EUDR Legality Document Assessment Methodology

How TRACT scores a document submitted as Article 9(1)(h) legality evidence under the EU Regulation on Deforestation-free Products (EUDR)

TRACT

Version 0.2

Table Of contents

1 Purpose and Scope

What the score is, and what it is not

The TRACT EUDR legality analyser produces a single **assessed confidence score (0–100)** for a document submitted as evidence of legality under Regulation (EU) 2023/1115.

It answers: how much weight should a competent reviewer place on this document as Article 9(1)(h) evidence, given what is visible on its face and what is known about the origin?

It does not answer: is this document genuine? Nothing is checked against an issuing register, and for most producer countries no public register exists to check against. A high score means *"this instrument is of a strong type, well-formed, from a plausible authority, in a context where it can be corroborated"* — not *"this document has been verified."*

> The distinction is load-bearing. Output must never state or imply that a document has been verified, confirmed genuine or validated, and every assessment carries an explicit `verification_status` sentence recording that no register was contacted.
>
> **The score is an input to human review, not a compliance determination.**

Legal basis

Every element of the method is anchored in a provision of the Regulation. The table below sets out which provision drives which part of the assessment.

Table 1: Provisions of Regulation (EU) 2023/1115 and their role in the method

PROVISION	ROLE IN THE METHOD
Art. 9(1)(h)	The obligation being evidenced: "adequately conclusive and verifiable information" that the relevant commodities were produced in accordance with the relevant legislation of the country of production, including any arrangement conferring the right to use the area. This is what the *evidential sufficiency* component measures against.
Art. 2(40)(a)–(h)	Defines "relevant legislation" as eight categories: land use rights; environmental protection; forest-related rules; third parties' rights; labour rights; human rights under international law; free, prior and informed consent; and tax, anti-corruption, trade and customs. Pass 1 maps every document against all eight.
Art. 10(2)(g)	"The source, reliability, validity, and links to other available documentation" of the Article 9(1) information — the basis for the four document-level components.

PROVISION	ROLE IN THE METHOD
Art. 10(2)(h)	Country-level concerns: corruption, prevalence of document and data falsification, lack of law enforcement, human rights violations, armed conflict and sanctions — the basis for the *country context* component.
Art. 10(2)(n)	Certification schemes are *complementary information* only. An ISPO or MSPO certificate is scored as corroboration, never as a substitute for Article 9 information.
Art. 10(3)	A valid FLEGT licence creates a presumption of compliance with Article 3(b) for in-scope wood. This is recognised in scoring, noting that it covers the legality limb only, not deforestation-free.
Art. 29	Country benchmarking tiers, an input to *country context*.

2 Inputs and Pipeline

Inputs

The assessment runs on three types of input:

1. **The document itself** — a PDF, JPG or PNG of the instrument submitted as legality evidence.
2. **The origin reference library** — the document types, issuing authorities and administrative levels recognised for each covered origin, together with what can be independently checked and how.
3. **The country risk profile** — benchmarking tier, corruption indicator, curated concerns and registry verifiability for the origin.

Country risk data

Country risk data is held for six origins: cocoa in Côte d'Ivoire and Ghana, coffee in Brazil and Vietnam, and oil palm in Indonesia and Malaysia. Where a document originates outside these six, no curated profile exists; the model is instructed to cap *country context* accordingly and to say so in the assessment.

Table 2: Country risk fields and their sources

FIELD	SOURCE
Benchmarking tier	Commission Implementing Regulation (EU) 2025/1093 of 22 May 2025
Corruption indicator	Transparency International Corruption Perceptions Index 2025
Concerns	Curated per origin and tied to specific Article 2(40) categories
Registry verifiability	Curated: an overall rating plus an itemised list of what can be checked, how, and at what access level

Benchmarking caveat

EUR-Lex blocks automated retrieval, so tiers were compiled from secondary reporting of the Annex. **Sources disagree on Ghana and Indonesia.** Every entry carries a tier-confidence field recording the disagreement, and the interface surfaces a caveat where sources conflict. Confirm at the official source before relying on a tier for a compliance decision.

Status of the list

The European Parliament objected to the benchmarking list on 9 July 2025 (373–289), criticising the methodology and the classification of major deforestation fronts as standard

rather than high risk. The objection did not block the instrument and it remains in force. The first scheduled review falls in 2026, so tiers may change.

Effect of the tier

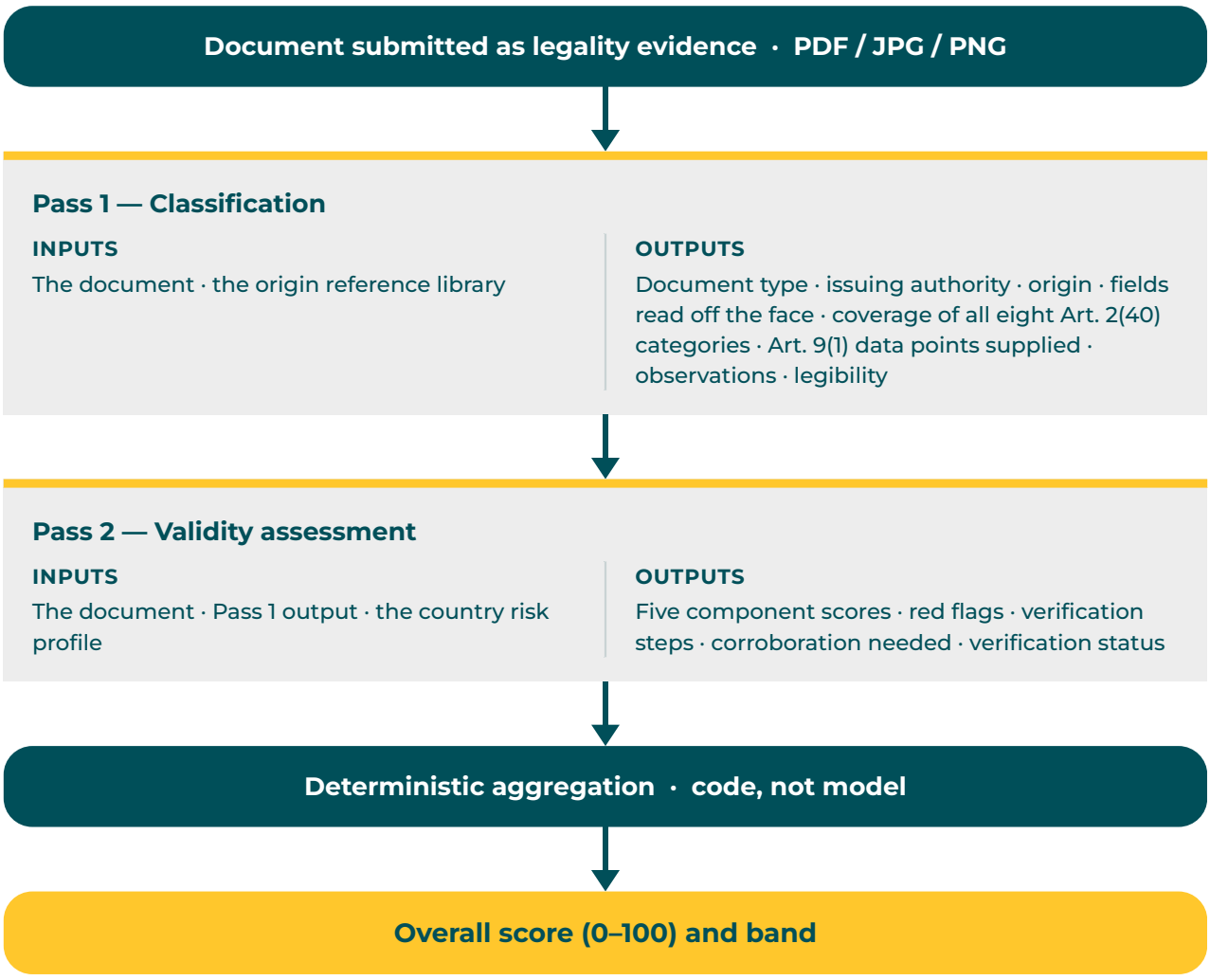
A low-risk tier permits Article 13 simplified due diligence and lowers the competent authority check rate — 1% of operators, against 3% standard and 9% high. It does **not** remove the Article 9(1) information obligation and does **not** lower the Article 9(1)(h) evidential standard. The tier is therefore an input to one component at 0.15 weight, not a shortcut.

The two-pass pipeline

Assessment runs as two sequential model passes followed by deterministic aggregation in code. Pass 2 depends on Pass 1's output, so the two cannot be parallelised. The document itself is re-sent in Pass 2 so that scoring is grounded in the artefact, not only in Pass 1's description of it.

Both passes use Claude Opus 5 with adaptive thinking at high effort and JSON-schema-constrained structured output.

Diagram 1: The assessment pipeline



3 The Five Components

Each component is scored 0–100 by the model, with a written finding citing specifics from the document. The components are designed to separate cleanly: a genuine, well-formed instrument can still be weak evidence for the purpose it was submitted for, and a single-dimension score would collapse that distinction.

Diagram 2: Component weights

Weights sum to exactly 1.00. They were set from the structure of Article 10(2), not fitted to labelled outcome data — see limitation 5.

3.1 Document integrity — weight 0.20

Is the document complete and properly executed on its face? We look for expected fields, signatures and seals where the instrument requires them, all pages supplied, legibility, no signs of alteration, and no partial extract presented as a whole.

Lowers the score: missing signature blocks, absent seals, a single page of a multi-page instrument, referenced annexes not attached, poor legibility, evidence of tampering.

3.2 Internal consistency — weight 0.20

Do the document's own contents agree with each other? Dates in sequence, areas and percentages that add up, the same party named consistently, identifiers in the correct format, an issuing authority matching the instrument type, and arithmetic that holds.

> **This is the component most likely to surface a fabrication.** In testing it independently caught that a specimen's declared Reserva Legal plus APP exceeded its declared remaining native vegetation — implying an undisclosed Forest Code deficit — a relationship not specified in the prompt.

3.3 Issuing authority plausibility — weight 0.20

Does the named issuing body exist, and could it lawfully issue this instrument for this place? This is checked against the origin reference library: the right authority for the document type, the right administrative level, and territorial competence over the stated location.

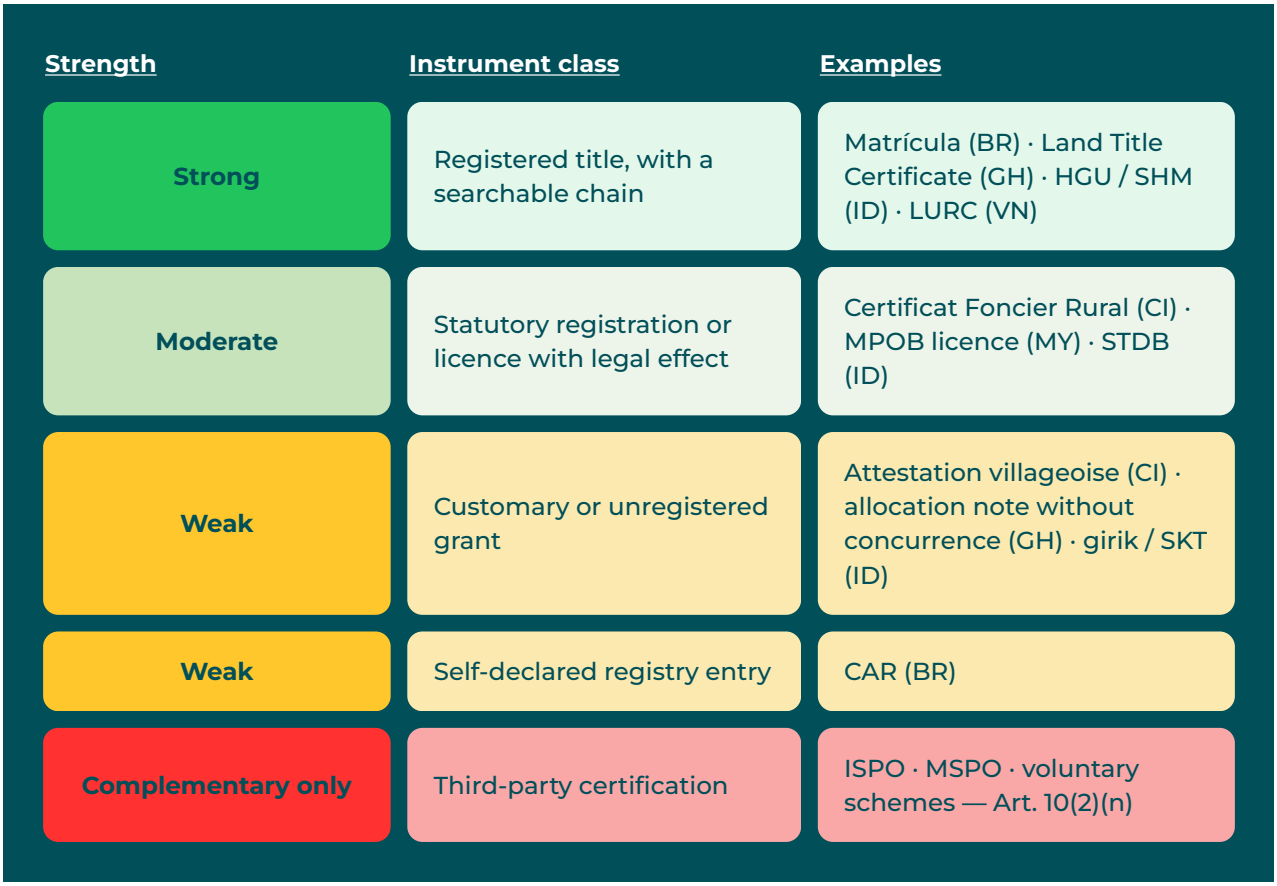
> **An instrument issued by a body that could not lawfully issue it is the single strongest indicator of a problem.** For example, a land certificate purporting to cover Indonesian forest estate that has never been released from it: BPN cannot lawfully title such land, so the certificate's existence is itself the red flag.

3.4 Evidential sufficiency for Art. 9(1)(h) — weight 0.25

How far does this document actually go towards "adequately conclusive and verifiable information"? This carries the highest weight because it speaks most directly to the statutory test.

The instrument is judged for what it is, not for what it is labelled. The governing hierarchy is set out below.

Diagram 3: Instrument strength hierarchy



3.5 Country context (Art. 10(2)(h)) — weight 0.15

How much does the operating environment reduce the weight this document can carry? **This component scores the environment, not the document.** A flawless document from a low-verifiability, high-falsification origin still scores lower here. The inputs are:

- **Art. 29 benchmarking tier** — low, standard or high
- **Corruption Perceptions Index 2025** score

- **Curated country and sector concerns** — documented falsification prevalence, enforcement gaps, tenure-system characteristics, sector-specific risks
- **Registry verifiability** — whether documents of this kind can be independently checked at all

> Verifiability is weighted heavily within this component, and in practice matters more than the benchmarking tier. Brazil (standard tier, CPI 35) scores *above* Vietnam (low tier, CPI 40) on country context, because Brazilian documents carry codes designed for public online validation while Vietnamese land records have no public register at all.

4 Scoring and Outputs

Aggregation and bands

The overall score is computed in code, not by the model. A reviewer can reproduce the number by hand from the five published component scores — which is the point.

```
overall = round( Σ ( component score × component weight ) )   clamped to [0, 100]
```

The band is derived from the computed score, never supplied independently.

Diagram 4: Score bands and what they mean

Score	Band	Meaning
80–100	Substantial	Can carry significant weight in the due diligence file, subject to the verification steps.
60–79	Moderate	Useful evidence, but not sufficient on its own for the categories it covers.
40–59	Limited	Weak evidence. Needs corroboration before it can support a negligible-risk conclusion.
0–39	Insufficient	Should not be relied on as legality evidence in its current form.

Table 3: Worked example — Vietnam Land Use Right Certificate, calibration run

COMPONENT	SCORE	WEIGHT	CONTRIBUTION
Document integrity	66	0.20	13.2
Internal consistency	70	0.20	14.0
Issuing authority plausibility	86	0.20	17.2
Evidential sufficiency	70	0.25	17.5
Country context	56	0.15	8.4
Overall		**1.00**	**70.3 → 70 (moderate)**

The determinism boundary

Deciding what belongs to the model and what belongs to code is the central design choice of this methodology.

Table 4: What is owned by code, and what is owned by the model

OWNED BY CODE	OWNED BY THE MODEL
Weights and their sum	Component scores
The weighted-sum arithmetic	Written findings justifying each score
Band thresholds and derivation	Red flags
0–100 clamping	Verification steps and their feasibility tags
Component deduplication and gap-filling	Corroboration requirements
Canonical component ordering	Overall rationale and decisive factor

Why the arithmetic moved into code

In development the model was asked to compute the weighted overall itself. On a test document it returned 26 where its own component scores implied 34. Both fell in the same band, so the conclusion was unaffected — but a score a reviewer cannot reproduce is not defensible in a compliance file. The model now supplies judgement only.

Robustness handling

The model has been observed to emit the same component twice. Post-processing keeps the first occurrence of each, restores canonical order, and fills any omitted component with

a zero score and an explicit “not assessed” finding, so a malformed response degrades visibly rather than silently skewing the total.

Schema constraint

Structured outputs reject minimum and maximum on integer fields, so the 0–100 range is stated in the field description and enforced by clamping after parsing.

Calibration anchors

Anchors are specified in the prompt to prevent clustering:

- A well-executed, internally consistent **registered title from a moderate-risk origin sits at 75–85, not 95**. The residual reflects that nothing has been verified.
- A **self-declared registry printout offered as tenure evidence belongs at 25–40**, however cleanly it is printed.
- **Above 85 is reserved** for instruments that are both intrinsically strong and independently checkable.
- There is an explicit instruction not to cluster in the middle, and to distinguish clearly between strong and weak instruments.

> **Specimen rule.** A document watermarked as a specimen, sample, template or draft, or declaring no legal effect, is decisive: document integrity and evidential sufficiency score very low, a high-severity red flag is raised, and the rationale says so plainly.

Outputs beyond the score

The score alone is not actionable. Each assessment also returns four further outputs.

Red flags

Specific indicators that the document may not be what it purports to be, or may not support what it is being used to support. Each is rated low, medium or high severity, with an explanation of why it matters. The model is instructed not to manufacture flags.

Verification steps

Ranked, concrete next actions, each tagged by feasibility. Steps that do not depend on the supplier are prioritised.

Table 5: Feasibility tags for verification steps

TAG	MEANING
Public online	Checkable now, free, without the supplier
Published	In a published list or register
In-country request	Requires a formal application in-country
Supplier dependent	Cannot proceed without the supplier's cooperation
Not verifiable	No independent source exists

Steps must be specific. *"Verify the title"* is rejected; *"run a SICAR consultation on CAR number MG-3143906-... and compare the returned polygon area against the 148.7320 ha declared"* is the standard.

Corroboration needed

The documents that should sit alongside this one to close the Article 9(1)(h) gap for the categories it touches.

Verification status

A plain sentence recording that no issuing register was contacted, and what that means for reading the score.

5 Validation and Limitations

Validation to date

Three synthetic specimens were assessed, each constructed with a known defect. Calibration mode was used so that the scoring exercised instrument strength rather than bottoming out on the specimen watermark.

Table 6: Validation results against synthetic specimens

SPECIMEN	DOC	CONSIST	AUTH	EVID	COUNTRY	OVERALL	DESIGNED DEFECT — DETECTED?
Vietnam LURC	66	70	86	70	56	**70** moderate	Front page only; reverse-side encumbrances unknown — **yes**
Brazil CAR	76	62	82	28	62	**60** moderate	Self-declared registry offered as tenure evidence — **yes**
Ghana allocation note	45	68	55	28	42	**47** limited	No Lands Commission concurrence; incomplete chain — **yes**

The Brazil row demonstrates the components separating cleanly: a genuine, well-formed federal instrument (integrity 76, authority 82) that is nonetheless weak evidence for the purpose it was submitted for (evidential 28).

> **Not yet validated:** calibration against real documents. Synthetic specimens cannot establish whether the absolute scores are correctly pitched for production use — only that the components discriminate in the expected direction.

Calibration mode

A synthetic specimen declares on its face that it has no legal effect, which correctly craters two components — and makes the fixtures useless for testing whether the scoring distinguishes a strong instrument from a weak one. Passing a calibration flag to the analysis endpoint instructs the model to score the instrument on its merits as if genuine, ignoring specimen markings for scoring purposes. All other assessment is unchanged.

Calibration mode is a test-harness flag, deliberately not exposed in the interface. It must never be used on a real document — it suppresses a decisive red flag.

Known limitations

1. **No authentication.** No document is checked against an issuing register. The score reflects intrinsic quality and context only.
2. **Absolute calibration unproven.** Component discrimination is demonstrated; the pitch of the absolute numbers is not yet validated against real documents.
3. **Benchmarking tiers are secondary-sourced** and disputed for Ghana and Indonesia.
4. **Six origins only.** Documents from elsewhere are assessed without a curated country profile and without an authority reference library, which materially weakens two components.
5. **Weights are a judgement, not an empirical fit.** They were set from the structure of Article 10(2), not derived from labelled outcome data. They are a defensible starting point, to be revised once real-world review outcomes exist.
6. **Single-document scope.** Each document is scored in isolation. The tool does not yet assess whether a *set* of documents collectively satisfies Article 9(1)(h) for a given plot or consignment — which is the actual compliance question.
7. **Non-determinism.** The model may return slightly different component scores on repeated runs of the same document. Aggregation is deterministic; the inputs to it are not.

Change control

Any change to weights, bands or component definitions changes the meaning of previously issued scores. Bump the version at the head of this document, record the change, and do not compare scores across versions without noting it.

Table 7: Extension points

CHANGE	WHAT TO DO
New origin	Add an entry to the origins reference and to the country risk profile. Prompts and interface read from both; no other change required.
New component or weight	Update the component definitions. Weights must sum to 1.00.
Band thresholds	Update the band definitions, and note the change against the version.

Appendix — data sources

SOURCE	USED FOR	AS OF
Regulation (EU) 2023/1115, consolidated text 02023R1115 — 26.12.2025	All legal definitions and obligations	Dec 2025
Commission Implementing Regulation (EU) 2025/1093	Art. 29 country benchmarking tiers	May 2025
Transparency International CPI 2025	Corruption indicator	2025
Curated origin reference library	Document types, issuing authorities, verifiability	Aug 2026

Version 0.2