SUPPLEMENTARY INFORMATION: CODEBOOK

Article title: No need, or no category? Recording refusal of generative AI in European official statistics
Journal: AI & SOCIETY

Author names, affiliations and contact details are withheld from this file because the article is under double-anonymous review. They are supplied through the submission interface.

1. WHAT THESE FILES ARE

Online Resource 1 (ESM_1.csv) The analysis dataset. One row per country and population subgroup for which both survey items are complete: 960 rows.

Online Resource 2 (ESM_2.csv) Country-level results. Source of Table 4 and Figures 1, 3 and 4.

Online Resource 3 (ESM_3.csv) Subgroup results with interval estimates and exact sign tests. Source of Table 3 and Figure 2.

Online Resource 4 (this file) Definitions, provenance and method notes.

2. DATA SOURCE

Two tables published by Eurostat, both for reference year 2025, both released 17 April 2026 and accessed 13 August 2026. Both are freely available and require neither an account nor a licence.

isoc_ai_iaiuxr	Individuals: reasons for not using generative AI tools. Read at unit PC_IND_IUAIX, percentage of AI non-users. https://doi.org/10.2908/isoc_ai_iaiuxr
isoc_iiu_iuxr	Individuals: reasons for not using the internet. Read at unit PC_IND_IU3X, percentage of internet non-users.

Reuse of Eurostat statistical data is authorised for commercial and non-commercial purposes provided the source is acknowledged, under the Commission Decision of 12 December 2011 on the reuse of Commission documents (2011/833/EU).

Eurostat revises online tables in place. Anyone repeating this analysis should compare the "updated" timestamp of the payload against 17 April 2026.

3. VARIABLES IN ONLINE RESOURCE 1

geo	Country code, Eurostat geo dimension
ind_type	Population subgroup code, Eurostat ind_type dimension
time	Reference year, 2025 throughout
Internet item, each as a percentage of internet non-users:	
inet_no_need	No need
inet_too_difficult	Difficult to use
inet_disability	Issues with accessibility for persons with impairments or disabilities
inet_privacy_security	Concerns about security or privacy
inet_cost	Cost of the connection or equipment

inet_opposed_distrust	Opposed to it or do not trust the information
inet_other	Other reasons

AI item, each as a percentage of AI non-users:

ai_no_need	No need
ai_unaware	Did not know they existed
ai_no_skill	Did not know how to use
ai_privacy_security	Concerns about privacy, security or safety
ai_other	Other

Category labels above are Eurostat's published labels. The questionnaire wording differs slightly and is quoted in full in Section 4 of the article.

4. DERIVED QUANTITIES IN ONLINE RESOURCES 2 AND 3

no_ai_counterpart_sum	inet_opposed_distrust + inet_disability + inet_cost. These are the three internet categories with no counterpart in the AI item.
recording_gap	no_ai_counterpart_sum minus ai_other.
recording_gap_rescaled	The same difference after each country's seven internet shares are proportionally rescaled to total 100, which removes the arithmetic inflation caused by the multiple-response format.
internet_shares_sum	Sum of the seven internet shares. Exceeds 100 because the item permits multiple responses.
ci_lower, ci_upper	95 per cent percentile bootstrap interval, 10,000 replicates resampling countries, seed 20260813.
sign_test_p_exact	Exact two-sided sign test on the paired differences.

5. TWO METHOD NOTES THAT AFFECT REPRODUCTION

Arithmetic. Eurostat publishes these shares to two decimal places, which is exact in decimal but not representable in binary floating point: 3.55 is held as 3.5499999999999998, so rounding it in floating point sends it down rather than up. All arithmetic is carried out in exact decimal and each reported figure is rounded once, half away from zero. Reproducing the tables with floating-point arithmetic will shift several cells by 0.1.

The summed quantity. The internet item allows multiple responses, so summing three of its categories counts a respondent once for each category endorsed. That sum is an upper bound on the share of non-users citing at least one of the three, and is not that share. The share itself requires the joint distribution across categories, which published aggregates do not carry. The rescaled column is a conservative approximation under an explicit proportionality assumption, not a proven lower bound. Both are reported.

6. SAMPLE FUNNEL

Each step is a data availability constraint, not an analytic choice. Nothing is imputed.

AI item published	36 countries
Internet item published	24 countries
Both items pairable	21 countries
Country-level (IND_TOTAL) cell present	20 countries
Analysed at country level	19 countries

Ireland and Luxembourg publish no internet reason cells for the required subgroups. Norway publishes six of the seven internet categories but omits the disability category, so no Norwegian cell is complete on both items. The Netherlands publishes subgroup cells but no country total. Finland reports zero on every internet reason category.

Counts of published countries exclude the European aggregates present in each table's country dimension.

7. SUBGROUP CODES USED IN ONLINE RESOURCE 3

Y65_74	Aged 65 to 74	overlap stratum
Y55_74	Aged 55 to 74	overlap stratum
I0_2	Low or no formal education	overlap stratum
RETIR_OTHER	Retired or not in labour force	overlap stratum
IND_DEG3	Rural areas	overlap stratum
I5_8	High formal education	contrast stratum
Y25_54	Aged 25 to 54	contrast stratum
IND_DEG1	Cities	contrast stratum

Overlap strata are those in which internet non-users and AI non-users are both well represented. Contrast strata are those in which they overlap least.

8. ANALYSIS CODE

The code that downloads the source tables, builds Online Resource 1 and computes Online Resources 2 and 3 is archived in the repository named in the data availability statement, which is supplied through the submission interface. It uses only the Python standard library for the analysis, so no result depends on a statistical package version.