

<슬라이드 1>

안녕하세요, AI코스 대전 3팀 **생활 법률 자문 챗봇 서비스**의 발표를 맡게 된 발표자 최인용입니다.
팀원 구성은 다음과 같습니다.

<슬라이드 2>

우선 목차입니다. 발표는 다음과 같은 순서로 진행하겠습니다.

<슬라이드 3>

첫 번째로 저희가 **생활 법률 자문 챗봇 서비스**를 기획하게 된 배경과 의도를 먼저 말씀드리겠습니다.
저희는 일상생활 속에서 임대차 분쟁이나 근로계약, 소비자 권리 등 법률과 관련된 문제를 자주 접하게 됩니다.

하지만 막상 관련 정보를 찾으려고 하면, 전문적인 법률 용어나 복잡한 절차 때문에 **정보를 찾는 것 자체가 어렵고 막막한 경우가 많습니다.**

또한, 변호사 상담은 짧게는 30 분, 많게는 수십만 원의 비용이 들어가는 경우도 있어서,
시간과 비용 면에서 진입장벽이 매우 높습니다.

이처럼 **일반 시민들의 법률 접근성이 부족한 현실** 속에서, 정보를 제대로 찾지 못하고 문제를 더 키우는 경우도 많습니다.

더불어, 팬데믹 이후로는 **24 시간 언제든지 접근할 수 있는 비대면 서비스**에 대한 수요도 점점 증가하는 추세입니다.

이런 점들을 종합적으로 고려해보았을 때,
생활 법률 정보를 손쉽게 제공하는 자동화된 비대면 상담 시스템,
바로 생활 법률 자문 챗봇이 필요하다고 판단했습니다.

<슬라이드 4>

타겟 사용자는 크게 두 부류로 나뉩니다.

첫 번째는,

계약 문제나 근로 문제, 소비자 분쟁 등 **생활 속 법적 고민을 가진 일반 시민들**입니다.

이들은 대부분 **법률 전문가가 아니기 때문에**, 문제 상황이 발생했을 때 어디서부터 정보를 찾아야 할지 막막해하는 경우가 많습니다.

두 번째는,

이러한 시민들에게 상담을 제공하는 **법률 종사자**입니다.

이들은 반복적인 초기 문의에 시간을 많이 소모하게 되는데,

이를 효율적으로 정리해주는 도구가 필요합니다.

그렇다면, 이 사용자들이 바라는 점은 무엇일까요?

우선, 일반 사용자는 **법적 고민에 대해 1 차적으로 방향을 제시해줄 수 있는 서비스**를 원합니다.

꼭 변호사가 아니더라도, 챗봇을 통해 문제의 유형이나 필요한 절차 정도는 스스로 파악하고 싶은 거죠.

또한, 앞서 말씀드렸듯이 **법률 접근성 부족과 정보 비대칭 문제** 역시 큰 요구사항 중 하나입니다.

이 문제를 해소하기 위해, 저희는 챗봇이 사용자의 질문에 대해 명확하고 신뢰도 있는 답변을 제공할 수 있도록 기획했습니다.

<슬라이드 5>

저희가 사용한 데이터는 **AI Hub의 민사법 LLM Instruction Tuning용 데이터셋**입니다.
해당 데이터셋은 판결문 기반으로 구성되어 있으며, 주제별 빈도를 분석한 결과 다음과 같습니다.

왼쪽의 바 차트는 84587 개의 전체 데이터에 대하여 사건명을 기준으로 빈도를 정렬한 것이고,
오른쪽 파이차트는 주요 사건 유형의 비율을 시각화한 것입니다.

가장 많은 비중을 차지하는 항목은 **손해배상**으로 전체의 약 33.5%를 차지하고 있고,
그 뒤로 대여금, 부당이득금, 구상금, 공사대금 등 금전 관련 사건들이 높은 비중을 보이고 있습니다.

<슬라이드 6>

다음은 질문과 답변 텍스트의 길이 분포입니다.

좌측은 질문 텍스트 분포인데, 평균적으로 **30~50 자** 사이에 밀집되어 있고,
우측의 답변 텍스트는 평균 **150~250 자**로 분포되어 있습니다.

이러한 분포를 기반으로, 모델 입력에 적합한 **최적 토큰 범위**를 추정하여 프롬프트 설계 및 사전학습 토큰 구성에 반영했습니다.

<슬라이드 7>

저희가 활용한 판결문 데이터는
평균 텍스트 길이가 약 **10,000 자**, 많게는 **15 만 자**에 이르기 때문에
데이터를 직접 임베딩하기에는 **길이 제한과 효율성 측면에서 어려움이 있었습니다.**

이에 따라, 텍스트를 **chunk 단위로 분할**하는 전처리 과정을 진행했습니다.

구체적으로는 **chunk size** 는 5,000 자, **overlap** 은 300 자로 설정하여,
의미 단절을 최소화한 슬라이딩 윈도우 방식으로 데이터를 나눴습니다.

이렇게 분할된 각 chunk 는 이후 벡터화 및 저장 작업에 활용됩니다.

이 과정을 통해 판결문 전체의 맥락을 잃지 않으면서도,
효율적인 정보 검색 및 응답 생성을 위한 기반 데이터를 준비했습니다.

<슬라이드 8>

저희가 개발한 **생활 법률 자문 챗봇**은,
일상생활에서 발생할 수 있는 임대차 분쟁, 근로계약 문제, 소비자 권리 침해 등과 같은
법률 고민에 대해 **빠르고 신뢰할 수 있는 1 차 판단을 제공하는 서비스**입니다.

사용자는 챗봇에게 자연어로 질문을 입력하면,
챗봇은 관련 법률 정보나 판례를 바탕으로
간단하고 이해하기 쉬운 답변을 제공합니다.

이 서비스를 통해
법률 지식이 부족한 일반 시민도 보다 쉽게 정보에 접근할 수 있고,
상담 전 단계에서 **문제 해결 방향을 미리 가늠해볼 수 있는 도구**로 활용할 수 있습니다.

<슬라이드 9>

이 슬라이드는 저희 **생활 법률 자문 챗봇의 전체 시스템 아키텍처**를 보여줍니다.
왼쪽의 클라이언트부터 오른쪽의 AI 응답 생성까지, 전체 요청 흐름은 다음과 같은 순서로
이루어집니다.

가장 먼저,
사용자(Client)는 **Vue.js**로 구성된 **프론트엔드 앱**을 통해 질문을 입력합니다.
이 요청은 **AWS EC2** 상에 배포된 **백엔드 서버**로 전달됩니다.

EC2 내에서는
요청이 **Nginx**를 거쳐 **Gunicorn**, 그리고 **FastAPI 애플리케이션**으로 전달됩니다.

FastAPI 앱 내부에서는 사용자의 질문을 처리하기 위해
먼저 **PostgreSQL 기반 Vector DB**에 질의를 수행합니다.
이때, 질문은 임베딩되어 벡터 형태로 변환되며,
벡터 DB에서는 질문과 가장 유사한 **상위 4 개의 관련 데이터**를 검색해옵니다.

검색된 데이터와 사용자 질문은 함께
OpenAI의 LLM API (GPT-4o-mini)에 전달되며,
LLM은 이 정보를 바탕으로 최종 답변을 생성하게 됩니다.

생성된 응답은 다시 FastAPI → Gunicorn → Nginx → Vue.js 순으로 전달되어,
사용자에게 최종 결과가 출력됩니다.

<슬라이드 10>

이제 저희 서비스의 **주요 기능과 기술적 구현 방식**에 대해 설명드리겠습니다.

우선 전체 시스템은 **LLM, Vector DB, RAG, 클라우드 컴퓨팅**의 네 축을 기반으로 구성되어 있습니다.

먼저, 사용자의 질문이 들어오면
OpenAI의 **GPT-4o-mini** 모델을 통해 자연어를 이해하고,
적절한 프롬프트 형식으로 변환하여 처리합니다.

하지만 LLM 특성상, 학습되지 않은 정보에 대해서는
환각(Hallucination) 문제가 발생할 수 있기 때문에,
이를 보완하기 위해 **RAG(Retrieval Augmented Generation)** 기법을 도입했습니다.

사용자의 질문은 embedding 과정을 거쳐,
벡터 형태로 변환되어 **PostgreSQL 16.9 + vector extension**을 활용한 Vector DB에서
관련 context를 검색한 후, 이를 기반으로 답변을 생성합니다.

이렇게 하면, **판례나 법령 등 실제 문서에 기반한 응답 생성이 가능해지고**,
기존 LLM 단독 응답 대비 신뢰성과 정확도가 높아집니다.

또한, 클라우드 환경을 통해
24시간 365일 무중단 서비스 제공,
보안성과 확장성 확보,
그리고 **원격 서버를 활용한 안정적인 배포 환경도** 갖추고 있습니다.

마지막으로,
이 모든 기술 요소들이 통합되어
사용자의 법률 질문에 대해 **판례 기반의 빠르고 신뢰도 있는 답변**을 생성하는 것이
저희 시스템의 핵심 구현 구조입니다.

<슬라이드 11>

다음은 저희 챗봇의 성능을 정량적으로 평가하기 위해 사용한 **4 가지 주요 지표**를 정리한 내용입니다.

첫 번째는 **Faithfulness**, 즉 **충실성**입니다.

이 지표는 생성된 답변이 검색된 context 의 정보에 **얼마나 기반하고 있는지**를 측정합니다.

LLM 이 문맥과 동떨어진 답변을 생성하지 않도록 하기 위한 핵심 기준입니다.

두 번째는 **Answer Semantic Similarity**, 즉 **답변 의미 유사성**입니다.

이는 생성된 답변이 **이상적인 정답과 얼마나 유사한지**,

즉 의미적 거리를 코사인 유사도로 측정한 지표입니다.

세 번째는 **Context Precision**, **문맥 정확도**입니다.

검색된 context 중에서 실제로 **답변 생성에 직접 사용된 유의미한 정보가 얼마나 포함되어 있는지**를 평가합니다.

즉, 불필요한 context 를 얼마나 줄였는지를 보는 지표라고 보시면 됩니다.

네 번째는 **Context Recall**, **문맥 재현율**입니다.

답변에 필요한 핵심 정보가 context 에 **얼마나 빠짐없이 포함되었는지**를 측정하는 지표입니다.

이 네 가지 지표는 각각

정확도, 문맥 활용도, 답변 근거의 충실성을 평가하기 위해 함께 사용되며,

RAG 기반 시스템의 **정확성 개선 여부를 객관적으로 판단하는 기준으로** 활용됩니다.

<슬라이드 12>

다음은 저희가 진행한 **정량 평가 및 성능 개선 결과**입니다.

먼저, 초기 평가에서는 **Context Precision** 이 0.49, **Context Recall** 은 0.41 로,
두 지표 모두 상대적으로 낮은 수치를 보였습니다.

이 지표들은 각각, 검색된 context 에 **답변 생성에 유의미한 정보가 얼마나 포함되었는지**,
그리고 **context 가 전체적으로 얼마나 필요한 정보를 커버했는지**를 나타냅니다.

즉, 초기에 설정한 프롬프트 구조에서는
검색된 문서가 LLM 의 답변 생성에 충분히 활용되지 못하고 있다는 한계가 있었습니다.

<슬라이드 13>

이에 따라, 프롬프트를 다음과 같이 개선했습니다.

“당신은 민법에 대해 잘 알고 있는 변호사입니다. 고객의 질문에 상세하게 답해줄 수 있다면, 일반인이 이해할 수 있도록 설명하세요. 관련 법조항이나 판례가 있다면 반드시 포함해주세요. 답변이 불가능한 경우 그 이유를 명확히 전달해주세요.”

이러한 구조는 **답변의 명확성, 구체성, 맥락 반영 정도**를 높이도록 설계되었습니다.

<슬라이드 14>

그 결과,

- **Context Precision** 은 0.49 → 0.88,
- **Context Recall** 은 0.41 → 0.61 로 크게 개선되었고,
- **Faithfulness** 도 0.80 → 0.85 로 소폭 상승했습니다.
- **Answer Similarity** 는 0.72 로 동일한 수준을 유지했습니다.

이러한 결과는, 단순히 RAG 기법을 적용하는 것에 그치지 않고
프롬프트의 질 또한 성능 향상에 결정적인 영향을 준다는 점을 보여주는 지표라 할 수 있습니다.

<슬라이드 15>

이번에는 **모델 성능에 대한 정성 평가 결과**를 소개드리겠습니다.

정성 평가는 총 세 가지 항목으로 구성됩니다.

바로 **정확성**, **관련성**, 그리고 **명확성**입니다.

먼저 **정확성**은

생성된 답변이 실제 문서나 판례 등 참고한 자료와 **얼마나 일치하는지**를 평가합니다.

평가 결과, 전체의 약 **75%가 정확하다고** 판단되었습니다.

두 번째는 **관련성**입니다.

답변이 실제 검색된 context와 **얼마나 밀접하게 연결되어 있는지**를 확인한 항목입니다.

이 항목에서는 대부분의 답변이 **문맥적으로 적절한 정보를 기반으로 생성되었음을** 확인할 수 있었습니다.

마지막으로 **명확성**은,

답변이 일반 사용자에게 **이해하기 쉽고 논리적으로 잘 정리되어 있는지**를 평가한 기준입니다.

여기에서는 **83%의 응답이 명확하다고** 평가되었습니다.

이처럼 정성 평가 결과는

정량 평가로는 놓치기 쉬운 **사용자 관점의 품질 요소들 — 즉, 이해도, 신뢰도, 적합성 — 을 확인하는 데 유의미한 역할**을 합니다.

또한 이 결과를 통해, 저희 챗봇이 단순히 기술적으로 동작하는 수준을 넘어,

실제로 활용 가능한 수준의 응답 품질을 갖추고 있음을 입증할 수 있었습니다.

<슬라이드 16>

앞서 말씀드린 정량·정성 평가에 이어,
이번에는 **RAG 기법의 적용 유무에 따른 실제 답변 차이**를 비교해보겠습니다.

먼저 이 예시는 "명예훼손죄에서 공연성이 인정되기 위한 요건"에 대한 질문입니다.

왼쪽은 **RAG 적용 답변**,
오른쪽은 **RAG 미적용 답변**입니다.

RAG 를 적용했을 경우,
챗봇은 질문에 대해 단순히 일반적인 개념을 설명하는 수준을 넘어서,
관련 법조항(형법 제 307 조)과 함께 **구체적인 대법원 판례**까지 인용하며,
답변의 **정확도와 신뢰도 모두를 충족**하고 있습니다.

<슬라이드 17>

예를 들어,
“대법원 2000. 5. 16. 선고 99도 5622 판결”이나
“2011. 9. 8. 선고 2010도 7497 판결”을 통해
공연성 판단 기준이 실제로 어떻게 적용되는지 설명하고 있죠.

<슬라이드 18>

반면, RAG 를 적용하지 않은 경우에는,
LLM 모델이 제시한 판례를 검색한 결과,
정확한 조문이나 판례가 아닌 사례를 제시하였으며,
답변도 **다소 일반적이고 추상적인 수준**에 머무르게 됩니다.

즉, 두 답변 모두 일견 그럴듯해 보일 수는 있지만,
법률 서비스라는 특성상 “**근거 없는 설명**”은 신뢰할 수 없는 답변이 되기 쉽습니다.

이러한 비교는
RAG 가 단순히 성능 수치를 올려주는 도구가 아니라,
실제 서비스 품질을 결정짓는 핵심 요소임을 보여줍니다.

<슬라이드 19>

마지막으로, 저희 서비스가 가져올 **기대효과와 향후 계획**에 대해 말씀드리겠습니다.

먼저 **기대효과**입니다.

이 챗봇은 **일반 시민, 프리랜서, 소상공인** 등
법률에 대한 접근이 쉽지 않은 사용자들에게
비용 부담 없이 1차 판단 자료를 제공할 수 있습니다.

이를 통해 법률 상담 이전에
문제의 성격과 대응 방향을 스스로 가늠할 수 있는 기회를 제공하고,
정보 격차와 디지털 법률 문해력 문제를 해소하는 데 기여할 수 있습니다.

또한 공공기관이나 고객 서비스 부서처럼
반복적인 법률 문의가 많은 곳에서는
업무 효율화에도 실질적인 도움을 줄 수 있습니다.

그리고 **향후 계획**으로는 크게 세 가지를 준비하고 있습니다.

첫째, **멀티턴 대화** 기능을 도입해
사용자의 **이전 질문과 맥락을 반영한 응답** 생성을 구현할 예정입니다.
이를 통해 챗봇 사용 경험을 더 자연스럽게 몰입감 있게 개선할 수 있습니다.

둘째, **데이터셋 확장**입니다.
현재는 AI Hub의 민사법 데이터로 중심으로 운영되고 있지만,
앞으로는 **국가법령정보센터, 대법원, 국회 등 공공 데이터셋**을 추가로 수집하여
답변의 폭과 정확도를 함께 높일 계획입니다.

셋째, **Retriever 고도화**입니다.
법률 도메인에 특화된 임베딩 모델을 활용해,
더 정밀하고 정합성 높은 context retrieval을 가능하게 할 예정입니다.

저희는 이 프로젝트를 통해,
생활 속 법률 고민을 덜어주고,
누구나 쉽게 접근할 수 있는 실용적인 리걸테크 서비스로 확장해 나가고자 합니다.