

Annex C. Glossary: plain labels and the codebook's codes

This annex is the only place where the codebook's variable and value codes appear; Annex B names candidates by their identifiers in the frame files. It maps every plain label this paper prints to the codebook's variable and value, so that a methods reader can check any figure against the dataset. The program that fills this paper prints Tables C1 to C5 from the same dictionary it uses everywhere else.

Table C1. Items and their values.

Plain label (item)	Codebook variable	Plain label of a value: codebook value
Training-content summary	tcs_published	
		Published, identified as the summary: identified_summary_public
		Related document only: adjacent_document_public
		Summary behind a login or another barrier: identified_summary_gated
		Related document behind a login or another barrier: adjacent_document_gated
		Summary mentioned, with no location a reader can reach: referenced_no_location

Plain label (item)	Codebook variable	Plain label of a value: codebook value
Summary or related document served live when read	tcs_live_serve_status	Served live when read: served_live
		Reached only in an archive copy: archive_only
		Not retrieved by our reading, live or from an archive copy: neither
Where the summary or related document sits on the provider's site	tcs_findability	Linked from the site's main menus: nav_named_document
		One click from a page in the main menus: one_click_from_nav_named
		Three clicks from the home page, through standing pages: deeper_in_site_hierarchy
		Reached only through a dated post: single_post_link
		Not reached within three clicks of the home page: not_found_in_bounded_crawl
		No document to read: none_stated
The document names the model it covers	tcs_identifies_model	Could not be determined (the link crawl could not be completed): unknown
		Names specific models: names_specific_models
		Names a model family: names_model_family
		Covers the provider in general: covers_provider_generally
The document states the kinds of training content	tcs_modalities_stated	No document to read: none_stated
		Lists them: stated_enumerated
		States them in general terms: stated_general
The document states the categories of sources	tcs_source_categories_stated	No document to read: none_stated
		Lists them: stated_enumerated
		States them in general terms: stated_general
The document names particular sources	tcs_named_sources_stated	No document to read: none_stated
		Names datasets: named_datasets
		Names websites or licensors: named_domains_or_licensors
		Gives examples of kinds of source only: class_examples_only
		Names no source: none_named
The document refers to the AI Office template	tcs_template_reference	No document to read: none_stated
		Names and cites the template: template_named_and_cited
		Mentions the template without citing it: template_mentioned_without_citation
		No reference to the template: no_template_reference
The document carries a date or a version	tcs_currency	No document to read: none_stated
		No document to read: none_stated
Copyright policy	copyright_policy_public	

Plain label (item)	Codebook variable	Plain label of a value: codebook value
		Policy published, open without an account: <code>policy_document_public</code>
		Policy said to exist, not published: <code>policy_described_not_published</code>
		Statements about copyright, no policy mentioned: <code>copyright_statement_only</code>
The copyright policy refers to EU copyright law	<code>copyright_policy_eu_law_referenced</code>	
		Cites the EU text and data mining reservation: <code>dsm_article_4_3_cited</code>
		Refers to EU copyright law: <code>union_copyright_law_referenced</code>
		Refers to copyright in general: <code>copyright_referenced_generally</code>
		No policy document to read, or no reference: <code>none_stated</code>
Route for rightsholders to reserve their rights against training	<code>rights_reservation_mechanism_documented</code>	
		A specific public route (crawler name, opt-out form or signal): <code>route_public_and_specific</code>
		A route described without specifics: <code>route_described_no_location</code>
Statement on the provider's own web crawling for training	<code>tcs_scraping_disclosed</code>	
		Crawler named, with how to block it: <code>crawler_named_and_exclusion_documented</code>
		Crawler named only: <code>crawler_named_only</code>
		Web collection stated, crawler not named: <code>scraping_stated_crawler_unnamed</code>
The copyright policy carries a date or a version	<code>copyright_policy_currency</code>	
		No policy document to read: <code>none_stated</code>
Statement that technical documentation is kept for the authorities	<code>techdoc_regulator_route_documented</code>	
		Statement, with a route for the authorities: <code>statement_and_route_public</code>
		Statement only: <code>statement_only</code>
Documentation for developers who build on the model	<code>downstream_integrator_documentation_public</code>	
		Behind an account: <code>account_required</code>
		Available on request: <code>on_request_route_documented</code>
Model card or other per-model documentation	<code>model_card_public</code>	
The model card carries a date or a version	<code>model_card_currency</code>	
Conditions of the open-source exemption stated	<code>oss_exemption_conditions_documented</code>	
		License and all three public elements stated: <code>all_elements_documented</code>
		Some of the elements stated: <code>some_elements_documented</code>
Statement on whether the model has systemic risk	<code>systemic_risk_designation_stated</code>	
		States the model has systemic risk: <code>states_systemic_risk_model</code>
		States the model does not have systemic risk: <code>states_not_systemic_risk</code>
		States a training-compute figure, no designation: <code>compute_figure_stated_no_designation</code>

Plain label (item)	Codebook variable	Plain label of a value: codebook value
Statement that users are told they are interacting with an AI system	interaction_disclosure_documented	
		Statement, with the notice or mechanism described: mechanism_documented
		Statement, no mechanism described: stated_without_mechanism
		States no notice is given because the AI nature is obvious: obviousness_relied_on
That statement carries a date or a version	interaction_disclosure_currency	
Whether the provider's documents say the system produces content	system_generative_status	
		Produces content, in explicit words: generation_stated_explicit
		Produces content, in ordinary product language: generation_stated_functional
		States it does not produce content: generation_stated_negative
		The documents read do not settle it: generation_unknown
The coded generative status agrees with the system's selection	designation_status_agreement	
		Agrees: agrees
Statement that the system's outputs are marked	marking_disclosure_published	
		Outputs marked, in a machine-readable form: marking_stated_machine_readable
		Outputs marked, form not stated: marking_stated_form_unspecified
		Marking planned: marking_stated_planned
The marking method is named	marking_method_named	
		Names a specification or standard: named_specification
		Describes a method, names no specification: method_described_unnamed
		No method stated: unstated
		No marking statement: none_stated
The kinds of output the marking covers	marking_modalities_covered	
		Lists the kinds of output marked: modalities_enumerated
		States that all outputs are marked: all_outputs_stated
		Not stated: unstated
		No marking statement: none_stated
List of the kinds of output stated as marked	marking_modalities_list	
Which models carry the marking, and from when	marking_scope_dated	
		Models and start date stated: scope_models_and_date_stated
		Models stated, no date: scope_models_stated_undated
		Date stated, models not: date_stated_models_unspecified
		Neither stated: unstated
		No marking statement: none_stated
Public route to check an output for the marking	marking_verification_route_documented	
		A public route to check outputs: public_route_documented
		A route behind an account: route_documented_account_required

Plain label (item)	Codebook variable	Plain label of a value: codebook value
		A route described, with no location: route_described_no_location
The marking statement carries a date or a version	marking_disclosure_currency	
		No marking statement: none_stated
Guidance for deployers on disclosing AI-generated content	deployer_guidance_published	
		Guidance citing Article 50: guidance_cites_article_50
		Guidance without a reference to Article 50: guidance_without_article_reference
What the provider says about its own role under the Act	provider_gpai_self_identification	
		States it provides a general-purpose AI model under the Act: states_gpai_provider
		Mentions the Act, states no role: references_act_without_role
		States it is not such a provider: declared_absent
		Does not mention the Act: no_reference_to_act
A page about the EU AI Act	eu_act_resource_page_public	
Statement on adhering to the code of practice	code_of_practice_signature_stated	
		States adherence, names chapters: signature_stated_chapters_named
		States adherence, names no chapter: signature_stated_chapters_unnamed
		States it does not adhere: declared_absent
Legal entity named by the provider	provider_legal_entity_stated	
How the model is released (open or closed weights)	release_form	
		Open weights: open_weight_stated
		Closed weights: closed_weight_stated
		Mixed: mixed_stated
		Not stated: unstated
Earliest documented date of outside access to the model	model_public_availability_first_documented_date	
Earliest documented public release date of the system	system_public_availability_first_documented_date	
A page in another language met during the reading	non_english_surface_encountered	
		Met and recorded: encountered_recorded
		None met: none_encountered
Own reading: when the AI nature of a system is obvious (Article 50(1))	interp_quote_50_1_obviousness	
Own reading: marking as far as technically feasible (Article 50(2))	interp_quote_50_2_feasibility	
Own reading: assistive editing functions (Article 50(2))	interp_quote_50_2_carveout	
Own reading: the exemptions and limits in Article 50(4)	interp_quote_50_4_scope	
Own reading: a sufficiently detailed summary (Article 53(1)(d))	interp_quote_53_1d_sufficiency	
Own reading: the AI Office template (Article 53(1)(d))	interp_quote_53_1d_template	
Own reading: the open-source exemption (Article 53(2))	interp_quote_53_2_exemption	

Plain label (item)	Codebook variable	Plain label of a value: codebook value
Own reading: systemic risk (Article 51(2))	interp_quote_51_2_systemic_risk	

Note: the plain labels the paper prints, mapped to the codebook's variable names and values.

Table C2. Values shared by many items.

Plain label	Codebook value (any item, unless the item maps it otherwise above)
Nothing found in the places we looked	none_stated
Could not be determined	unknown
Could not be opened	unretrievable
Behind a login or another barrier	gated
The provider's documents disagree	conflicting
The provider states it does not publish this	declared_absent
Not asked, because a rule removed the question	not_applicable
Not stated in the document	unstated
No provider in this group	(empty_group)
Dated and versioned	dated_and_versioned
Dated, no version	dated_only
Versioned, no date	versioned_only
Neither dated nor versioned	undated_unversioned
Published, open without an account	public_url
Published, open without an account	public_page
Mentioned, with no location a reader can reach	referenced_no_location

Table C3. Deadline groups, readings and index bands.

Plain label	Codebook token
Deadline had passed (training-content summary)	due_at_collection
Not yet due (training-content summary)	not_due_at_collection
Release date not settled (training-content summary)	not_determinable
Deadline had passed (marking)	due_at_collection
Not yet due (marking)	not_due_at_collection
System release date not settled (marking)	not_determinable
Reading A (application date): already due	due_at_collection (every record on which the items are coded)
Reading B (deferral): model released before 2 August 2025, due 2 August 2027	designated model available before 2 August 2025: not due until 2 August 2027
Reading B (deferral): model released on or after 2 August 2025, already due	designated model available on or after 2 August 2025: falls due as on the application-date reading
Reading B (deferral): model release date not settled	documents did not settle when the designated model became publicly available
Reading B (deferral): in no group	beside the three groups (D19 item 7): art53_stratum not_applicable, in no group
Full public record	full_public_record
Substantial public record	substantial_public_record
Partial public record	partial_public_record
Thin public record	thin_public_record
Minimal public record	minimal_public_record
Not computed	not_applicable

Table C4. The rules that remove or answer a question.

Codebook dependency rule	What it does, in plain words
D1a	no summary or related document, so the questions about its content and place are answered from that
D1b	the summary question could not be settled, so its dependent questions take the same answer
D1c	the summary could not be opened, so its dependent questions take the same answer
D2	no marking statement, so the questions about the marking are answered from that
D3	no system qualified for study, so the questions about a system are not asked
D4	the system is documented as not producing content, so the marking questions are not asked
D4b	the documents do not settle whether the system produces content, so the marking questions are not asked
D4c	the coded generative status contradicts the system's selection, so the marking questions are not asked
D5	no copyright policy document, so the questions about it are answered from that
D6	no model card, so its date question is answered from that
D7	the provider was not admitted as a provider of an AI system, so the system questions are not asked
D7b	the provider was not admitted as a provider of a general-purpose AI model
D8	no document address exists to test, so the live-serving question is not asked

Note: a provider "removed by a rule" is one for whom one of these rules took the question away; it leaves that item's denominator.

Table C5. Events of the collection and reasons for rejecting a candidate.

Plain label	Code
Reading withdrawn	attempt_withdrawn
Record coded again on its own copies	record_given_again
Reading set aside	pass_voided
Reader's note kept beside the record	note_line_returned
Missing read line completed	read_line_completed
Record read again in full on the day of the final-day re-check	recollection
Reading abandoned	pass_abandoned
Difference noted at the final-day re-check	recheck_difference
Line range of a quotation from a PDF recorded	quote_lines
All events	all_events
Rejected: not a provider of AI models on its own pages	g1_not_a_model_provider
Rejected: its own pages did not settle whether its model is general-purpose	g2_not_established
Rejected: its own pages describe its models as built for one task or a narrow set	g2_not_general_purpose
Rejected: not the party providing the model	g3_not_the_provider
Rejected: no English-language page of its own that a reader could reach	g6_not_first_party_reachable

Note: the event types of the events file and the reason codes of the eligibility tests.