

# 전문가 + 초거대 AI 조합형 반자동화 학습데이터 구축



## 메뉴얼 학습데이터 정형화

원천 데이터 (고장조치 메뉴얼)		
오류	발생 오류 내용	오류시 도어 상태
사이드 도어 DCU failure (DCU 오류)	DCU 하드웨어의 중고 장(CPU, 메모리, 디스크, 전원부 등) : DCU 전원 주입시 및 주기적인 차가 테스트 과정에서 발생	도어는 그 위치에서 프리상태(수동작동 가능상태)로 유지하며, 전체 기능이 정지된다. 기능적인 영향 없이 도어는 작동된다.
사이드 도어 DCU failure (DCU 오류)	DCU 하드웨어의 경고 장(FROM 등) : DCU 전원 주입시 및 주기적인 차가 테스트 과정에서 발생	도어는 그 위치에서 프리상태(수동작동 가능상태)로 유지하며, 전체 기능이 정지된다. 기능적인 영향 없이 도어는 작동된다.
사이드 도어 motor circuit failure	사이드 도어의 경고 부터 모터온류가 10% 간 증가하지 않음.	도어는 오류가 현시되는 한 개폐할 수 없다.

**정형화 되지 않은 예시**

DCU failure (DCU 오류) DCU 하드웨어의 중고장(CPU, 메모리, 디스크, 전원부 등) : DCU 전원 주입시 및 주기적인 차가 테스트 과정에서 발생 도어는 그 위치에서 프리상태(수동작동 가능상태)로 유지하며, 전체 기능이 정지된다. 기능적인 영향 없이 도어는 작동된다.

**질의응답 예시**

사이드 도어의 DCU 오류시 도어 상태를 알려주세요.

**정형화된 예시**

DCU 오류 내용 : DCU 하드웨어의 중고장(CPU, 메모리, 디스크, 전원부 등) : DCU 전원 주입시 및 주기적인 차가 테스트 과정에서 발생

기능적인 영향 없이 도어는 작동된다.

DCU 오류 내용 : DCU 하드웨어의 경고장(FROM 등) : DCU 전원 주입시 및 주기적인 차가 테스트 과정에서 발생

기능적인 영향 없이 도어는 작동된다.

DCU 오류 도어 상태 : 도어는 그 위치에서 프리상태(수동작동 가능상태)로 유지하며, 전체 기능이 정지된다. 기능적인 영향 없이 도어는 작동된다.

- 표/도식 등의 형태를 관계성 파악을 위해 정형화된 학습데이터를 학습함으로써 SLLM 답변 품질 강화

## 직접 수집 이미지 전처리 및 학습 데이터 라벨링

### 이미지 전 처리

제품질 이미지 제거

이미지 크기 조정

이미지 방향 조정(회전)

이미지 선명도 조정



이미지 라벨링툴 MakeSence

PyTorch  
딥러닝 프레임 워크

- 이미지 라벨링 전문 툴 활용 부품 이미지 라벨링 수행
- 인식도 향상을 위해 부품 당 라벨링 이미지 3,000장 학습

# 데이터 특성을 고려한 라벨링으로 학습데이터 활용성 향상

## 데이터 특성을 고려한 학습데이터 구축

- 문장, 나열형 정보, 표, 그림 등을 포함한 라벨링

\* 고장조치 매뉴얼의 중요 정보를 누락 없이 포함

### Type 01 서술형 문장

단문 및 장문의 서술 방식으로 작성된 형태



### Type 02 나열형 정보

다수의 항목들을 나열형으로 제시하는 형태



### Type 03 표 형식 정보

다단의 표 형식으로 정보가 제시되는 형태



### Type 04 그림정보

도식, 구조 등의 그림으로 정보가 표시되는 형태



유형	데이터 예시	라벨링 방안																
서술형 문장	<p>노어를 차체에서 분리하기 위해서는 차체와 연결되어 있는 모든 전기장치들부터 분리시켜야 한다.</p> <p>먼저 도어 엔진의 터미널 블록과 연결되는 차량의 전선을 분리하고,</p> <p>각종 조작을 위해 스위치 브레이켓에 부착되어 있는 스위치들을 분리한다.</p> <p>분리시 발생할 수 있는 물리적 충격으로부터 DCU와 모터를 보호하기 위하여 미리 분리해 놓아야 한다.</p>	<p>(변환) 완전한 의미가 포함되도록 문장 단위로 내용 가공</p> <ul style="list-style-type: none"> <li>문장의 주어에 해당하는 [고장진단 대상 부품]을 포함시키기 문장 완성</li> </ul> <p>(생성) 문장 내 질문 추출 및 답변 작성</p> <ul style="list-style-type: none"> <li>즉 출입문 교체에서 도어를 차체에서 분리하기 위해서는 차체와 연결되어 있는 모든 전기장치들부터 분리시켜야 합니까?</li> <li>즉 출입문 교체를 위한 분리 시 발생할 수 있는 물리적 충격으로부터 DCU와 모터를 보호하기 위해서 필요한 조치는 어떻게 되나요?</li> </ul>																
나열형 정보	<p>(기) 판넬 좌, 우의 밀착성을 조정한다. (②번 판넬 간 겹 조정부)</p> <p>ㄱ. 로크 너트의 풀림 상태에서 잠금이 되도록 판넬을 이동시켜, 잠금(후크)이 되도록 위치 시킨다.</p> <p>ㄴ. 후크의 잠금이 확인되면, 좌, 우 판넬을 밀착시키고, 로크 너트를 후크 스패너를 이용하여, 체결한다.</p> <p>ㄷ. 이때 스크류의 캡과 레버 안쪽의 블록(반달모양)이 0~5mm를 유지 하여야 한다.</p>	<p>(변환) 나열형 정보를 포함하여 문장 단위로 내용 가공</p> <ul style="list-style-type: none"> <li>판넬 좌, 우의 밀착성을 조정하기 위해서 로크 너트의 풀림 상태에서 잠금이 되도록 판넬을 이동시켜, 잠금이 되도록 위치시킨다.</li> </ul> <p>(생성) 문장 내 질문 추출 및 답변 작성</p> <ul style="list-style-type: none"> <li>판넬 좌, 우의 밀착성을 조정하기 위해서 필요한 조치는 무엇인가요?</li> <li>스크류의 캡과 레버 안쪽의 블록의 유지 간격은 무엇인가요?</li> </ul>																
표 형식 정보	<table border="1"> <thead> <tr> <th>순번</th> <th>코드 번호</th> <th>코드 설명</th> <th>고장조치</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>888</td> <td>경상</td> <td>경상 통작시 표시 된다.</td> </tr> <tr> <td>1</td> <td>F01</td> <td>DCU Major Failure</td> <td>DCU를 교체 한다.</td> </tr> <tr> <td>2</td> <td>F02</td> <td>DCU Minor Failure (EEPROM, SRAM)</td> <td>파라메타 실패시 기본값으로 임의 설정되며, 모니터링 프로그램을 이용하여 파라메타를 재설정 한다.</td> </tr> </tbody> </table>	순번	코드 번호	코드 설명	고장조치	0	888	경상	경상 통작시 표시 된다.	1	F01	DCU Major Failure	DCU를 교체 한다.	2	F02	DCU Minor Failure (EEPROM, SRAM)	파라메타 실패시 기본값으로 임의 설정되며, 모니터링 프로그램을 이용하여 파라메타를 재설정 한다.	<p>(변환) 표를 문장으로 변환하여 문장 단위로 내용 가공</p> <ul style="list-style-type: none"> <li>F01, DCU Major Failure, DCU를 교체 한다.</li> <li>F02, DCU Minor Failure(EEPROM, SRAM), 파라메타 실패시 기본값으로…(생략)</li> </ul> <p>(생성) 문장 내 질문 추출 및 답변 작성</p> <ul style="list-style-type: none"> <li>코드번호 8888은 정상 통작시 표시된다.</li> <li>F01은 DCU Major Failure로 DCU를 교체해야 한다.</li> <li>F02는 DCU Minor Failure로 파라메타 실패 시, 기본값으로 임의 설정되며, 모니터링 프로그램을 이용하여 파라메타를 재설정한다.</li> </ul>
순번	코드 번호	코드 설명	고장조치															
0	888	경상	경상 통작시 표시 된다.															
1	F01	DCU Major Failure	DCU를 교체 한다.															
2	F02	DCU Minor Failure (EEPROM, SRAM)	파라메타 실패시 기본값으로 임의 설정되며, 모니터링 프로그램을 이용하여 파라메타를 재설정 한다.															
그림 정보		<p>(변환) 문장 내 그림의 위치를 저장하여 메타데이터 관리</p> <p>(생성) 문장 내 질문 추출 및 답변 작성</p> <ul style="list-style-type: none"> <li>도어 구동부 설치에서 구동부와 레일 간의 평행이 되도록 하는 간격은?</li> <li>구동부가 레일의 중앙에 들어가도록 위치 시킬 때 수치는?</li> </ul>																

# 반복적 프롬프트 개선 프로세스 도입으로 지속적 품질 강화

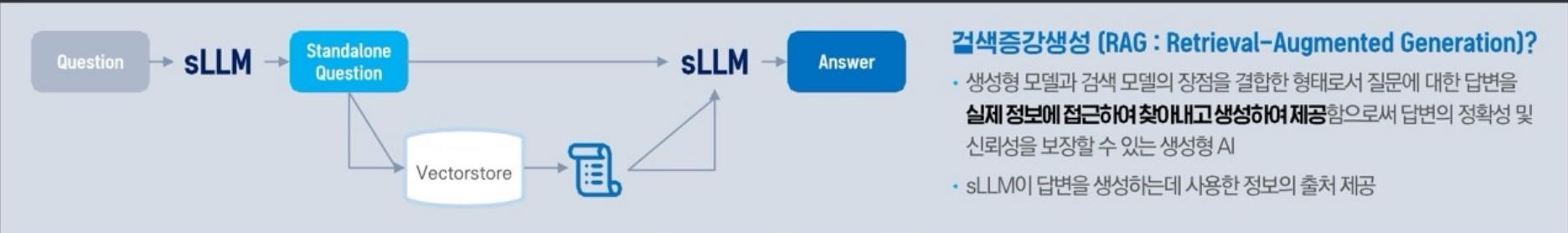
## 프롬프트 구축 개요

- 캐릭터 설정
  - AI에게 사용자의 배경, 지역, 담당업무 등의 정보 부여
- 제약사항 주입
  - 언어모델이 수행할 작업 설명, 요구사항 포함
- 답변 형식 제공
  - 원하는 형태의 결과물 형식을 지정



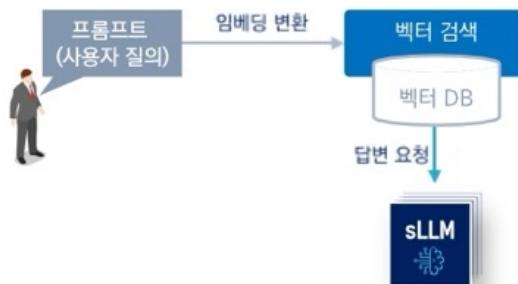
# 검색 증강 생성(RAG) 적용을 통한 생성 AI 신뢰성 향상

## RAG(Retrieval Augmented Generation) 개념도



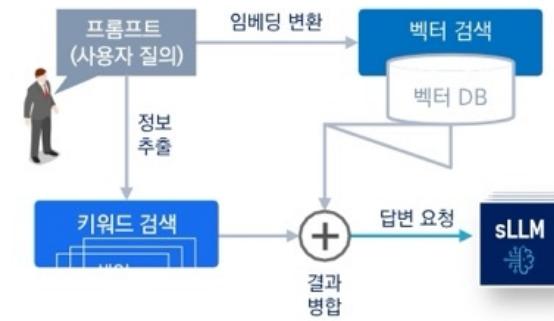
### TYPE 1 벡터 기반 Full 검색

- 사용자 질의를 임베딩 변환, 벡터 DB에서 유사도가 가장 높은 정보를 선택



### TYPE 2 키워드 및 벡터 검색 병합

- 키워드 포함 정보 중 유사도가 높은 정보 선택 or 키워드 검색 후 벡터 검색 기반 결과 랭킹



# QA 서비스용 학습 시나리오 상세 방안 수립

## 고장수리 절차서 차체 및 설비 메뉴얼

Q.

사이드 도어의 DCU 오류 시 도어 상태를 알려주세요.

A.

사이드 도어의 DCU 오류 시 도어 상태는 아래와 같습니다.

사이드 도어의 DCU 오류시 도어는 그 위치에서  
프리상태(수동작동가능상태)를 유지하며, 전체  
기능이 정지됩니다.

기능적인 영향 없이 도어는 작동됩니다.

관련문서      고장수리 절차서 차체 및 설비  
                    측 출입문 M00-0-U1234-T-678

질문을 입력하세요



## 1,941 건 원천데이터 중 라벨링 작업 데이터

필드명	예시	샘플												
문서 식별자	1234													
문서 번호	M00-0-U1234-T-678													
문서유형	고장수리절차서													
문서주제	측 출입문													
작성일자	2016-12-22													
문서명	고장수리 절차서 측 출입문(Side Door System)													
발행기관	한국철도공사													
작성자	홍길동													
목적	<p>본 문서는 일반형 전기동차에 장착된 측 출입문의 고장수리작업을 기술하고 있으며, 유지보수 요원용으로 작성되었다.</p> <p>이 문서를 이해한 유지보수 요원은 관련제품에 대한 고장수리를 수행할 수 있는 능력을 갖출 수 있게 될 것이다.</p>													
사용공구	독각렌치, 스페너, 후크 스파너, (-)드라이버, (+)드라이버, 토크レン치													
고장유형 (고장코드)	<table border="1"> <tr> <td>888</td> <td>상</td> <td>정상 출입문 동작</td> </tr> <tr> <td>F01</td> <td>상</td> <td>도어 제어부 결함</td> </tr> <tr> <td>F02</td> <td>상</td> <td>도어 제어부 기동으로 결함</td> </tr> </table>	888	상	정상 출입문 동작	F01	상	도어 제어부 결함	F02	상	도어 제어부 기동으로 결함				
888	상	정상 출입문 동작												
F01	상	도어 제어부 결함												
F02	상	도어 제어부 기동으로 결함												
고장수리 절차 및 방법	<p>1) DCU Failure(DCU 오류)</p> <table border="1"> <tr> <td>상태</td> <td>DCU Failure(DCU 오류)</td> </tr> <tr> <td>조치</td> <td>제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함), 제어부 결함으로 출입문이 정지되는 경우(도어 제어부 기동으로 결함)</td> </tr> <tr> <td>조치</td> <td>제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)</td> </tr> <tr> <td>조치</td> <td>제어부 결함으로 출입문이 정지되는 경우(도어 제어부 기동으로 결함)</td> </tr> <tr> <td>조치</td> <td>제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)</td> </tr> <tr> <td>조치</td> <td>제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)</td> </tr> </table>	상태	DCU Failure(DCU 오류)	조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함), 제어부 결함으로 출입문이 정지되는 경우(도어 제어부 기동으로 결함)	조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)	조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 기동으로 결함)	조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)	조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)	
상태	DCU Failure(DCU 오류)													
조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함), 제어부 결함으로 출입문이 정지되는 경우(도어 제어부 기동으로 결함)													
조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)													
조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 기동으로 결함)													
조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)													
조치	제어부 결함으로 출입문이 정지되는 경우(도어 제어부 결함)													

참조  
정보

## 학습데이터 중 Instruction tuning QA 데이터

INPUT

질의 내용

고장코드가 F01인 경우 메모리 로딩 및 초기화 실행을 수리하기 위해서 필요한 절차는 무엇인가요?

- 사용자의 별도의 절차 없이 이상이 생긴 메모리 영역은 출고 시 상태로 초기화 된다.
  - 출고시와 다르게 변경된 파라미터가 있다면 재설정 한다.
  - 동일한 DCU에서 계속해서 해당 고장이 발생할 경우 메모리의 수명 문제 또는 메모리 셀 불량의 가능성이 있으므로 교체가 필요하다.
- 조치 후 시험 : 모니터링 소프트웨어에서 파라미터를 변경하고 로딩하여 올바르게 저장되었는지 확인한다

컨텍스트  
(주어진 내용)

지시문

참고 내용

답변 형식

설명 방식

주어진 [컨텍스트]를 참고하여 질문에 대한 적절한 답변을 작성해주세요.  
답변을 작성할 때는 다음의 [지시사항]을 참고하여 작성해주세요.

- ✓ 해당 지침과 규정 기준에 맞는 답변으로 작성하세요.
- ✓ 출고하기 쉽고 실제 사용에 적용 가능한 방식으로 설명하세요.
- ✓ 답변을 작성할 때, [컨텍스트] 이외에 다른 내용은 쓰지 마세요.
- ✓ 답변은 [답변 형식]에 맞춰 작성하세요.

OUTPUT  
(답변내용)

별도 절차 없이 이상이 생긴 메모리 영역은 출고 시 상태로 초기화 됩니다.  
출고시와 다르게 변경된 파라미터가 있다면 재설정 합니다. 그럼에도 동일한  
DCU에서 해당 고장이 발생하는 경우 메모리의 수명 문제 또는 메모리 셀 불량의  
가능성이 있으므로 교체를 합니다. 조치 후에는 테스트로 검증합니다.

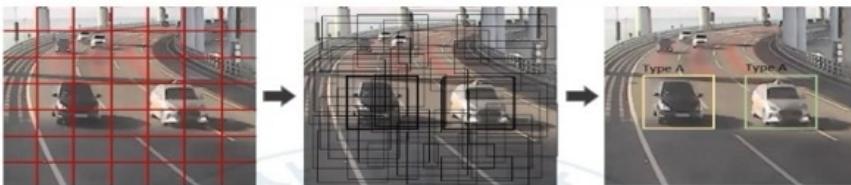
관련 문서 : 고장수리 절차서 차체 및 설비 측 출입문 M00-0-U1234-T-678

인스트럭션(Instruction) 데이터 셋 학습 통한 사용자 맞춤형 답변 생성

# 실시간 객체 탐지 최적모델 적용, 분석속도 및 정확도 확보

## YOLOv8

실시간 객체 탐지 및 이미지 분할 모델



YOLOv8 모델 활용 실시간 이미지 검출

### 최적화된 정확도-속도 트레이드오프

- 정확도와 속도 간의 최적의 균형을 유지하는데 중점을 두어 다양한 애플리케이션 영역의 실시간 물체 감지 작업에 적합

### 다양한 사전 학습 모델

- 다양한 작업 및 성능 요구 사항을 충족하는 다양한 사전 학습 모델을 제공

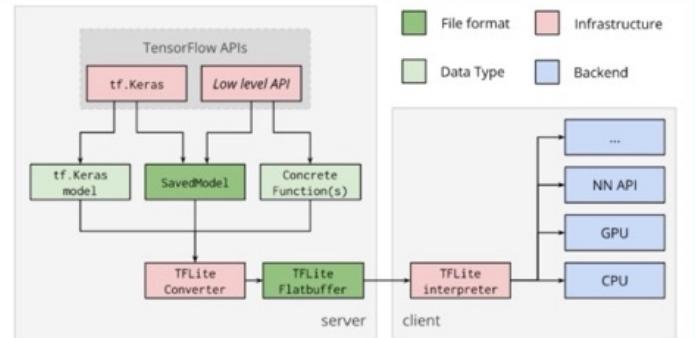
### 고급 백본 및 Neck 아키텍처

- 최첨단 백본 및 Neck 아키텍처를 사용하여 특징 추출 및 객체 감지 성능 향상

## TensorFlowLite

모바일 기기 위한 모델 최적화

모바일 기기 최적화  
TensorFlow Lite  
적용 구성도



### 지연 시간 축소

- 에지 추론 방식으로 서버와 왕복이 없어 지연 시간이 짧음

### 데이터 보호

- 에지 추론 방식으로 데이터는 네트워크에서 공유되지 않으므로 데이터 프라이버시에 대한 우려 해결

### 모델 크기 최적화

- 에지 장치의 리소스 제약을 고려한 최적화

### 인터넷 연결 불필요

- 인터넷 연결이 필요하지 않으므로 연결 문제가 없음

### 낮은 전력 소모

- 효율적인 추론과 네트워크 연결 불필요로 낮은 전력 소모

# 사람과 인공지능의 협업을 통한 공동 데이터 구축 기법 적용

## Self-Written with Human 방법론

사람이 작성한 초기 데이터(Seed Data)를 LLM(초거대언어모델)에 입력하여 학습 데이터를 증강하여 생성하고, 이를 사람이 다시 검수 및 보완하는 제안사의 학습데이터 구축 방법론

- ✓ LLM 모델을 연동하여 Instruction 데이터 셋 자동 증강 생성
- ✓ LLM 모델을 통한 추출 정보의 다양성 확보
- ✓ 사람의 재검수를 통한 품질 확보



### Why? Self-Written with Human 방법론이 필요할까?

- ✓ 대량 데이터에 대한 수동 작업 한계 극복 필요

Human Written	Self-Written	Self-Written With Human
고비용	저비용	
다양성/창의성 부족	환각/독설/편견 존재	저비용으로 다양하고 창의적인 고품질 데이터 구축
고품질 구축	저품질 구축	

## Self-Written with Human 방법론

사람이 수행한 초기 작업(Task)에 기반하여 기계(LLM)의 다양하고 창의적인 작업 결과를 활용합니다.

### Step 01 초기 데이터 작성

#### 원천데이터 입력

4호선 지하철 전동차 고장진단 매뉴얼 및  
경정비 관련 주요 부품 300개 선정

종류	분량
고장진단 매뉴얼	1,951종
경정비 부품 이미지	300종 (총 3,000장)

### 사람) 질문답변 작성-1차 수동

- ✓ 도어판넬 분해에 있어서 정소가 혐오한 경우 조치 방법은?
- ✓ 링크와 행거 플레이트가 연결된 볼트를 제거하고 편심 볼트를 끌어 분리하는 방법은?
- ✓ 수동 열림력 검사 항목과 조건은?

#### 다양성 및 창의력 부족

### Step 02 LLM 통한 증강 생성

#### LLM 입력 프롬프트 작성

아래 내용에서 질문들을 도출하여 작성해 주세요.

#### 컨텍스트[주어진 내용]

판넬과 행거 사이에 라이너를 삽입(볼트부위)하여 중앙 레버의 상, 하부 간격을 일정하게 조정한다.  
-라이너는 기본적으로 2(0.5, 1.5)이나, 치중의 조건이나 판넬의 위치 조정을 하며, 더하거나 줄여서 작업한다.

### Step 03 데이터 필터링

#### LLM 질문답변 작성-2차 자동

주어진 내용에서 도출한 질문들은 아래와 같습니다.

- ✓ 도어판 차체에서 분리하기 위해서 치우와 연결되어 있는 모든 전기 부품을 분리하는 방법은?
- ✓ 출입문 교체방법에서 전선 및 스위치 분리방법은?
- ✓ 스위칭 모듈 제작 과정에서 배이어셋을 렌지 분리방법은?
- ✓ 구성부품 고장수리 절차 및 방법에서 DCU 오류는?
- ✓ 열린 달릴 명령 후 모터에 과전류가 흐르면?

내용누락

부적합

### 지시문

주어진 [컨텍스트]를 바탕으로 [지시사항]을 참고하여 작성해주세요.

- ✓ 만들수 있는 모든 질문과 답변에 작성해주세요.
- ✓ 답변을 작성할 때 [컨텍스트]와 함께 다른 내용은 쓰지 마세요.
- ✓ 답변은 [답변]로 작성해주세요.

### 데이터필터링(수정 및 제거) 반자동

- ✓ 도어판 차체에서 분리하기 위해서 치우와 연결되어 있는 모든 전기 부품을 분리하는 방법은?
- ✓ 출입문 교체방법에서 전선 및 스위치 분리방법은?
- ✓ 스위칭 모듈 제작 과정에서 배이어셋을 렌지 분리방법은?
- ✓ 구성부품 고장수리 절차 및 방법에서 DCU 오류는?
- ✓ DCU 저항으로 유효하지 않은 데이터는 출고 시사 테로 초기화방법은?

내용보완

증복제거

### LLM 답변 요청

작성한 프롬프트를 LLM에 입력하여 답변 생성 요청

다양하고 창의적인 고품질 데이터 작성