

DATA 3 팀 프로젝트 최종 보고서

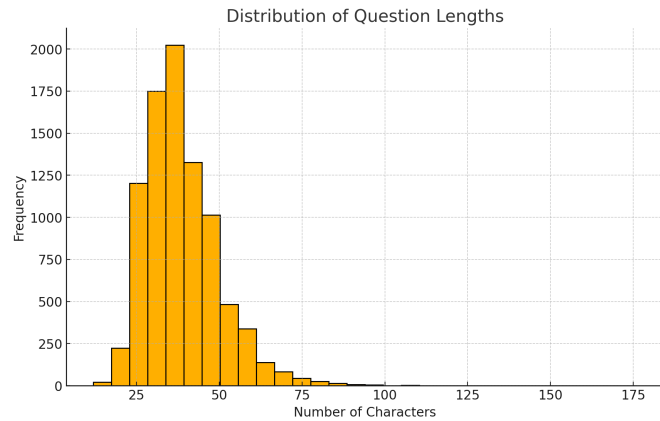
서비스 명 및 개요	<p>서비스명: HyDrive Chatbot – 현대자동차 차량 매뉴얼 기반 AI 챗봇</p> <p>서비스 개요:</p> <ul style="list-style-type: none"> ● 서비스가 해결하려는 문제점(Problem Statement): 현대자동차 차량은 기능이 점점 복잡해지고 있음에도 불구하고, 사용자들은 차량 매뉴얼의 방대한 분량과 어려운 용어로 인해 필요한 정보를 빠르게 찾지 못하고 있습니다. 또한 똑같은 기능이라도 차종에 따라 다르게 작동하는 경우 일일이 해당 매뉴얼을 찾아봐야 하는 번거로움이 있습니다. ● 제공하고자 하는 주요 기능 및 가치: 자연어 기반 질문에 대해 매뉴얼 내용을 빠르게 검색하여 직관적이고 신속한 응답을 제공합니다. 차량 모델별 특성을 반영한 사용자 맞춤형 정보를 제공하여 운전자가 다양한 기능을 이용할 수 있도록 합니다.
타겟 사용자 및 시장 분석	<p>타겟 사용자:</p> <ul style="list-style-type: none"> ● 일반 운전자: 경고등 확인, 블루투스 연결법, 스마트 크루즈 설정법 등 ● 즉각적인 차량 기능 질문 ● 고객 서비스팀: 고객 문의 응답을 빠르게 처리하고 신뢰도 있는 근거 제공 - 정비사/전문가: 매뉴얼 기반 고장 코드/점검 절차 확인, 정비 매뉴얼 링크 ● 확인 ● 고령자 · 초보 운전자: 어려운 용어를 쉽게 풀어 설명받고 안전 기능 숙지 ● 시장 분석: ● 자동차 기능의 복잡화: ADAS, HUD, 커넥티드카, OTA 등 운전자 보조 및 연결 기능이 증가하면서, 설명해야 할 기능과 사용자 문의가 폭발적으로 증가함 ● 문서 기반 고객지원의 한계: PDF 매뉴얼은 검색 효율이 낮고, 고객센터 인력도 반복적 질문 대응에 리소스 소모가 큼 ● 모바일/웹 기반 정보 소비 트렌드: 사용자는 앱 또는 챗봇 인터페이스를 통한 즉답(instant answer)을 선호함 ● 자동차 서비스 경험 차별화 경쟁: 주요 브랜드는 차량 구매 이후의 서비스 경험 개선을 위한 디지털 인터페이스 고도화에 집중하고 있음
목표 및 기대효과	<p>서비스 목표:</p> <ul style="list-style-type: none"> ● 현대 차량의 매뉴얼 정보를 AI 기반으로 구조화하여, 누구나 쉽게 접근 가능한 QA 챗봇을 구축 ● 질문-응답 시스템의 정확도, 반응속도, 사용자 만족도를 최적화한 차량 지식 서비스 구현 <p>기대효과:</p> <ul style="list-style-type: none"> ● 사용자에게 제공되는 이점 <ul style="list-style-type: none"> ○ 차량 매뉴얼 정보 접근성 향상 → 안전 운행 및 유지보수 효율 증가 - 반복적인 고객센터 문의 감소 → 시간 절약 및 비용 절감 ○ 고령자/초보 운전자도 직관적으로 차량 기능 파악 가능 ● 조직/팀/사회 기여 효과 <ul style="list-style-type: none"> ○ 고객 대응 인력의 효율 향상 → 업무 자동화 및 인력 비용 절감 ○ 사용자 자가 진단 및 경고 대응 가능 → 사고 예방 및 유지관리 효율

	<ul style="list-style-type: none">○ 증가○ 차량 정보 접근성 강화 → 브랜드 이미지 제고 및 고객 만족도 상승
데이터 구성 및 활용	<p>원천데이터 소스: 현대 자동차 제품 매뉴얼 문서 20 개</p> <ul style="list-style-type: none">● 기능이 다양해진 최신 차량 매뉴얼 중심● [('그랜저', 2025), ('그랜저', 2026), ('그랜저 Hybrid', 2025), ('그랜저 Hybrid', 2026), ('싼타페', 2024), ('싼타페', 2025), ('쏘나타 Hybrid', 2024), ('쏘나타 Hybrid', 2025), ('아반떼', 2025), ('아반떼', 2026), ('코나', 2024), ('코나', 2025), ('코나 Electric', 2023), ('코나 Electric', 2024), ('코나 Electric', 2025), ('투싼', 2024), ('투싼', 2025), ('투싼 Hybrid', 2024), ('투싼 Hybrid', 2025), ('팰리세이드 Hybrid', 2026)] <p>원천 데이터 형식: PDF 문서 파일</p> <p>데이터 처리 방법:</p> <ul style="list-style-type: none">● 데이터 수집: 현대차 매뉴얼 홈페이지에서 Selenium 을 사용한 웹 크롤링을 통해 제품 매뉴얼 데이터 확보● 데이터 전처리:<ul style="list-style-type: none">○ PyMuPDF 활용하여 PDF 파일을 구조화된 JSON 파일로 변환.○ 이 과정에서 목차 페이지 탐색 후 목차 추출○ 제목별 섹션 내용을 본문에서 추출○ 내용 정리 후 JSON 으로 저장○ 변환한 데이터에서 HTML 태그와 문법 정제하여 순수 텍스트 데이터로 변환 <p>QA 데이터 생성 방법:</p> <ul style="list-style-type: none">● QA 데이터셋 생성<ul style="list-style-type: none">○ 20 개의 변환된 매뉴얼 데이터를 Upstage Solar LLM 을 사용하여 QA Pair 생성○ 미리 목차를 기반으로 사용자 질문 유형을 분류하여, 해당 질문 유형에 필요할 것 같은 목차들을 샘플링<ul style="list-style-type: none">■ 고객 질문 유형 분류 (카테고리화)<ul style="list-style-type: none">● 기능 설명 (편의장치, 시동 및 주행, 운전자 보조)● 문제 해결 (클러스터의 경고등 및 표시등, 비상시 응급 조치)● 차량 관리 방법 (안내 및 차량 정보, 안전 및 주의사항, 정기 점검)○ 총 15198 개의 데이터셋을 생성● 중복 제거<ul style="list-style-type: none">○ 자동차 매뉴얼 특성 상, 차종 구분 없이 안내되는 내용과 차종 특성에 따라 달라지는 내용으로 구성됨○ 공통되는 내용은 model 구분 없이, 차별화되는 내용은 model 을 포함하여 QA Pair 을 저장하여 중복되는 데이터를 처리

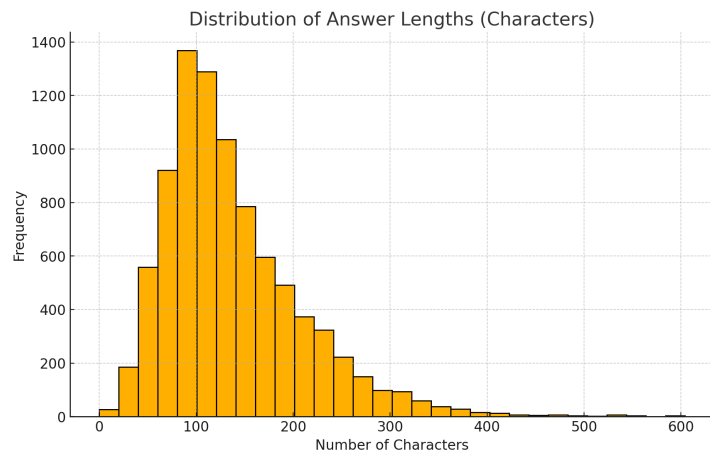
○ 8705 개의 최종 데이터셋을 구축

데이터 통계:

- 데이터 개수: 8705 개
- 데이터 길이 통계:
 - Question
 - 평균 길이: 약 38.6 자
 - 길이 분포

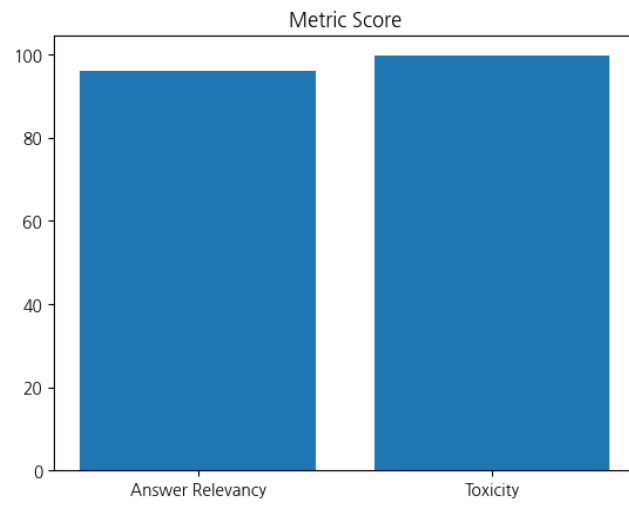


- Answer
 - 평균 길이: 137.5 자
 - 길이 분포

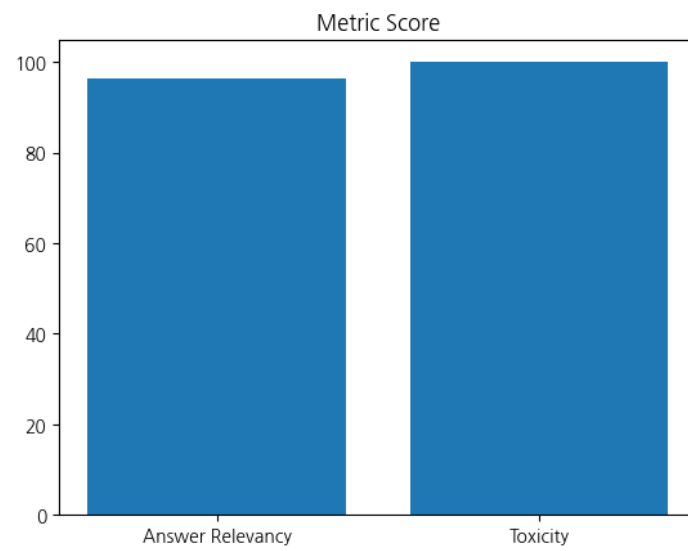


- QA 데이터 개수
 - category 별
 - 기능 설명 5393
 - 차량 관리 방법 2428
 - 문제 해결 880
 - 기타 4
 - 제품 별

	<div data-bbox="526 268 1292 716"><p>제품별 QA 개수</p><table border="1"><thead><tr><th>제품 모델</th><th>QA 개수</th></tr></thead><tbody><tr><td>전차형</td><td>2050</td></tr><tr><td>특판 (2025)</td><td>450</td></tr><tr><td>특판 (2024)</td><td>450</td></tr><tr><td>전차형 (2025)</td><td>400</td></tr><tr><td>특판 Hybrid (2025)</td><td>380</td></tr><tr><td>코나 (2025)</td><td>380</td></tr><tr><td>특판 Hybrid (2024)</td><td>380</td></tr><tr><td>헬리콥터 Hybrid (2024)</td><td>380</td></tr><tr><td>전차형 (2024)</td><td>380</td></tr><tr><td>코나 Electric (2026)</td><td>350</td></tr><tr><td>코나 Electric (2023)</td><td>350</td></tr><tr><td>코나 Hybrid (2026)</td><td>350</td></tr><tr><td>코나 Hybrid (2025)</td><td>350</td></tr><tr><td>코나 Electric (2024)</td><td>320</td></tr><tr><td>코나 Hybrid (2025)</td><td>320</td></tr><tr><td>코나 Hybrid (2025)</td><td>320</td></tr><tr><td>코나 Hybrid (2025)</td><td>320</td></tr><tr><td>코나 Hybrid (2024)</td><td>320</td></tr><tr><td>코나 (2024)</td><td>320</td></tr><tr><td>아반떼 (2026)</td><td>150</td></tr><tr><td>아반떼 (2025)</td><td>100</td></tr></tbody></table></div>	제품 모델	QA 개수	전차형	2050	특판 (2025)	450	특판 (2024)	450	전차형 (2025)	400	특판 Hybrid (2025)	380	코나 (2025)	380	특판 Hybrid (2024)	380	헬리콥터 Hybrid (2024)	380	전차형 (2024)	380	코나 Electric (2026)	350	코나 Electric (2023)	350	코나 Hybrid (2026)	350	코나 Hybrid (2025)	350	코나 Electric (2024)	320	코나 Hybrid (2025)	320	코나 Hybrid (2025)	320	코나 Hybrid (2025)	320	코나 Hybrid (2024)	320	코나 (2024)	320	아반떼 (2026)	150	아반떼 (2025)	100
제품 모델	QA 개수																																												
전차형	2050																																												
특판 (2025)	450																																												
특판 (2024)	450																																												
전차형 (2025)	400																																												
특판 Hybrid (2025)	380																																												
코나 (2025)	380																																												
특판 Hybrid (2024)	380																																												
헬리콥터 Hybrid (2024)	380																																												
전차형 (2024)	380																																												
코나 Electric (2026)	350																																												
코나 Electric (2023)	350																																												
코나 Hybrid (2026)	350																																												
코나 Hybrid (2025)	350																																												
코나 Electric (2024)	320																																												
코나 Hybrid (2025)	320																																												
코나 Hybrid (2025)	320																																												
코나 Hybrid (2025)	320																																												
코나 Hybrid (2024)	320																																												
코나 (2024)	320																																												
아반떼 (2026)	150																																												
아반떼 (2025)	100																																												
<div data-bbox="255 1187 399 1299">데이터 품질 평가 및 결과</div>	<div data-bbox="430 873 654 918">데이터 품질 평가 방식:</div> <div data-bbox="478 940 1276 1433"><ul style="list-style-type: none">● QA 평가셋 구성 방식:<ul style="list-style-type: none">○ 정량 평가(Deepeval 활용)<ul style="list-style-type: none">■ 샘플링 : 무작위 + 다중 층화 방식으로 각 300 건 평가■ 평가 지표:<ul style="list-style-type: none">■ Answer Relevancy■ Toxicity○ 정성 평가 방식(gpt 기반 자동 평가)<ul style="list-style-type: none">■ 전체 데이터셋의 10% (870 건) 대상■ 평가 항목:<ul style="list-style-type: none">■ 매우 적절함: 질문과 답변이 명확하게 부합함■ 부분적으로 적절함: 일부 정보는 유효하지만 불완전하거나 맥락이 약함■ 관련 없음: 질문과 답변이 명확히 불일치함</div> <div data-bbox="430 1456 654 1500">데이터 품질 결과 보고:</div> <div data-bbox="478 1523 750 1635"><ul style="list-style-type: none">● metric score<ul style="list-style-type: none">○ 랜덤 샘플링</div>																																												



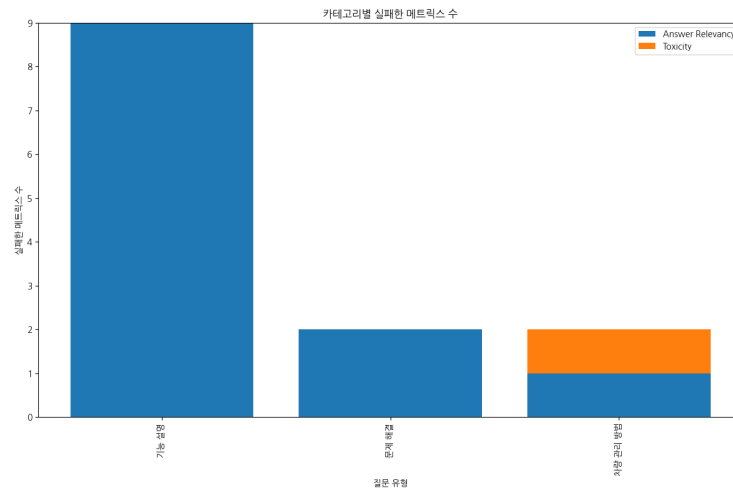
○ 다중 층화 샘플링



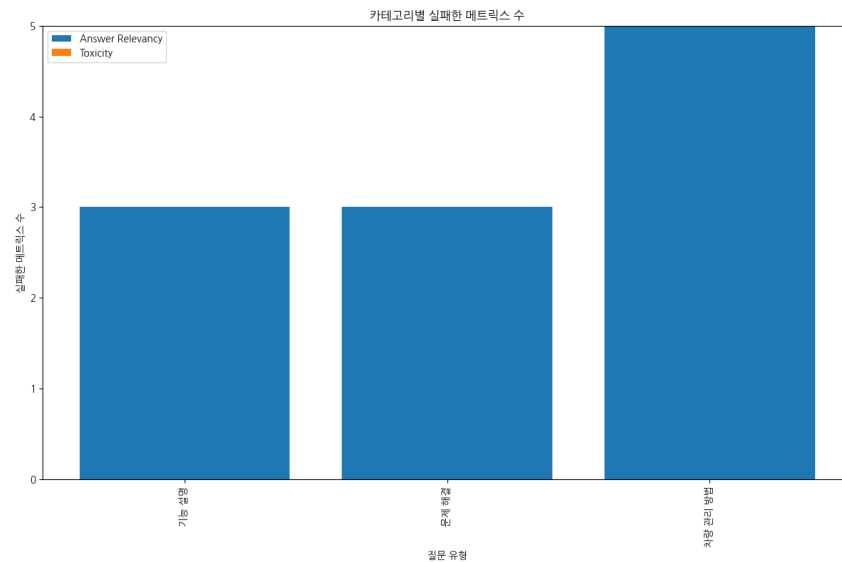
● 카테고리별 평가

○ metrics 별 부적격 데이터 수

■ 랜덤 샘플링

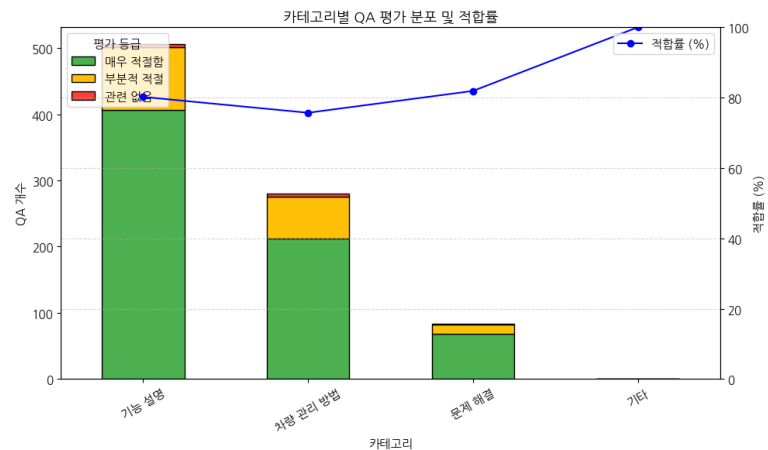


■ 다중 층화 샘플링



● 정성 평가

- 매우 적절함: 687 개 (78.9%)
- 부분적으로 적절함: 172 개 (19.8%)
- 관련 없음: 11 개 (1.3%)
- 전체적으로 98.7%의 QA 가 질문과 일정 부분 이상 부합함으로써 기본 품질 수준 확보
- 기능 설명: 가장 많은 QA 를 보유하면서도 높은 적합률 (약 80.2%)
- 차량 관리 방법: 75.7% 매우 적절함으로, 양호한 품질
- 문제 해결: 전체 수는 적지만 81.9% 매우 적절함 → 효율적인 QA 구성
- 기타: 샘플 수가 너무 적어 평가 불가
- 문제 해결 카테고리는 QA 수는 적지만 가장 높은 품질률을 보였고, 기능 설명은 전체 QA 중 가장 많은 수를 차지하면서도 품질 역시 우수



● 평가 결과

- 전체적인 품질 수준은 70~100% 범위에서 유지되며, QA 생성 품질은 전반적으로 기본 이상 수준 확보
- Answer Relevancy와 Toxicity 메트릭 기준으로 정량 평가 수행
- HallucinationMetric은 별도로 측정되지 않았으며, 주요 메트릭은 Answer Relevancy 중심
- QA 개수는 전 차종 공통 QA 2,055 건을 포함해, 투싼, 쏘나타페, 코나, 그랜저, 쏘나타 등 주요 모델에 300~400 개 수준으로 균등하게 분포
- 카테고리별 현황:
 - 기능 설명 : 가장 많은 QA 보유 (5,393 개) → 사용자 질문 빈도가 높고 성과도 양호
 - 차량 관리 방법 : 2,428 개로 중간 수준
 - 문제 해결 : 880 개로 가장 적음, 일부 카테고리에서 품질 보강 필요

품질 개선 사항 및 개선 방법:

● 단기 개선 과제

- 카테고리별 미흡 항목 보완: 문제해결 관련 QA가 적고 불완전한 경우 존재 → 품질 및 다양성 향상 필요
- 정량 메트릭 정확도 향상: Answer Relevancy 기준 외, Hallucination 관련 평가 지표 추가 필요
- 부분적으로 적절한 QA 점검:
 - 설정 단계 누락, 불완전한 설명 등 → 검토 및 보완 필요
 - 관련 없음 QA 제거 또는 수정:
 - 문맥 일치하지 않는 응답 사례는 우선 제거 조치
- 카테고리별 컨텍스트 보강:
 - 차량 관리 방법 등 일부 항목에서 답변의 정보 밀도가 낮음

● 중장기 발전 방향

- QA 카테고리별 표준 강화: 기능 설명, 문제 해결, 차량 관리 등 각 카테고리별로 프롬프트를 설계하여 명확한 QA 생성 기준 수립
- 우수 사례 학습 및 피드백 루프 구축: '매우 적절함' QA 활용하여 데이터 학습/보정
- 제품별 품질 모니터링 확대: QA 수가 많은 모델(예: 투싼, 쏘나타페, 그랜저 등)에 대해 정기적 성능 분석 체계 마련

	<ul style="list-style-type: none">○ 자동화된 평가 파이프라인 구축: QA 생성 → 메트릭 평가 → 수동 검수까지 자동화된 평가 루프 설계 필요○ 모델별 매뉴얼 다양성 확보 필요 : 현재는 최신 년도 매뉴얼을 사용하고 있지만 구입 후 보유 기간이 긴 차량의 특성 상, 오래된 연도 모델의 매뉴얼도 추가할 필요 있음● 품질 관리 체계 개선<ul style="list-style-type: none">○ 정기적인 품질 점검 체계화: LLM 기반 QA 의 지속적 성능 저하 방지 위해 주기적 품질 리포트 및 검토 체계 필요○ 데이터 생성 표준화: PDF 구조화 → QA 생성 → 중복 제거 → 분류까지의 워크플로우 고도화○ 카테고리/제품군별 관리 전략 수립: 사용자 활용도가 높은 모델 중심으로 차별화된 QA 전략 적용
--	--