

자동차 법적 규제 준수 여부 판단을 위한 LLM 개발 프로젝트

타겟 도메인 및 법체계 선정과
구현 기능 선정을 위한 LLM 평가 방안 제시

CarBot

빅데이터 종합설계
with 현대자동차

발표자:

안성규, 박상범,
김세희, 문버리

발표일:

2024년 5월 27일

목차

- 이전 발표 리뷰
- 부품 선정
 - 리콜 사례 세부 분석
 - 최신 기술 분야
 - 해석의 모호함
 - 정리 및 결론
- 법률 선정
 - 유럽 - UN ECE
 - 미국 - NHTSA
 - 한국 - 국제 조화 추진 체계
 - 정리 및 결론
- 기능 선정 및 LLM 성능 평가 계획

지난주 리뷰

국내 자동차 관련 법률 리서치

- 자동차관리법/대기환경보전법 리서치
- 광범위한 법률을 모두 깊게 다룰수는 없기에 세분화된 특정 도메인을 선택할 필요 확인

한국 자동차 리콜 사례 분석

- 국토부에서 제공한 차량 리콜 사례를 분석하여 주요 원인 부품 확인
- 법률 해석 실패 혹은 개발 단계와 직접적으로 관련 있는 사례에 대해 세부 분석 필요 확인

다음 목표

- 어떤 도메인의 법률을 학습시킬 것인가?
- 어느 국가의 법률을 학습시킬 것인가? 어떤 언어를 학습시킬 것인가?
- LLM의 주요 기능으로 유권해석이 가능한 모델을 만들 것인가?

도메인 선택

부품 선정

-한국 자동차 리콜 사례 세부 분석

리콜 사례 중 법적 관련 없는 사례 필터링

- 리콜 사유에 ‘불량’ 이 들어가지 않고 시정 대수가 100대 이상인 건수 : 3449개
 - 리콜 사유 중복 제외 : 1044개

1차 필터링된 사례 라벨링 작업

• 문제 유형

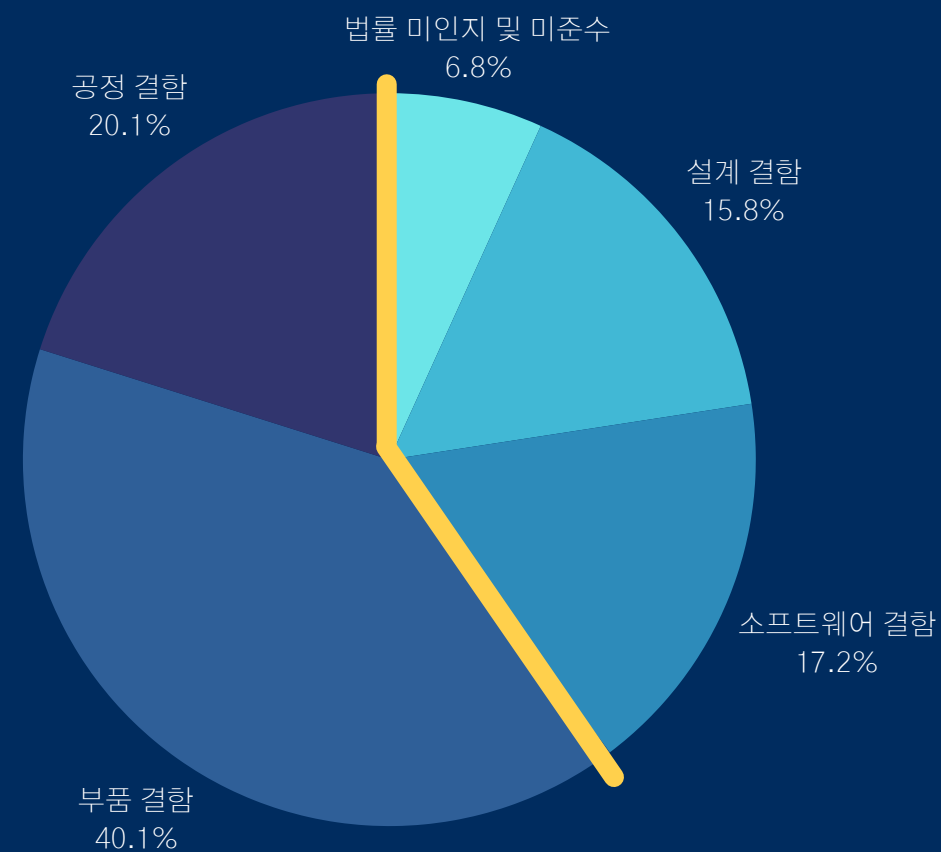
1. 규제 법률 사항 미인지 및 부적합
2. 설계 결함: 설계 상의 구조적 문제 및 필요 기능 미구현
3. 소프트웨어 결함: 탑승자와 상호작용하고 차량을 제어하는 소프트웨어 문제
4. 부품 결함: 조달 부품의 요구 성능 미충족 (내구성, 강도)
5. 공정 결함: 생산 및 조립 과정에서 발생하는 불량 (볼트 체결, 용접, 절단 공차, 이송)

• 부품 유형

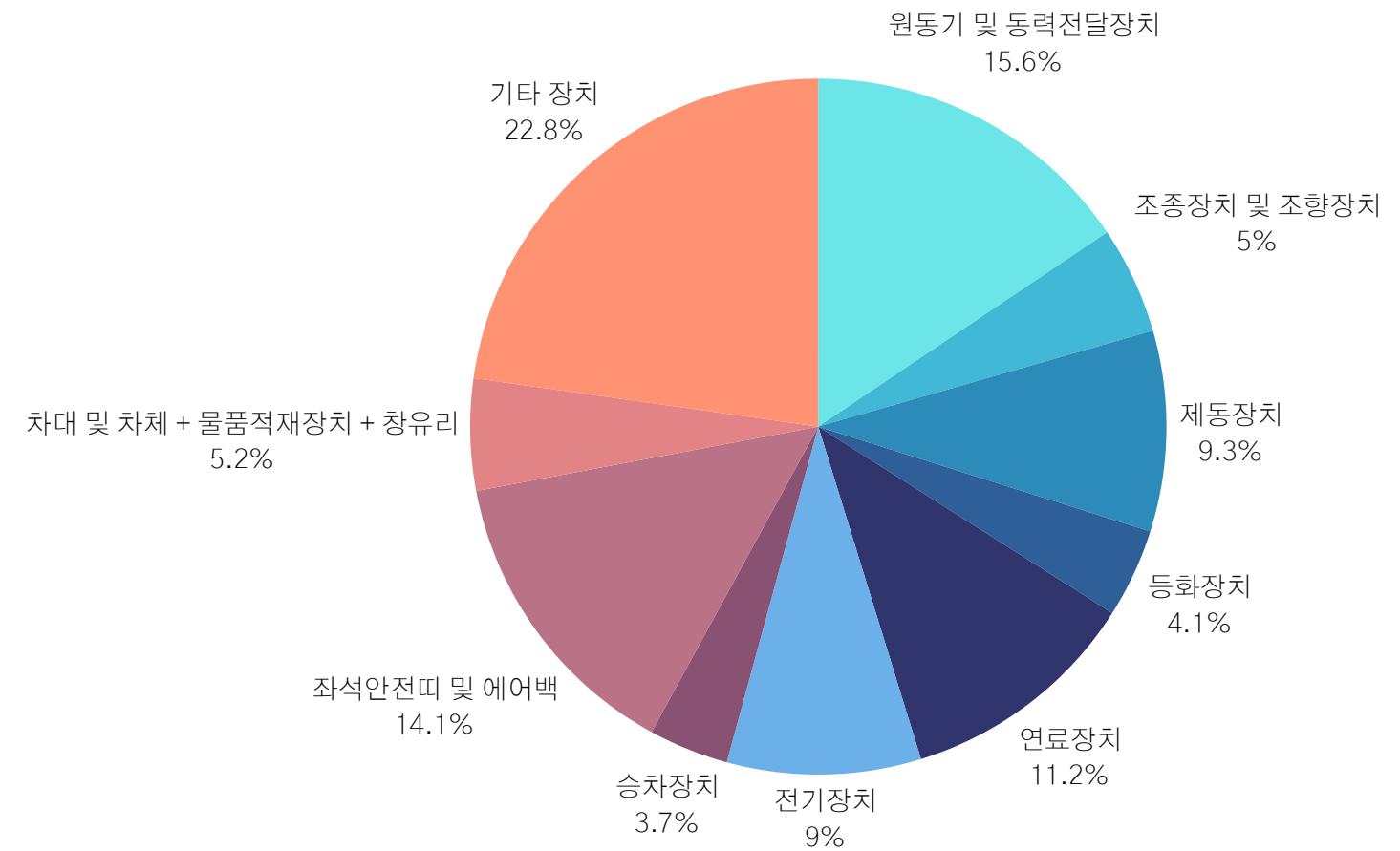
- 특정 불가 / 원동기 및 동력전달장치 / 주행장치 / 조종장치 및 조향장치 / 제동장치 / 완충장치 (or 현가장치 or 현수장치) / 등화장치 / 연료장치 / 배기가스 발산장치 / 소음방지장치 / 전기장치 / 승차장치 / 좌석안전띠 및 에어백 / 차대 및 차체 + 물품적재장치 + 창유리 / 후부반사기 / 간접시계장치 / 시야 확보장치 / 계기판 / 소화기 및 방화장치 / 견인장치 및 연결장치 / 기타 장치

부품 선정

- 1차 필터링&라벨링된 사례: 1044
 - 문제 유형이 다음과 같은 경우: 423
 - 규제 법률 사항 미인지 및 부적합
 - 설계 결함
 - 소프트웨어 결함



- 앞과 같이 필터링된 데이터에서의 원인 부품 유형



*기타 장치: 특정 불가 / 주행장치 / 완충장치 (or 현가장치 or 현수장치) / 배기가스 발산장치 / 소음방지장치 / 후부반사가 / 간접시계장치 / 시야 확보장치 / 계기판 / 소화기 및 방화장치 / 견인장치 및 연결장치 / 그 외 장치

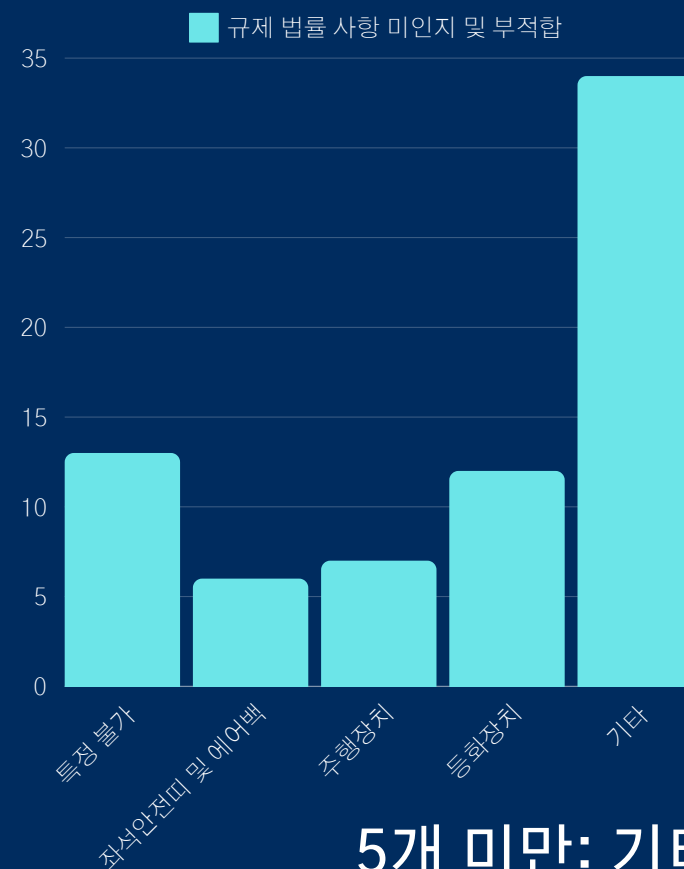
30개 미만: 기타 장치
기타 제외 원동기 및 동력전달장치가 가장 수가 많음.
상위 3개 장치들을 합하면 약 40%를 차지함.

부품 선정

• 앞과 같이 필터링된 데이터에서의 각 문제 유형별 원인 부품 유형

◦ 규제 법률 사항 미인지 및 부적합

*기타: 원동기 및 동력전달장치 / 조종장치 및 조향장치 / 제동장치 / 완충장치 (or 현가장치 or 현수장치) / 연료장치 / 배기가스 발산장치 / 소음방지장치 / 전기장치 / 승차장치 / 차대 및 차체 + 물품적재장치 + 창유리 / 후부반사기 / 간접시계장치 / 시야 확보장치 / 계기판 / 소화기 및 방화장치 / 견인장치 및 연결장치 / 기타 장치

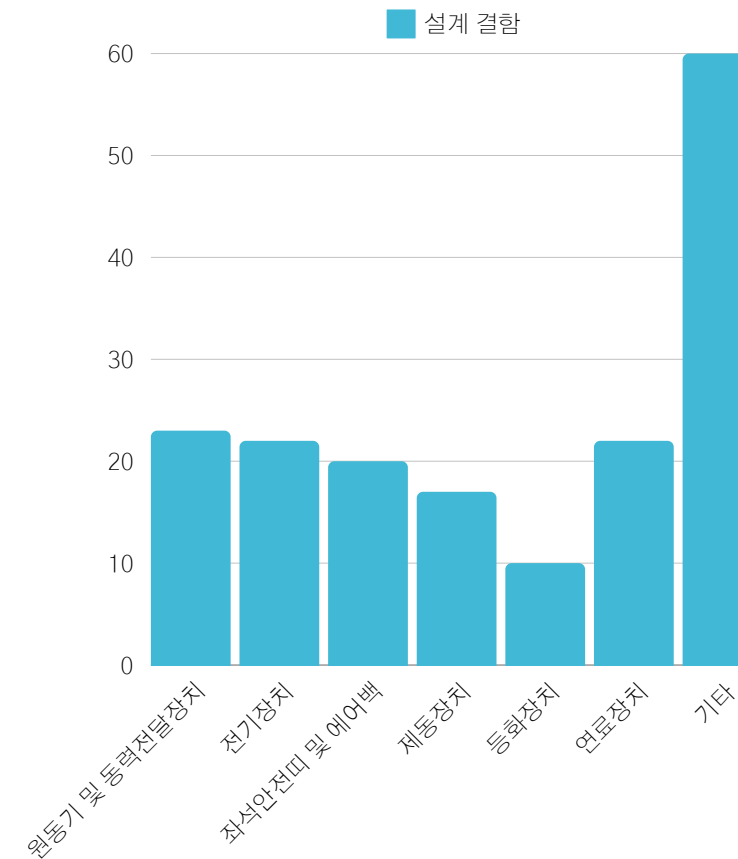


5개 미만: 기타 장치
기타 제외 특정불가(내용 부족 or 복합 원인-연비, 중량 등-, 자기인증 미부착)가 가장 수가 많았고 13개였음.

출처: 자동차리콜센터

• 앞과 같이 필터링된 데이터에서의 각 문제 유형별 원인 부품 유형

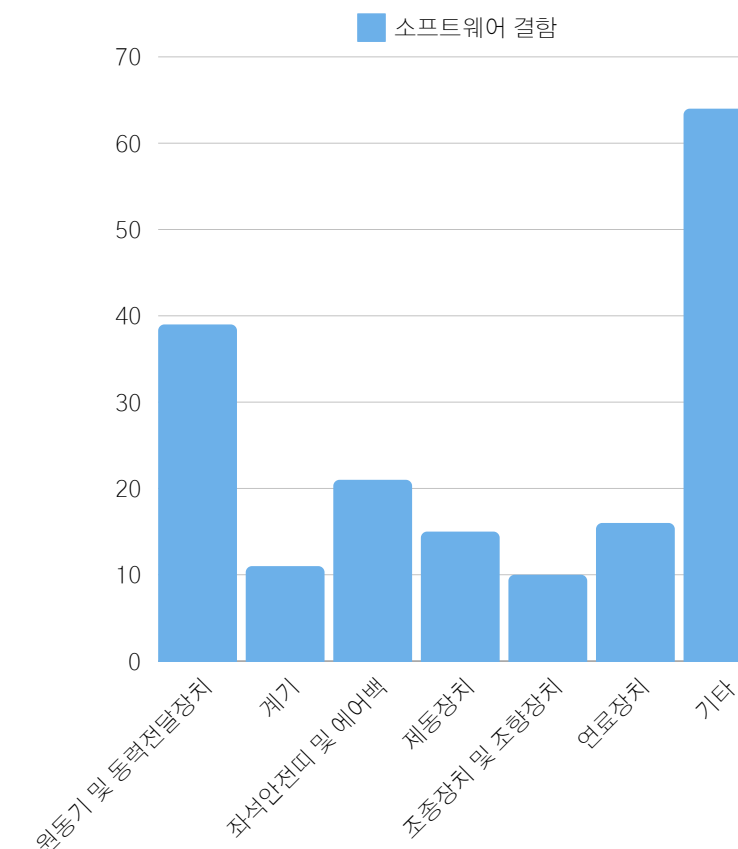
◦ 설계 결함



기타: 특정 불가 / 주행장치 / 조종장치 및 조향장치 / 완충장치 (or 현가장치 or 현수장치) / 배기가스 발산장치 / 소음방지장치 / 승차장치 / 차대 및 차체 + 물품적재장치 + 창유리 / 후부반사기 / 간접시계장치 / 시야 확보장치 / 계기판 / 소화기 및 방화장치 / 견인장치 및 연결장치 / 기타 장치

10개 미만: 기타 장치
기타 제외 20개 넘는 게 원동기 및 동력전달장치, 전기장치, 좌석안전띠 및 에어백, 연료장치가 있었음.

◦ 소프트웨어 결함



기타: 특정 불가 / 주행장치 / 완충장치 (or 현가장치 or 현수장치) / 등화장치 / 배기가스 발산장치 / 소음방지장치 / 전기장치 / 승차장치 / 차대 및 차체 + 물품적재장치 + 창유리 / 후부반사기 / 간접시계장치 / 시야 확보장치 / 소화기 및 방화장치 / 견인장치 및 연결장치 / 기타 장치

10개 미만: 기타 장치
기타 제외 원동기 및 동력전달장치가 가장 수가 많았음.
20개 이상은 제일 많은 두 개뿐.

안전으로 돌아가기

자동차 미래 기술 관련 부품

• 배경

: 앞으로의 자동차 법률도 현재 개발 되고 있는 기술에 한해서 규제가 추가될 것이고, 여러 자동차 부품들 또한 미래 자동차 기술에 따라 변화 될 것이기 때문에 부품을 선정할 때도 앞으로의 미래 자동차 기술을 고려해야한다고 생각

• 미래 자동차 기술 및 관련 부품

자율주행자동차



조종장치 : 스티어링 휠 제어, 페달 컨트롤
원동기 및 동력 전달장치 : 변속기

전기 자동차



원동기 및 동력전달장치: 감속기, 구동모터
전기장치: 인버터, 온보드 차저

사이버 보안



전 영역에 걸침

법적 해석이 ‘모호한’ 부품 분야가 있을까?

• 조사 배경

: 자동차 제작과 관련한 법적 규정이 명확히 정의되지 않아 중의적인 해석이 가능한 경우, 이는 제작 과정에 혼선을 야기하며 추후 법적 리스크(리콜, 과징금 등)으로 이어질 가능성이 있음.

: LLM이 이러한 규정 해석의 ‘모호함’을 개발 단계에서 먼저 인지하고 안전한 해석을 위한 가이드 라인을 제공할 수 있다면 제작자는 법적 리스크 해소에 상당한 도움을 받을 수 있음.

• 방법

: 해석에 있어 ‘모호한’ 부품 영역이 존재한다면 관련한 법적 질의 혹은 분쟁이 이미 많이 있었을 것을 가정

: 다음과 같은 자료를 이용하여 관련한 사례들을 조사하였음

◦ 자동차 부품과 관련한 법률(자동차 관리법, 자동차 부품법 등)의 재개정안

자동차 부품과 관련한 법률의 재개정안

• 선택 배경

: 자동차 제작과 관련하여 규정 해석이 ‘모호한’ 영역이 있다면 관련한 문제 제기나 피드백이 있었을 것이고, 이러한 해석의 ‘모호함’을 해소하기 위해 구체화된 개선안이 개정안에 반영될 것임.

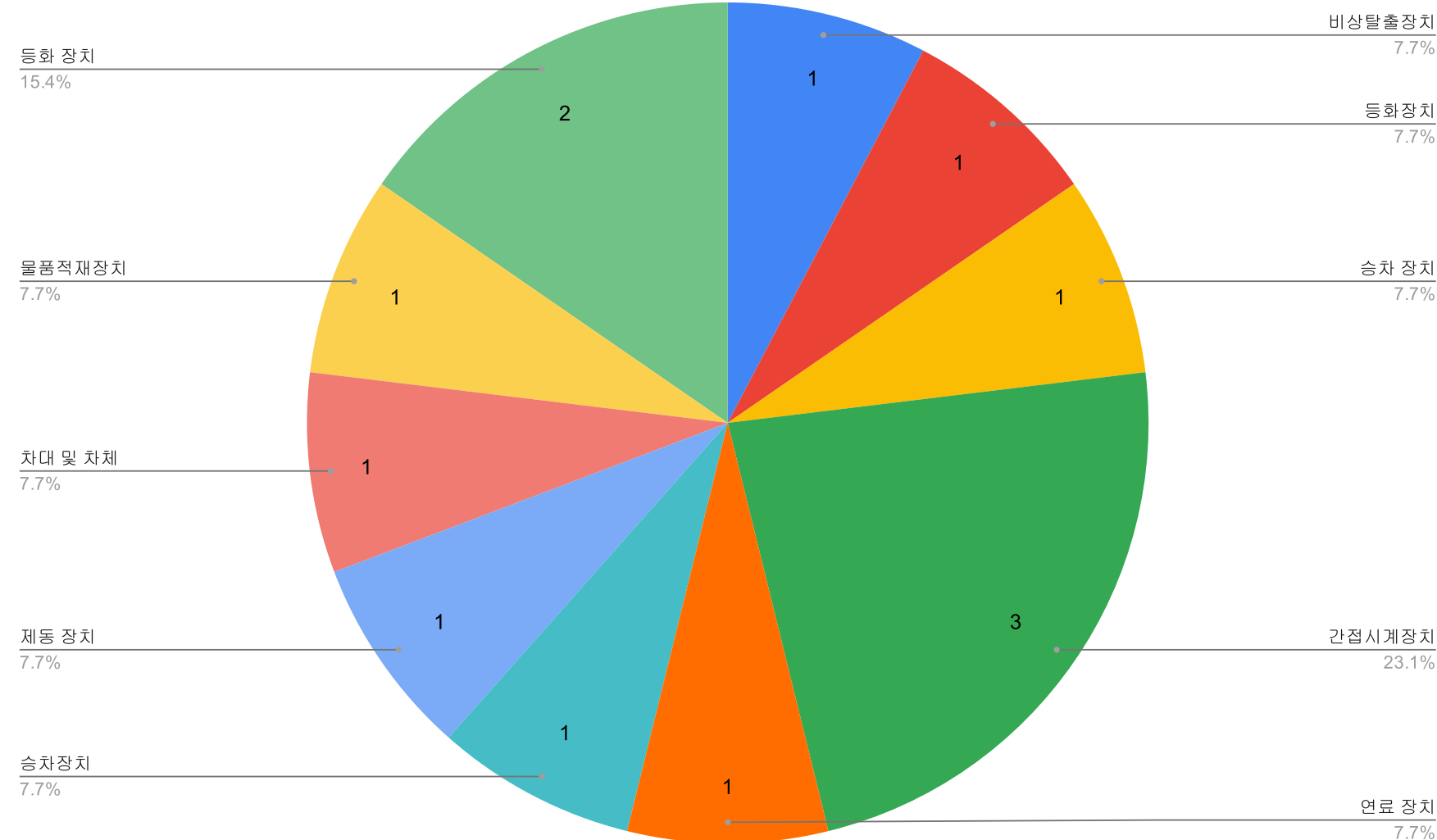
• 자료 출처 & 검색 방법

: 법제처, ‘국가 법령 정보 시스템’ 이용

: 자동차 부품 관련 법률 (자동차 관리법, 자동차 부품 등)에 대하여 키워드 검색 진행

: 근래 5개년 (2020 ~ 2024) 동안의 6개의 재개정안 – 74개의 항목에 대한 개정 확인

자동차 부품과 관련한 법률의 재개정안



• 기준의 명확화

: 개정 항목 중 개정 이유로 “명확화” 혹은 “세부화”를 명시한 경우는 74건 중 13건이 존재.

: 비중에 있어 특정 항목이 지배적이지 않음

: 예시)

해당 규정의 시험방법 등을 명확히 하기 위하여 그림의 변화 및 누락된 시험 방법을 추가 (승차 장치)

화물차 재질 및 물품적재장치 파쇄 구조 설정 및 평가 시험 규정 명확화(물품적재장치)

자동차 부품과 관련한 법률의 재개정안

• 최신 기술 분야에 대한 법적 체계 정비

: 개정 항목 중 신기술에 대한 법적 기준 신설을 위한 경우는 74건 중 10건이 존재.

: 예시)

- 구동축전지(배터리) 안전기준 시험방법의 일부 항목 추가
- 수소가스 및 천연가스 연료를 사용하는 차량충돌강 45도 초과 승합자동차의 전복시험에 대한 세부시험절차

• 국제 기준과 조화

: 개정 항목 중 국제 기준과 조화를 위한 경우는 74건 중 12건이 존재.

: 예시)

- 부분전측충돌 등 및 측면충돌에 사용되는 변형벽의 구조성 세부시험방법 등을 국제기준과 조화
- 자동차 휠 표기기준 및 시험방법 등에 대한 국제기준과 조화

타겟 도메인 선정

- 리콜 사례 세부 분석

전체 부분

가장 많이 리콜한 장치는 원동기 및 동력전달 장치, 연료 장치, 제동 장치, 전기 장치, 조종 장치 순으로 나타남

세부 부분

법률 미준수에서는 주행장치 및 등화장치, 설계 결함에서는 원동기, 전기장치, 연료장치, 소프트웨어 결함에서는 원동기, 연료, 제동, 조종 순으로 많이 나타남

- 자동차 미래기술

자동차 미래기술인 자율주행, 전기 자동차 분야와 관련한 주요 자동차 부품들은 전기장치, 원동기 및 동력 전달장치, 조종 장치와 많이 연관이 되어 있었음

- 해석의 “모호함”

자동차 부품 관련 법률적 모호함이 많이 나타날 수 있는 영역을 자동차 부품과 관련한 재개정 사례로 살펴 보았음.

법률적 해석의 어려움을 겪는 특정 부품 분야가 별도로 존재하지는 않았지만 신기술과 국제 조화를 위한 재개정 사례들을 확인

미래기술과 관련된 장치들 중 가장 많이 리콜이 되고 있는 사례인
원동기 및 동력전달 장치와 전기장치 그리고 제동 및 조종 장치 순으로 살펴볼 예정

국외 법체계와 한국의 국제 조화 체계

UNECE WP.29 : 유럽

• 기관 소개

차량에 대한 기술 규정의 글로벌 차원의 조화(조정, 일치) 또는 개발을 목표로 활동을 추구하고 차량 안전 개선, 환경 보호, 에너지 효율 증진 및 도난 방지 성능 향상을 목적으로 하는 규제를 개발 목적

• 개최 주기

정기 모임 : 매년 정기적으로는 총 3번 (3월, 6월, 11월) 모여 자동차 안전 기준 및 규제 관련 결의안 채택

비정기 모임 : 필요에 따라 개최 (최근 매년 또는 2년마다 비정기 모임 가짐)

• 영향력

UNECE WP.29에서 발의한 규약은 EU회원국인 경우 법적인 구속력을 가지므로 자국법에 해당 규약을 적용 시켜야 함

• 세부적 안전기준

세부적인 안전기준은 ISO 국제 표준 기구에서 선정한 자동차 안전 국제 표준을 사용함

예) 자율발렛주차(조종장치 관련) : UN.R79 - ISO/AWI 23374-1, 전기차 배터리 : UNECE WP.29 GTR 13-ISO 12405-1

UNECE WP.29 : 비유럽

• UNECE WP.29 관련 협정 체결문

	1958 협정	1998 협정
목적	안전기준의 국제조화와 형식승인의 상호 인정	통일된 안전기준의 배정 (형식승인 개념 배제)
체약국	59개국 (유럽연합, 한국, 일본)	38개국 (한국, 미국, 중국, 유럽, 일본 등)
기준적용	체약국 선택 적용	체약국 의무 적용 (미채택 시 사유 매년 제출)
회의 진행	연간 총회 3회, 6개 전문분과회의 각 2회 개최	6개 전문분과 회의 시 분야별 동시 진행
차이	차량 설계 및 제조 관련 안전 기준에 초점을 맞춤	1958협정보다 훨씬 더 광범위한 규제를 포함 (차량 배기가스 배출, 연료 소비, 소음 등 환경 관련 규제)
시사점	한국, 일본은 1958협정 및 1998협정 체결국가로서 해당 기관에서 공표한 국제 표준을 준수하고 따라 가고 있음, 반면 중국, 미국은 1998협정에만 체결하고 있어 세부적인 부분은 다를 수 있지만 전반적인 추세와 흐름은 UNECE WP.29를 따름	

NHTSA: 미국

- 독자적인 기관 존재 - NHTSA (미국 고속도로 교통안전청)

미국 교통부 산하 소속으로 차량 성능 표준과 주 또는 지방 정부와의 파트너십을 시행함.
자동차 관련 법률에 대해서 유권해석하는 단체이기도 함.

- UNECE와 ISO와의 관계

- UNECE 회원국 아님 - 하지만 그 조화를 위한 노력은 하고 있음
 - 1998 협정은 함께 체결하여 UN GTR 부분은 따르고 있음
- 인증 표준 센터 공지에 따르면 2024 국제표준화기구 자동차 기능안전*(ISO/TC 22/SC 32/WG 8) 표준 회의에 참여
- 법률 기사로 ISO, SAE International 등과 유사한 신뢰 가능한 기관의 안전기준 도입을 확인할 수 있었음.

- 자율주행 자동차 관련 법률 가장 먼저 시행한 나라

2009년 자율주행 자동차 시범 운행이 시작 후, 2011년 8월 수동모드로 전환되어 사람이 운전 하던 구글 자율주행자동차 사고가 발생 -> 법적 규율의 필요성 대두 -> 입법 활동 활발

자동차 안전기준 국제조화를 위한 노력

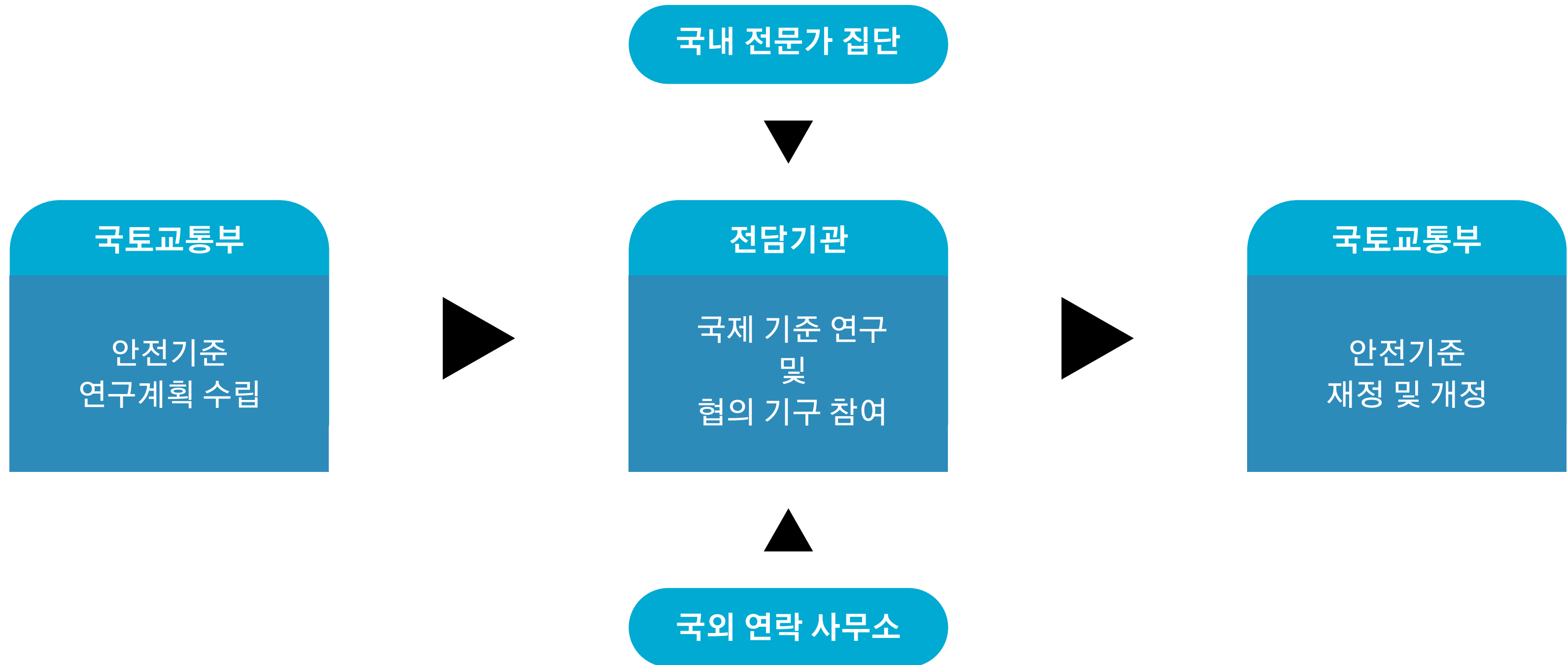
• 개요

- : 자동차 무역에서 교역국 간의 상이한 안전기준은 무역장벽으로 작동하며 개발 및 생산에 있어 비효율을 초래함.
- : 이를 해소하고 자국의 자동차산업 경쟁력 확보를 위해 안전기준의 국제조화가 세계적인 추세임.
- : 한국은 앞서 언급했던 1958 협정과 1988 협정의 체결국이며 다양한 논의 기구에 참여하고 있음.

• 주요 협의체

- UN/ECE/WP29
- APEC TPT-WG (아시아 및 평양 지역 회원국: 한국, 일본, 중국, 미국 등)
- Public and Private Joint Forum in Asian Region (일본, 중국, 인도 및 ASEAN 국가들)

자동차 안전기준 국제조화를 위한 노력



타겟 법체계 선정

- UN ECE WP.29은 유럽의 자동차 규제 규정 당국으로서 많은 전문가들의 협의를 통해 선도적으로 자동차 안전 규제를 만들고 ISO(국제표준규격)을 반영함
- 미국의 경우, UNECE의 회원국은 아니나 1998 협정(UN GTR)을 따르며, ISO(국제표준규격)를 반영함으로써 직간접적으로 UN ECE의 내용을 자국법에 반영하고 있음.
- 한국은 1958 협정과 1988 협정을 모두 체결하였으며 다양한 국제 협의체에 참여하여 자국법을 국제 기준과 조화 시키기 위해 많은 노력을 가하고 있음.

사실상 국제 표준으로 작동하고 있는
UN ECE WP.29의 UN Regulations 과 UN GTR을 타겟 법체계로 선정

구현 기능 선택을 위한 LLM 평가 방안

구현 목표 기능의 종류와 수준 선정을 위한 모델 성능 평가

- 구현하고자 하는 서비스의 핵심 기능들은 LLM에 의해 수행됨.
따라서 LLM의 성능은 목표 기능의 종류와 수준 결정에 있어 중요한 고려 사항임.
- 평가의 객관성과 신뢰성을 위해 체계적으로 구성된 벤치 마크를 사용
 - 널리 사용되는 상용 벤치 마크 (MMLU, DROP 등)
 - 법률 분야에 대한 특화 벤치 마크 (LegalBench)

후보 모델 선정

- Chat-GPT 3.5

- : 높은 접근성과 준수한 성능
- : 상대적으로 저렴한 이용료 (\$0.50 / 1M tokens)

- Chat-GPT 4o (omni)

- : GPT 4보다 소폭 앞선 성능 수준
- : 다국어 입출력에 대해서는 성능 향상
- : GPT 4의 절반 수준의 이용료 (\$5 / 1M tokens)

- Meta-Llama-3-7B

- : 최신 오픈 소스 모델
- : 작은 크기로 인한 범용성 존재

- Meta-Llama-3-70B

- : 최신 오픈 소스 모델
- : GPT 3.5 이상의 성능

적용 벤치 마크

- MMLU

- : Massive Multitask Language Understanding
 - : 다양한 주제에 대한 언어 이해력을 평가
(과학, 수학, 기술, 사회학, 인문학 등)

- BIG-Bench Hard

- : 고급 언어 기반 문제 해결
 - : 현재 LLM 기능을 넘어서는 수준을 기준으로 평가

- DROP

- : Discrete Reasoning Over Paragraphs
 - : 문단을 이해하고 주어진 정보를 기반으로
복잡한 질문에 대답

- GSM-8K

- : grade School Math 8K
 - : 초등 수준의 수학 능력

적용 벤치 마크

	GPT 3.5	GPT 4o	LLAMA 3 7B	LLAMA 3 80B
MMLU	70.0	88.7	66.6	79.5
DROP	64.1	83.4	58.4	79.7
GSM-8K	57.1	97.0	79.6	93.0

법률 분야에 특화된 성능을 측정하기 위한 벤치 마크, LegalBench

LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models



Neel Guha, Julian Nyarko, Daniel E. Ho, Christopher Re, Adam Chilton, Aditya Narayana, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore, Diego Zambrano, Dmitry Talisman, Enam Hoque, Faiz Surani, Frank Fagan, Galit Sarfaty, Gregory M. Dickinson, Haggai Porat, Jason Hegland, Jessica Wu et al. (20 additional authors not shown)

Published: 26 Sept 2023, Last Modified: 03 Nov 2023 NeurIPS 2023 Datasets and Benchmarks Poster Everyone Revisions BibTeX

Keywords: large language models, law, benchmarks, legal reasoning, foundation models

TL;DR: We work with the legal community to develop a benchmark of 160+ tasks to evaluate legal reasoning in large language models

Abstract:

The advent of large language models (LLMs) and their adoption by the legal community has given rise to the question: what types of legal reasoning can LLMs perform? To enable greater study of this question, we present LegalBench: a collaboratively constructed legal reasoning benchmark consisting of 162 tasks covering six different types of legal reasoning. LegalBench was built through an interdisciplinary process, in which we collected tasks designed and hand-crafted by legal professionals. Because these subject matter experts took a leading role in construction, tasks either measure legal reasoning capabilities that are practically useful, or measure reasoning skills that lawyers find interesting. To enable cross-disciplinary conversations about LLMs in the law, we additionally show how popular legal frameworks for describing legal reasoning—which distinguish between its many forms—correspond to LegalBench tasks, thus giving lawyers and LLM developers a common vocabulary. This paper describes LegalBench, presents an empirical evaluation of 20 open-source and commercial LLMs, and illustrates the types of research explorations LegalBench enables.

Supplementary Material: [pdf](#)

Submission Number: 413

법률 분야에 특화된 성능을 측정하기 위한 벤치 마크, LegalBench

- Issue-spotting

: 주어진 정보의 특정 법적 문제 혹은 특정 법률 영역에 대한 관련 여부 판단 능력을 평가

- Rule-recall

: 주어진 법률 문제에 대해서 적용해야 할 올바른 법률 규칙을 식별하는 능력을 평가

- Rule-application

: 특정 역할(ex. 판사)을 설정하고 주어진 상황에 대해서 적절한 역할 수행을 시도하는지 평가

- 위 지표를 포함한 총 6가지 지표에 대해서 정량적, 정성적 평가를 진행

감사합니다!

[안전으로 돌아가기](#)