



유저					
	유저ID	id	Domain	INT	NOT NULL
이름	name	Domain	VARCHAR(50)	NOT NULL	
유저정보	firebase_uid	login	VARCHAR(100)	NOT NULL	

유저테이블에 등록된 유저(firebase_uid)에 한해,
토큰을 발급받아 모델 학습 서비스를 이용 가능.

프로젝트					
	프로젝트ID	id	Domain	INT	NOT NULL
프로젝트명	name	Domain	VARCHAR(255)	NOT NULL	
설명	description	Domain	TEXT	NULL	
생성일	created_at	Domain	DATETIME	NULL	

데이터셋					
	데이터셋ID	id	Domain	INT	NOT NULL
소속 프로젝트	project_id	Domain	INT	NOT NULL	
데이터셋명	name	Domain	VARCHAR(255)	NOT NULL	
설명	description	Domain	TEXT	NULL	

데이터셋 스플릿					
	스플릿ID	id	Domain	INT	NOT NULL
데이터 구분	dataset_version_id	Domain	INT	NOT NULL	
비율	ratio	Domain	REAL	NOT NULL	
생성일	created_at	Domain	DATETIME	NOT NULL	

예셋					
	예셋ID	Key	Domain	INT	NOT NULL
타입	dataset_split_id	Domain	INT	NOT NULL	
파일 경로	storage_uri	Domain	TEXT	NOT NULL	
무결성	sha256	unique	CHAR(64)	NOT NULL	
파일 크기	bytes	Domain	BIGINT	NULL	
이미지 폭	width	Domain	INT	NULL	
이미지 높이	height	Domain	INT	NULL	
비디오 길이	duration_ms	Domain	INT	NULL	
비디오 FPS	fps	Domain	REAL	NULL	
비디오 프레임	frame	Domain	INT	NULL	
비디오 코덱	codec	Domain	TEXT	NULL	
생성일	created_at	Domain	DATETIME	NOT NULL	

유저테이블에 등록된 유저(firebase_uid)에 한해,
토큰을 발급받아 모델 학습 서비스를 이용 가능.

모델					
	모델ID	id	Domain	INT	NOT NULL
프로젝트ID	project_id	Domain	INT	NOT NULL	
모델명	name	Domain	VARCHAR(255)	NOT NULL	
작업	task	Domain	TEXT	NULL	
설명	description	Domain	TEXT	NULL	
생성일	created_at	Domain	DATETIME	NOT NULL	

데이터셋 버전					
	버전ID	id	Domain	INT	NOT NULL
소속 데이터셋	dataset_id	Domain	INT	NOT NULL	
온톨로지 버전ID	ontology_version_id	Domain	INT	NOT NULL	
버전 태그	version_tag	Domain	VARCHAR(50)	NOT NULL	
수정 불가 여부	is_frozen	Domain	BOOLEAN	NOT NULL	
생성일	created_at	Domain	DATETIME	NOT NULL	

온톨로지 버전					
	온톨로지 버전ID	id	Domain	INT	NOT NULL
프로젝트ID	project_id	Domain	INT	NOT NULL	
버전명	version_tag	Domain	VARCHAR(50)	NOT NULL	
수정 불가 여부	is_frozen	Domain	BOOLEAN	NOT NULL	
생성일	created_at	Domain	DATETIME	NOT NULL	

라벨 클래스					
	클래스ID	id	Domain	INT	NOT NULL
온톨로지 버전ID	ontology_version_id	Domain	INT	NOT NULL	
리밸이름	display_name	Domain	VARCHAR(100)	NOT NULL	
형태	shape_type	Domain	TEXT	NULL	
표시색상	color	Domain	VARCHAR(20)	NOT NULL	
포즈스켈레톤정의	keypoint_spec	Domain	JSON	NULL	
생성일	created_at	Domain	DATETIME	NOT NULL	

주석					
	주석ID	id	Domain	INT	NOT NULL
대상 예셋ID	asset_id	Domain	INT	NOT NULL	
라벨 클래스ID	label_class_id	Domain	INT	NOT NULL	
자동 생성 모델 ID	model_version_id	Domain	INT	NOT NULL	
형태	geometry_type	'bbox', 'polygon', 'keypoint'	ENUM	NOT NULL	
좌표/포인트 데이터	geometry	Domain	JSON	NOT NULL	
점검회여부	is_normalized	Domain	BOOLEAN	NOT NULL	
생성 주석	source	Domain	TEXT	NULL	
모델 신뢰도	confidence	Domain	REAL	NOT NULL	
라벨리	annotator_name	Domain	TEXT	NOT NULL	
생성	created_at	Domain	DATETIME	NOT NULL	

모델 아티팩트					
	모델 아티팩트ID	id	Domain	INT	NOT NULL
소속 모델 버전ID	model_version_id	Domain	INT	NOT NULL	
형식	format	Domain	TEXT	NOT NULL	
실행 환경	device	Domain	TEXT	NOT NULL	
정밀도	precision	Domain	FLOAT	NOT NULL	
파일 경로	storage_uri	Domain	TEXT	NOT NULL	
무결성	unique	Domain	CHAR(64)	NOT NULL	
파일 크기	size_bytes	Domain	BIGINT	NOT NULL	
ONNX용 오프셋	opset	Domain	INT	NULL	
IR 버전	ir_version	Domain	INT	NULL	
환경 호환성	compat	Domain	JSON	NULL	
생성일	created_at	Domain	DATETIME	NOT NULL	

평가지표					
	평가ID	id	Domain	INT	NOT NULL
조회평균	f1_score	Domain	FLOAT	NOT NULL	
정밀도	precision	Domain	FLOAT	NOT NULL	
재현율	recall	Domain	FLOAT	NOT NULL	
평균정확도평균50	mAP_50	Domain	FLOAT	NOT NULL	
평균정확도평균95	mAP_50_95	Domain	FLOAT	NOT NULL	

배포					
	배포ID	id	Domain	INT	NOT NULL
모델 버전ID	model_version_id	Domain	INT	NOT NULL	
이름	name	Domain	VARCHAR(50)	NOT NULL	
배포 위치	target	'edge', 'cloud'	ENUM	NOT NULL	
엔드포인트	endpoint_url	Domain	VARCHAR(100)	NOT NULL	
배포일	deployed_at	Domain	DATETIME	NOT NULL	

AI 학습·어노테이션 관리 시스템 ERD 보고서

시스템 개요

본 시스템은 **Auto Annotation**, **Self Annotation**, 모델 학습, 모델 변환, 평가, 증분학습까지를 지원하는 사내 내부용 AI 데이터 관리 툴입니다.

특징

- 단일 서버 환경, 컨테이너 미사용
- DB에는 영속 데이터만 저장 (dataset, annotation, model, version)
- 학습 progress는 인메모리(WebSocket/SSE)
- 체크포인트 파일은 S3/Local 경로에 모두 저장
- Ontology(논리관계)를 통해 라벨 정의 버전 관리
- Incremental Learning(증분학습) 및 지식 보존(Distillation/Rehearsal) 전략 지원

데이터 흐름 요약

```
[데이터 업로드]
↓
[dataset_version (draft)] 생성
↓
[asset] 등록 → [auto/self annotation]
↓
[split(train/val/test)] 배분 → freeze
↓
[model_version] 학습 (ontology, dataset_version 참조)
↓
[checkpoints] 파일 저장 (5epoch마다)
↓
[best/final 모델 → model_artifact 등록]
↓
[증분학습 시 parent_model_version 참조]
```

핵심 개념 요약

개념	역할	비고
Dataset / DatasetVersion	데이터셋의 버전 스냅샷	frozen 후 수정 불가
Asset	이미지/영상 등 개별 데이터	dataset_split에 속함
Annotation	라벨링 정보 (bbox, keypoints 등)	human/model 구분

개념	역할	비고
Ontology / LabelClass	라벨 정의의 버전 스냅샷	클래스, 형태, 색상 등 포함
Model / ModelVersion	학습/추론 모델 및 버전 관리	parent_model_version으로 종분 연결
ModelArtifact	학습 결과물 파일 (ckpt, onnx 등)	중간 checkpoint는 DB 비저장
Run(인메모리)	현재 실행 중 상태 (progress 등)	WS로 실시간 업데이트

2. ERD 개요

```

project
  |- dataset — dataset_version — dataset_split — asset — annotation —
  |   label_class — label_ontology_version
  |
  |— model — model_version — model_artifact
      |—(self) parent_model_version_id
      |   init_from_artifact_id ——————
  
```

3. 테이블 상세 설계

1. project

컬럼	타입	제약	설명
id	UUID	PK	프로젝트 ID
name	VARCHAR(255)	NOT NULL	프로젝트명
description	TEXT		설명
default_ontology_version_id	UUID	FK → label_ontology_version.id	기본 온톨로지
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

2. dataset

컬럼	타입	제약	설명
id	UUID	PK	데이터셋 ID
project_id	UUID	FK → project.id	소속 프로젝트
name	VARCHAR(255)	NOT NULL	데이터셋명
description	TEXT		설명

3. dataset_version

컬럼	타입	제약	설명
id	UUID	PK	버전 ID
dataset_id	UUID	FK → dataset.id	상위 데이터셋
version_tag	VARCHAR(50)	UNIQUE	예: "v1.0"
ontology_version_id	UUID	FK → label_ontology_version.id	사용 온톨로지
is_frozen	BOOLEAN	DEFAULT false	수정 불가 여부
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

4. dataset_split

컬럼	타입	제약	설명
id	UUID	PK	스플릿 ID
dataset_version_id	UUID	FK → dataset_version.id	소속 버전
split	TEXT	CHECK('train','val','test','unassigned')	데이터 구분
ratio	REAL		비율
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

5. asset

컬럼	타입	제약	설명
id	UUID	PK	에셋 ID
dataset_split_id	UUID	FK → dataset_split.id	소속 스플릿
type	TEXT	CHECK('image','video')	에셋 타입
storage_uri	TEXT	NOT NULL	파일 경로(S3/Local)
sha256	CHAR(64)	UNIQUE	무결성
bytes	BIGINT		파일 크기
width	INT		이미지 폭
height	INT		이미지 높이
duration_ms	INT		(video) 길이
fps	REAL		(video) FPS
frame_count	INT		(video) 프레임 수

컬럼	타입	제약	설명
codec	TEXT		(video) 코덱
exif	JSONB		메타데이터
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

6. label_ontology_version

컬럼	타입	제약	설명
id	UUID	PK	온톨로지 버전 ID
project_id	UUID	FK → project.id	프로젝트
version_tag	VARCHAR(50)		버전명
is_frozen	BOOLEAN		수정 불가
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

7. label_class

컬럼	타입	제약	설명
id	UUID	PK	클래스 ID
ontology_version_id	UUID	FK → label_ontology_version.id	온톨로지 버전
key	VARCHAR(64)	UNIQUE	내부식별자
display_name	VARCHAR(100)		라벨이름
shape_type	TEXT	CHECK('bbox','polygon','mask','keypoints','polyline')	형태

컬럼	타입	제약	설명
color	VARCHAR(20)		표시색상
keypoint_spec	JSONB		포즈스켈레톤정의
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

8. annotation

컬럼	타입	제약	설명
id	UUID	PK	어노테이션 ID
asset_id	UUID	FK → asset.id	대상 이미지
label_class_id	UUID	FK → label_class.id	라벨 클래스
geometry_type	TEXT	CHECK('bbox','polygon','mask','keypoints','polyline')	형태
geometry	JSONB	NOT NULL	좌표/키포인트 데이터
is_normalized	BOOLEAN	DEFAULT false	정규화 여부
source	TEXT	CHECK('human','model')	생성 주체
confidence	REAL		모델 신뢰도
model_version_id	UUID	FK → model_version.id NULL	자동 생성 모델
annotator_name	TEXT		라벨러

컬럼	타입	제약	설명
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

9. model

컬럼	타입	제약	설명
id	UUID	PK	모델 ID
project_id	UUID	FK → project.id	프로젝트
name	VARCHAR(255)	NOT NULL	모델명
task	TEXT	CHECK('det','pose','seg','cls','ocr')	태스크
description	TEXT		설명
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

10. model_version

컬럼	타입	제약	설명
id	UUID	PK	모델 버전 ID
model_id	UUID	FK → model.id	상위 모델
parent_model_version_id	UUID	FK(self) NULL	증분학습 부모
init_from_artifact_id	UUID	FK → model_artifact.id NULL	초기 가중치
training_data_version_id	UUID	FK → dataset_version.id	학습 데이터
ontology_version_id	UUID	FK → label_ontology_version.id	라벨 체계
framework	VARCHAR(50)		PyTorch 등
framework_version	VARCHAR(50)		버전
metrics	JSONB		평가 지표
training_config	JSONB		하이퍼파라미터
incremental_info	JSONB		증분 전략
is_frozen	BOOLEAN		수정 불가
created_at	TIMESTAMPTZ	DEFAULT now()	생성일시

11. model_artifact

컬럼	타입	제약	설명
id	UUID	PK	아티팩트 ID
model_version_id	UUID	FK → model_version.id	소속 모델 버전
format	TEXT	CHECK('ckpt','onnx','trt','torchscript','openvino','tflite','onnx-external')	형식
device	TEXT	CHECK('cpu','cuda','tensorrt','mps','npu','web')	실행 환경
precision	TEXT	CHECK('fp32','fp16','int8','mixed')	정밀도
storage_uri	TEXT	NOT NULL	파일 경로
sha256	CHAR(64)		무결성
size_bytes	BIGINT		파일 크기
opset	INT		ONNX 전용
ir_version	INT		IR 버전
compat	JSONB		환경 호환성
created_at	TIMESTAMPTZ	DEFAULT now()	생성일 시



온톨로지(ontology) 운영 요약

항목	설명
정의	라벨 체계(클래스, 형태, 구조 등)의 버전 스냅샷
역할	데이터셋·모델의 라벨 규칙 일관성 보장
관계	dataset_version ↔ ontology_version (1:1), model_version ↔ ontology_version (1:1)
변경 시	클래스 추가/수정 시 새 ontology_version 발행
증분학습	새 클래스 추가 시 model_version.parent_model_version_id로 연결, ontology 버전 갱신

체크포인트 / 학습 진행률 관리

항목	저장 위치	비고
중간 체크포인트 (5epoch마다)	파일 (S3/local)	DB 비저장
best/final checkpoint	파일 + model_artifact 등록	대표 모델
manifest.json	파일	체크포인트 메타 정보
실행 상태(progress)	인메모리(WS/SSE)	DB 비저장

버전 관리 및 증분학습 관계

- 새 데이터 추가 → dataset_version v1.1 생성
- 동일 클래스 체계 → 기존 ontology_version 유지
- 새 클래스 추가 → 새 ontology_version 발행
- 증분학습 시:
 - model_version.parent_model_version_id = 이전 모델
 - init_from_artifact_id = 이전 모델의 best.ckpt
 - incremental_info = {"strategy": "distillation + rehearsal"}



운영 정책 요약

정책	설명
dataset_version.is_frozen = true	해당 버전 이후 수정 금지
새 에셋 추가 시	항상 DRAFT 버전에서 추가 후 freeze
온톨로지 변경 시	새 버전 발행, 데이터/모델 양쪽 갱신
체크포인트 관리	파일 기반 관리, DB는 best/final만 기록
인덱스	asset(dataset_split_id), annotation(asset_id), model_version(model_id) 등

결론

이 설계는 다음을 충족합니다:

- 데이터셋·모델·라벨의 **재현성과 일관성 보장**
- **체크포인트는 파일, 진행률은 인메모리, DB는 영속 메타데이터만**
- 증분학습, 라벨 버전, 모델 버전 관리까지 자연스럽게 연계 가능