-team/legal->

[bert-%EB%B2%95%EB%A5%A0-%EB%8F%84%EB%A9%94%EC%9D%B8%EC%97%90-%ED%8A%B9%ED%99%94%EB%90%9C-%EC%96%B8%EC%96%B4%EB%AA%A8%EB%8D%B8-%EA%B0%9C%EB%B0%9C%EA%B8%B0-940d64a6f908](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC940d64a6f908/)