| **RAG 파이프라인 구축 프로세스 보고서 양식** | |
| --- | --- |
| **서비스 명  및 개요** | **서비스명**: 생활 법률 자문 챗봇  **서비스 개요**: 임대차 분쟁, 근로계약, 소비자 권리 등 다양한 생활법률 이슈에 대해 정보를 찾기 어렵고, 변호사 상담은 시간과 비용의 부담이 큰 상황  이러한 문제를 해결하고자 구축된 데이터셋을 활용하여, 사용자의 질문에 대해 정확하고 맥락에 맞는 답변을 제공하는 법률 자문 챗봇 서비스를 설계  사용자가 “전세 계약 해지 시 보증금 반환은 어떻게 되나요?”와 같은 질문을 하면, 관련 법령 조항과 유사한 판례, 정부기관의 Q&A를 기반으로 근거 있는 답변 제공 |
| **타겟 사용자 및 시장 분석** | **타겟 사용자**:  - **예상 사용자 유형**:  계약, 근로, 소비자 법적 문제를 고민으로 가지고 있는 사용자(일반 시민)  법적 상담을 해주는 법률 종사자  - **주요 요구사항 및 사용 목적**:  요구사항 : 문제 상황에 대한 설명  사용목적 : 문제 상황에 대해 법률 상담이 아닌 AI를 활용한 1차 판단  **시장 분석**:  AI 시장은 많은 기업들이 AI 개발에 투자하지만 수익성으로 이어지지 않고 있는 추세. AI가 많이 활용될 수 있는 곳으로 의료, 법률 분야가 대두 되고 있으며 해당 분야를 리걸테크(Legal Tech)라고 부름. 해외의 경우, 법률 AI 서비스하고 있는 회사들 존재. 하지만 국내의 경우, 법적 문제로 인해 법률 AI 서비스 개발에 투자를 못하고 있음. |
| **목표 및 기대효과** | **서비스 목표**:  - 생활 법률에 대한 접근성 향상  - 정확하고 신뢰할 수 있는 법률 정보 제공  - 사용자 질문에 빠르게 대응하여 분쟁 예방 및 정보 격차 해소  **기대효과**:   * 일반 시민, 프리랜서, 소상공인 등이 법률 전문가 도움 없이도 문제 상황에 대한 1차 판단 가능 * 공공기관, 고객 서비스 부서의 반복 질의 자동화를 통한 업무 효율 향상 * 사회 전체적으로 법률 정보에 대한 접근성과 이해도를 높여 디지털 법률문해력 향상에 기여 |
| **데이터**  **구성 및 활용** | **원천데이터 소스**: :AI 허브 제공 데이터 (민사법 LLM 사전학습 및 Instruction Tuning 데이터)  **원천 데이터 형식:** JSON 형식, Training / Validation Data, LLM 학습을 위한 질문 – 답변 쌍으로 구성  **데이터 처리 방법:**      1. 텍스트의 길이를 분석해 chunk 단위로 분해하여 임베딩  질문, 답변 데이터는 길이가 짧아 chunk 단위로 나눌 필요 없었음.  판결문 데이터의 경우 평균 길이가 10,000자, 최대 길이 150,000자로 임베딩을 위해 chunk 단위로 나눌 필요가 있었음.  => chunk size: 5000, overlap : 300 으로 데이터를 나눔  2. RAG 기법 활용을 위한 질문 데이터 임베딩  OpenAI의 “text-embedding-3-small” 모델을 이용해 임베딩  이후, PostgreSQL에 vector extension을 활용, 벡터 데이터 저장  사용자의 질문이 들어오면 임베딩한 질문 데이터를 기반으로 RAG 기법 활용 |
| **RAG  파이프라인 설계** | **데이터 최적화:**  Chunk Size: 5000  Overlap: 300  **벡터 데이터베이스 구축 및 임베딩:**  벡터 DB : PostgreSQL (vector extension 사용)  임베딩 모델 : text-embedding-3-small  **Retriever 및 Reranker 구현:**  VectorStore Retriever 이용  **LLM 프롬프트 설계 및 답변 생성, 평가 :**  **1/ Task : QA / 챗봇**  **2/ 프롬프트 예시**  **“””**  당신은 민법에 대해서 잘 알고 있는 변호사입니다.  고객의 질문에 상세하게 답을 해줄 수 있다면, 일반인이 알아들을 수 있게 풀어서 설명하세요.  관련 문서에 법조항이 있다면 법조항을 포함하여 답변을 해주세요. 법조항은 많이 언급할 수 있으면 좋습니다. 그렇지만 너무 관련이 없는 것을 언급하지는 마세요.  만약 질문에 대한 답변이 없다면 "해당 경우에 대한 판례가 없기 때문에 답변을 드리기 어렵습니다" 라고 답하세요.  판례를 찾을 수 있다면 판례를 포함하여 답변을 해주세요.  “””  **3/ 답변 생성 : OpenAI API**  Q: 손해배상을 받을 수 있는 경우와 그 요건은 무엇인지 설명해 주세요. A: 손해배상을 받을 수 있는 경우와 그 요건에 대해 설명드리겠습니다.  손해배상은 다른 사람의 불법행위로 인해 피해를 입은 경우, 그 피해를 보상받기 위한 법적 절차입니다. 민법 제750조에 따르면, "타인의 권리를 침해한 자는 그로 인하여 발생한 손해를 배상할 책임이 있다"고 규정되어 있습니다. 즉, 손해배상을 청구하기 위해서는 몇 가지 요건을 충족해야 합니다.  ### 손해배상의 요건  1. \*\*불법행위의 존재\*\*:  - 손해배상을 주장하기 위해서는 먼저 불법행위가 존재해야 합니다. 불법행위란 고의 또는 과실에 의해 타인의 권리를 침해하는 행위를 말합니다. 예를 들어, 누군가의 재산을 고의로 훼손하거나, 교통사고를 일으켜 타인의 신체에 상해를 입히는 경우가 이에 해당합니다.  2. \*\*실제 손해의 발생\*\*:  - 불법행위로 인해 실제로 손해가 발생해야 합니다. 손해는 재산적 손해(예: 물건의 파손, 치료비 등)와 비재산적 손해(예: 정신적 고통 등)로 나눌 수 있습니다. 예를 들어, 납품된 물품에 결함이 있어 손해가 발생하였다면, 그 결함으로 인해 실제로 피해가 발생했음을 입증해야 합니다.  3. \*\*인과관계의 입증\*\*:  - 손해와 불법행위 사이의 인과관계가 입증되어야 합니다. 즉, 피해자가 입은 손해가 가해자의 불법행위로 인해 발생했음을 증명해야 합니다. 예를 들어, 교통사고로 인해 부상을 입었다면, 그 부상이 사고로 인해 발생했음을 입증해야 합니다.  ### 요약  손해배상을 청구하기 위해서는 다음 세 가지 조건이 필요합니다:  1. 불법행위가 존재해야 합니다 (민법 제750조).  2. 불법행위로 인해 실제 손해가 발생해야 합니다 (민법 제751조).  3. 손해와 불법행위 사이의 인과관계가 입증되어야 합니다.  이러한 요건을 충족해야만 손해배상을 받을 수 있습니다. 만약 증거가 부족하거나 요건이 충족되지 않는 경우, 손해배상의 주장은 인정되지 않을 수 있습니다. |
| **RAG  파이프라인 평가 및 결과** | **평가방법**  **성능 평가용 데이터**  **o** 법률 질문-대답 쌍으로 이루어진 데이터셋 100개 (학습 데이터 50개 + 검증 데이터 50개)  **정량 평가 : RAGAS 평가 지표**  **o** faithfulness(충실성)  : 생성된 답변이 가지고 있는 지식으로 얼마나 뒷받침 되는 지에 대한 비율  **o** answer semantic similarity(답변 의미 유사성)  : 생성된 답변이 이상적인 답변과 얼마나 유사한 지  **o** context precision(문맥 정확도)  : 검색된 데이터 중에서 중에서 답변 생성에 도움이 되는 관련 정보가 얼마나 포함되어 있는지  **o** context recall(문맥 재현율)  : 검색된 데이터가 답변에 필요한 정보를 얼마나 포함하고 있는지  **정성 평가**  **평가 항목** - 정확성: 생성된 답변이 제품 매뉴얼의 내용과 얼마나 일치하는가? - 관련성: 답변이 검색된 문서/데이터와 관련이 있는가? - 명확성: 답변이 쉽게 이해되고 논리적으로 명확한가?  **평가 절차** - 각 질문에 대해 생성된 답변을 매뉴얼의 실제 내용을 기준으로 비교 및 검토 - GPT에게 평가 항목에 대한 정의와 실제로 평가한 예시를 주고 100개 전부에 대해 T or F 로 평가 데이터를 생성 - 생성된 평가 데이터에서 10개를 랜덤하게 뽑아 검증  **평가 결과**   * **정량 평가**     - context\_precision : 0.8825 - context\_recall : 0.6092 - faithfulness : 0.8476 - answer\_similarity : 0.7253   * **정성 평가**   - 정확성 : TRUE 75 FALSE 25 - 관련성 : TRUE 76 FALSE 24 - 명확성 : TRUE 83 FALSE 17 |
| **결론 및  향후 발전 방향** | **결론**  벡터 DB를 구축하고 이를 바탕으로 LLM을 통해 답변을 생성하는 RAG 파이프라인을 설계. 법률, 판례 데이터를 기반으로 사용자의 질문에 대한 신속하고 정확한 답변을 제공. 이를 바탕으로 고객의 법률 고민을 덜어주고 1차 판단의 근거를 제공해 문제 해결 시간을 단축시켜주는 서비스를 제공함.  **향후 발전 방향**  **-** 멀티턴 대화 지원  사용자의 이전 대화에 기반한 답변 생성으로 사용자 경험 개선  - 답변 성능 향상  데이터셋 확장, RAG Retriever 고도화를 통한 LLM의 답변 성능 향상  **데이터 추가 수집 혹은 업데이트**  - Retriever 고도화 예정  법률 도메인 데이터를 학습한 임베딩 모델을 활용해 데이터셋 조회 성능 향상  - 법령 범위 확대  민법, 지식재산권법 뿐만 아니라 형사법과 다른 법령 데이터도 추가하여 확장 예정  - 데이터셋 추가  AI Hub 뿐만 아닌 국가법령정보센터(<https://law.go.kr/>), 대법원([https://portal.scourt.go.kr/)](https://portal.scourt.go.kr/)%EC%97%90) 등에 있는 법률 데이터를 추가로 활용 |