

# Compliance dossier — training data (AI Act, Article 10)

**Automatically generated skeleton** from the dataset analysis. The "to be completed by the expert" blocks require human review (provenance, compliance judgment, mitigation measures).

## 1. Identity & purpose

- **System / model:** \_\_\_\_\_
- **Dossier version:** \_\_\_\_\_
- **Analysis date:** \_\_\_\_\_
- **Dataset:** src
- **Intended purpose:** \_\_\_\_\_

\_To be completed by the expert — intended purpose and what the data is meant to represent\_

## 2. Provenance & legal basis

Origin of each source, licences/contracts, and — for personal data — initial purpose and GDPR legal basis.

\_To be completed by the expert — provenance per source (separate flow §16), licences, GDPR legal basis\_

## 3. Composition (filled automatically)

- **Volume:** 303 rows, 14 columns.
- **Variables:**

| Column   | Type   |
|----------|--------|
| age      | DOUBLE |
| sex      | DOUBLE |
| cp       | DOUBLE |
| trestbps | DOUBLE |
| chol     | DOUBLE |
| fbs      | DOUBLE |
| restecg  | DOUBLE |
| thalach  | DOUBLE |
| exang    | DOUBLE |
| oldpeak  | DOUBLE |
| slope    | DOUBLE |
| ca       | DOUBLE |
| thal     | DOUBLE |
| target   | BIGINT |

- **Declared sensitive attributes:** age, sex.

\_To be completed by the expert — geographic / contextual / behavioural scope\_

## 4. Preparation (automatic observations)

---

- **Missing values:** ca (4) -> handling strategy to be documented.

\_To be completed by the expert — transformation log: collection, cleaning, labelling, enrichment, aggregation\_

## 5. Quality (filled automatically)

---

- **Completeness:** 4 missing value(s) (ca).
- **Accuracy / outliers:** chol (1).

\_To be completed by the expert — representativeness and relevance to the purpose\_

## 6. Bias

---

### Automatic analysis

| Characteristic | Measure | Target    | Verdict   |
|----------------|---------|-----------|-----------|
| Completeness   | 99.9 %  | "e 95.0 % | compliant |
| Uniqueness     | 100.0 % | "e 99.0 % | compliant |
| Validity       | 100.0 % | "e 98.0 % | compliant |

**sex** — « target »: **disparate impact** (ratio) 2.41 · **statistical parity** (gap) 13.1 % (!)

| Group | Count | « target » rate |
|-------|-------|-----------------|
| 1.0   | 206   | 22.3 %          |
| 0.0   | 97    | 9.3 %           |

### Judgment & measures

\_To be completed by the expert — impact on health/safety/fundamental rights and detection/prevention/mitigation measures\_

## 7. Gaps & limitations

---

Detected points to examine as potential gaps:

- [low] missing — ca
- [low] missing — tha1
- [low] outliers — chol
- [high] bias — sex

\_To be completed by the expert — identified and addressed gaps; out-of-scope uses\_

## 8. Governance & traceability

---

- **Analysis tool version:** 0 . 4 . 1
- **Dataset fingerprint (schema + volume):** a5aed4387668f53f

\_To be completed by the expert — responsibilities, audit log, versioning, maintenance/updates\_