AI Researchers
Embedding Responsible AI in Research

AI READINESS TOOLKIT

UC3-01 Use case 3: Data preparation and content analysis

Data quality and audit: training dataset

Seventeen fictional training records used in the workshop exercise

AT A GLANCE

The seventeen fictional participant records used in the Data quality and audit workshop. The set contains deliberate rule failures, formatting inconsistencies, a review-only outlier, an exact duplicate and a pair of conflicting records, so that participants can practise an audit before any cleaning takes place.

TYPE
Training dataset, workshop annex

FORMAT
PDF

CONTRIBUTING PARTNER
FEP, University of Porto

WHO IT IS FOR
Facilitators and participants running the Data quality and audit workshop

FRAMEWORK BARRIERS
Trust in AI, Know-how

HUMAN-ONLY BOUNDARY

No unpublished or sensitive data in unapproved tools. Interpretation and conclusions stay human.

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

Data Quality and Audit - Synthetic Raw Dataset

17 fictional participant records × 11 variables: quantitative data only

Exercise instruction: this raw dataset contains intentional quality problems. Audit and document them before proposing any change.

PARTICIPANT_ID	AGE	COUNTRY	DISCIPLINE	TRAINING_HOURS	PRE_SCORE	POST_SCORE	COMPLETED	SATISFACTION	START_DATE	COMPLETION_DATE
P001	29	Portugal	Economics	6	62	78	Yes	4	2026-03-03	2026-03-20
P002	34	PT	Engineering	8	70	85	Y	5	2026-03-04	2026-03-21
P003		Spain	Social Sciences	4	55	71	yes	4	2026-03-05	2026-03-22
P004	222	France	Economics	5	66	74	Yes	3	2026-03-05	2026-03-23
P005	41	Germany	Data Science	7,5	73	88	Yes	5	2026-03-06	2026-03-24
P006	38	portugal	Engineering	-2	68	80	Yes	4	2026-03-06	2026-03-24
P007	27	Italy	Economics	60	64	69	Yes	3	2026-03-07	2026-03-25
P008	45	Spain	data science	9	81	108	Yes	5	2026-03-07	2026-03-25
P009	31	France	Social Sciences	6	59	76	Maybe	4	2026/03/08	2026/03/26
P010	36	Germany	Engineering	7	72		Yes	4	2026-03-08	2026-03-26
P011	52	Portugal	Economics	3	61		No		2026-03-09	
P012	24	Italy	Data Science	5	67	79	Yes	4	2026-03-12	2026-03-10
P013	40	Spain	Biochemistry	8	75	86	Yes	7	2026-03-10	2026-03-27
P014	33	Portugal	Engineering	6	69	82	Yes	4	2026-03-11	2026-03-28
P015	29	France	Social Sciences	5	63	77	Yes	4	2026-03-11	2026-03-29
P001	29	Portugal	Economics	6	62	78	Yes	4	2026-03-03	2026-03-20
P014	33	Portugal	Engineering	6	69	82	Yes	6	2026-03-11	2026-03-30

About this resource

TITLE	Data quality and audit: training dataset
TOOLKIT REFERENCE	UC3-01, use case 3: Data preparation and content analysis
DEVELOPED BY	FEP University of Porto, Fábio Duarte
SUGGESTED CITATION	AI Researchers (2026). Data quality and audit: training dataset. AI Readiness Toolkit, UC3-01. FEP University of Porto, Fábio Duarte. airesearchers.eu
QUALITY ASSURANCE	Reviewed by the AI Researchers consortium for quality, categorisation and alignment with the AI Engagement Framework. External peer review by senior academics is under way, and the item may be updated as a result.
VERSION	1.0, November 2026
PROJECT	AI Researchers: Embedding Responsible AI in Research. Erasmus+ Cooperation Partnership in Higher Education, KA220-HED-66835D2B. October 2025 to March 2028.
CONSORTIUM	University of Lodz (coordinator), FEP University of Porto, UPF Barcelona School of Management, ACEEU, NIPI (OPI PIB), European E-learning Institute and Momentum Marketing Services

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

AI Researchers
Embedding Responsible AI in Research